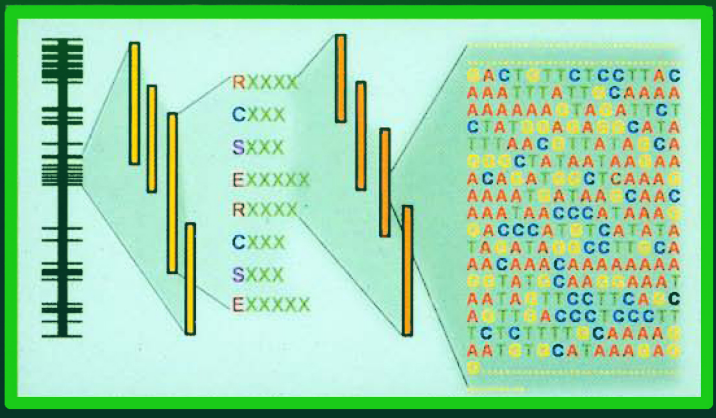
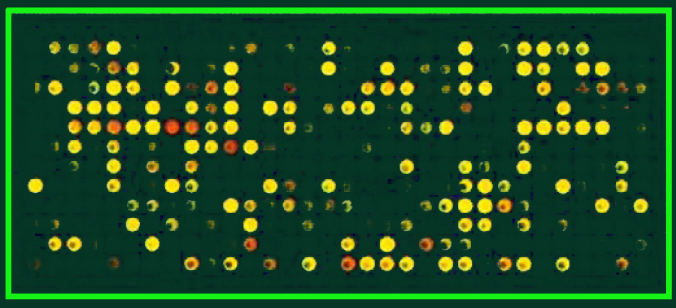
Rice Genetics IV

Edited by G.S. Khush, D.S. Brar, and B. Hardy





IRRI

SPI

Rice Genetics IV

Edited by G.S. Khush, D.S. Brar, and B. Hardy

IRRI
INTERNATIONAL RICE RESEARCH INSTITUTE

2001

SP
Science Publishers, Inc.

CIP data will be provided on request.

SCIENCE PUBLISHERS, Inc.
Post Office Box 699
Enfield, New Hampshire 03784
United States of America

Internet site: *http://www.scipub.net*

sales@scipub.net (marketing department)
editor@scipub.net (editorial department)
info@scipub.net (for all other enquiries)

ISBN 1-57808-167-X

Suggested citation:
Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. Enfield, NH (USA). Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Cover design: Juan Lazaro IV
Page makeup and composition: Mayanne A Wensceslao

Cover: A dense molecular genetic map; BAC, YAC, PAC, and EST libraries; and genomic sequences have become key resources for functional genomics of rice (figure above courtesy of M. Yano). Microarray is another new technology for understanding gene expression for complex traits. The gene expression pattern is shown on a glass slide after hybridization with fluorescently labeled probes derived from rice plants under salts stress (photo below courtesy of H. Bohnert).

Published by Science Publishers, Inc., Enfield, NH, USA.
Printed in India

Contents

FOREWORD **vii**

ACKNOWLEDGMENTS **viii**

OPENING ADDRESS **ix**
W.G. Padolina

Overview

Rice genetics from Mendel to functional genomics **3**
G.S. Khush and D.S. Brar

Application of Mendelian genetics in rice breeding **27**
J.N. Rutger and D.J. Mackill

The Rockefeller Foundation's International Program on Rice Biotechnology **39**
J.C. O'Toole, G.H. Toenniessen, T. Murashige, R.R. Harris, and R.W. Herdt

Molecular markers, genetic diversity, and evolution

Evolution and domestication of rice **63**
H. Morishima

Rice: a central genome for the genetics of all cereals **79**
M. Gale, G. Moore, and K. Devos

Phylogeny of the genus *Oryza* as revealed by molecular approaches **89**
Song Ge, Tao Sang, Bao-Rong Lu, and De-Yuan Hong

Miniature inverted repeat transposable elements help create genomic diversity in maize and rice **107**
S.R. Wessler, A. Nagel, and A. Casa

Microsatellite markers in rice: abundance, diversity, and applications **117**
S.R. McCouch, S. Temnykh, A. Lukashova, J. Coburn, G. DeClerck, S. Cartinhour, S. Harrington, M. Thomson, E. Septiningsih, M. Semon, P. Moncada, and Jiming Li

Molecular mapping and marker-assisted selection for major-gene traits in rice **137**
D.J. Mackill and Junjian Ni

QTL mapping in rice: a few critical considerations **153**
Zhi-Kang Li

Genetic and molecular basis of heterosis in rice **173**
Qifa Zhang, Jinping Hua, Sibin Yu, Lizhong Xiong, and Caiguo Xu

Structural and functional genomics

The International Rice Genome Sequencing Project: progress and prospects **189**
T. *Sasaki, T. Matsumoto, T. Baba, K. Yamamoto, J. Wu, Y. Katayose, and K. Sakata*

Strategies and techniques for finishing genomic sequence **197**
M. de la Bastide, D. Johnson, V. Balija, and W.R. McCombie

Sequence-tagged connector/DNA fingerprint framework for rice genome sequencing **215**
R.A. Wing, Yeisoo Yu, G. Presting, D. Frisch, T. Wood, Sung-Sick Woo, M. Arief Budiman, Long Mao, Hye Ran Kim, T. Rambo, E. Fang, B. Blackmon, J.L. Goicoechea, S. Higingbottom, M. Sasinowski, J. Tomkins, R.A. Dean, C. Soderlund, W.R. McCombie, R. Martienssen, M. de la Bastide, R. Wilson, and D. Johnson

Naturally occurring allelic variations as a new resource for functional genomics in rice **227**
M. Yano

Deletion mutants for functional genomics: progress in phenotyping, sequence assignment, and database development **239**
H. Leung, C. Wu, M. Baraoidan, A. Bordeos, M. Ramos, S. Madamba, P. Cabauatan, C. Vera Cruz, A. Portugal, G. Reyes, R. Bruskiewich, G. McLaren, R. Lafitte, G. Gregorio, 239J. Bennett, D. Brar, G. Khush, P. Schnable, G. Wang, and J. Leach

Generation of T-DNA insertional tagging lines in rice **253**
Gynheung An, Jong-Seong Jeon, Sichul Lee, Ki-Hong Jung, Sung-Hoon Jun, Dong-Hoon Jeong, Jinwon Lee, Seonghoe Jang, Shin-Young Lee, Kiyoung Yang, Byoungho Lee, Sinok Moon, Kyungsook An, Min-Jung Han, Jung-Hwa Yu, Namok Lee, Su-Young An, Sun-Hee Park, Eun-Sik Song, In-Soon Park, and Hyun-Sook Lee

Transposons and functional genomics in rice **263**
R. Greco, P.B.F. Ouwerkerk, C. Sallaud, A. Kohli, C. Favalli, T. Beguiristain, L. Colombo, E. Pè, P. Puigdomènech, E. Guiderdoni, P. Christou, J.H.C. Hoge, and A. Pereira

Retrotransposons of rice as a tool for the functional analysis of genes **279**
H. Hirochika, A. Mayo, M. Yamazaki, S. Takeda, K. Abe, R. Hirochika, G.K. Agrawal, T. Watanabe, K. Sugimoto, T. Sasaki, K. Murata, K. Tanaka, K. Onosato, A. Miyazaki, Y. Yamashita, and N. Kojima

Bioinformatics and the rice genome **293**
B.A. Antonio, K. Sakata, and T. Sasaki

Gene isolation and function

Enhancing deployment of genes for blast resistance: opportunities from cloning a resistance gene/avirulence gene pair **309**
B. Valent, G.T. Bryan, Y. Jia, L. Farrall, S.A. McAdams, K.N. Faulk, and M. Levy

Molecular signaling in disease resistance of rice **323**
K. Shimamoto, A. Takahashi, and T. Kawasaki

Structure, function, and evolution of disease resistance genes in rice **335**
Guo-Liang Wang, Hei Leung, and P. Ronald

Isolation of candidate genes for tolerance of abiotic stresses **345**
H.J. Bohnert, S. Kawasaki, C.B. Michalowski, Hong Wang, N. Ozturk, M. Deyholos, and D. Galbraith

Molecular dissection of cell death in rice **365**
H. Uchimiya, M. Kawai, M. Yamaguchi, and M. Umeda

Molecular tools for achieving synthetic apomixis in hybrid rice **377**
J. Bennett, X.-Z. Bi, A. Kathiresan, and G.S. Khush

Transformation

Engineering for virus resistance in rice **405**
N.M. Upadhyaya, Z. Li, M-B. Wang, S. Chen, Z-X. Gong, and P.M. Waterhouse

Transgenic approaches for generating rice tolerant of dehydration stress **423**
Z.Q. Cheng, J. Targolli, J. Su, C.K. He, F. Li, and R. Wu

High-level expression of C_4 photosynthetic genes in transgenic rice **439**
M. Matsuoka, H. Fukayama, M.S.B. Ku, and M. Miyao

Transgene integration, organization, and expression in cereals **449**
P. Christou, R.M. Twyman, Xiangdong Fu, E. Wegel, A. Kohli, and E. Stoger

Gene silencing and its reactivation in transgenic rice **465**
T.C. Hall, S.P. Kumpatla, P. Kharb, L. Iyer, M. Cervera, Y. Jiang, T. Wang, G. Yang, P. Teerawanichpan, J. Narangajavana, and J. Dong

Workshop reports
Rice molecular breeding workshop **484**
Functional genomics workshop **485**
Bioinformatics workshop **487**

RICE GENETICS COOPERATIVE **488**

Foreword

This is the Fourth International Rice Genetics Symposium in the series of symposia held at IRRI every five years. The first, held in 1985, led to the birth of the Rice Genetics Cooperative (RGC). The RGC took the lead in organizing these symposia and greatly enhanced international collaboration. In the same year, the Rockefeller Foundation established its International Program on Rice Biotechnology, which has played a major role in advancing frontiers of knowledge on cellular and molecular genetics of rice, international collaboration, and human resource development. In the second symposium, a unified system of numbering rice chromosomes and linkage groups was adopted. The orientation of classical and molecular maps was one of the many highlights of the third symposium. The fourth symposium brought together 520 rice scientists from 32 countries and provided an excellent forum for scientists from developed and developing countries to share information on the latest advances in rice science and to develop collaborative research arrangements.

The symposium featured 31 plenary lectures in seven sessions, 130 contributory papers in six concurrent sessions, and 240 poster presentations on different aspects of rice genetics. World famous geneticists delivered plenary lectures covering a wide range of topics from classical genetics to the most advanced research on sequencing of the rice genome and functional genomics. Various sessions provided an important forum for reviewing the latest advances in rice research and for in-depth discussion and exchange of information on classical genetics, biosystematics and evolution, molecular markers, transformation, genome organization, gene isolation, regulation of gene expression, sequencing of the rice genome, and bioinformatics.

Also during the symposium, three workshops were held on molecular breeding, functional genomics, and bioinformatics. These workshops led to initiatives among the national agricultural research and extension systems to join the international program in these three areas.

I am pleased to see that the plenary lectures have been published in this book. The concurrent and poster presentations will be published in a supplementary volume.

I would like to thank the organizing committee and other colleagues at IRRI who have devoted a great deal of time to organizing this symposium. IRRI would especially like to acknowledge the Rockefeller Foundation for its financial support for this symposium.

RONALD P. CANTRELL
Director General
International Rice Research Institute

Acknowledgments

We would like to thank the following members of the organizing committee for the Fourth International Rice Genetics Symposium: John Bennett, Swapan Datta, Mike Jackson, Zhikang Li, and Hei Leung. We would also like to give special thanks to Ronald P. Cantrell, Ren Wang, and William Padolina for their scientific and financial support. Financial support provided by the Rockefeller Foundation is gratefully acknowledged. Thanks are due to Ike Navarro for logistics support. We also thank Yollie Aranguren, Elma Nicolas, Emily Alcantara, and George Reyes for secretarial help and processing of the manuscripts. We appreciate the valuable services provided by different teams in the following areas: Communication and Publications Services, Food and Housing Services, Visitors, Exhibition, and Conference Services, Physical Plant Services, and a secretarial pool.

Opening address

W.G. Padolina

Rice is the principal food of nearly half of the world's people and more than 90% of the crop is grown in developing countries, where food supply is an acute problem. Much success has been gained over the past 35 years. Rice production has more than doubled, from 257 million t to 596 million t in 1999. This phenomenal increase can be attributed to the large-scale adoption of improved rice varieties and technology developed at IRRI and by national rice improvement programs. These gains were made using science based on Mendelian genetics and conventional breeding techniques.

An independent study by a team led by Dr. Robert Evenson of Yale University and Dr. Hans Gregersen of the University of Minnesota found that the development of improved rice varieties between 1970 and 1995 had substantial impact in four major areas. Their findings indicate that, were it not for the development of improved varieties, rice prices for consumers could have been up to 41% higher, rice-producing nations would be importing up to 8% more food, millions of hectares of forest and other fragile ecosystems would have been lost, and from 1.5% to 2% more children would have been malnourished in developing countries. This seemingly small figure in percentage terms translates into millions of better-fed children in reality.

While we can all be proud of some of these gains, our work is not yet done. The work to increase rice production averted large-scale famines predicted by some experts. However, the specter of food shortages looms once again. Estimates made by experts indicate that the rate of increase in rice production has slowed down to a point where it is below the rate of increase in rice consumers. Agricultural economists estimate that we have to produce much more rice by 2030.

No new areas can be used for rice farming. Arable land is becoming more and more scarce for various reasons. To cope with these constraints, new rice varieties that combine higher yield potential with superior grain quality are needed.

Einstein once said that problems could not be solved with the same kind of intelligence that created them. We are here to discuss new ideas and innovative approaches so that we can find solutions to providing enough good-quality food for all.

Recent advances in cellular and molecular biology have come in the "nick of time" to provide us with new tools for developing rice varieties of the future. Only 15 years ago, the use of these modern tools in rice research was considered far behind that of other important food crops such as maize, wheat, and tomato. However, the past 15 years have seen a major explosion in knowledge about rice genetics and rice is now considered a model plant for research on cereal crops. Several fortunate developments have led to this.

First, the Rice Genetics Cooperative (RGC) was organized in 1985 during the First International Rice Genetics Symposium here at IRRI. It has greatly enhanced international collaboration in rice genetics. The RGC formulated the rules for gene symbolization in rice and it monitors gene symbols and coordinates linkage mapping. It has established two genetic stock centers and publishes the Rice Genetics Newsletter annually. It also formulated a uniform chromosome numbering system during the Second International Rice Genetics Symposium.

Second, the Rockefeller Foundation established the International Program on Rice Biotechnology in 1985 and has contributed greatly to advancing the frontiers of knowledge in rice genetics and building scientific capacity in rice biotechnology, especially in developing countries.

Third, the establishment of the Rice Genome Research Program (RGRP) in Tsukuba, Japan, has contributed significant new knowledge on rice molecular biology far beyond anyone's expectations.

IRRI has been privileged to participate actively in the RGC, RGRP, and the Rockefeller Program on Rice Biotechnology.

Sir Aaron Klug, Director of the MRC Laboratory of Molecular Biology in Cambridge, once said, "Scientific breakthroughs don't come from committees, they come from the brains and hands of individuals who need a place to work and who need an infrastructure to support them."

It is in this light that we at IRRI feel honored to host this gathering of the world's best minds in rice genetics. We hope that this meeting will provide for the free exchange of ideas to enable everyone to work effectively in their chosen areas of research.

Like the previous three symposia, we look forward to interesting discussions and the birth of new ideas and approaches. This symposium features 31 invited plenary lectures, 130 papers in concurrent sessions, 240 poster presentations, and three workshops on molecular breeding, functional genomics, and bioinformatics. We trust that, by providing a venue for these discussions, IRRI can fulfill its role, which in the words of Dr. Ronald P. Cantrell, IRRI's director general, "is to objectively evaluate new strategies and options," especially in the field of biotechnology. After this symposium, we hope to be better informed to handle the issues and concerns that continue to permeate global debates on biotechnology, particularly on genetically modified organisms. Furthermore, the impact of rapid population growth on the use of the finite resources of our Earth speaks eloquently of the need to review our strategies for development. The notion of sustainable development has been discussed widely. In fact, one can hardly find anyone who is opposed to it. But, 13 years after the concept of sustainable development came to the fore, it remains a vague and imprecise concept. I am fully aware that technical knowledge such as genetics alone will not improve sustainability. But new knowledge can always be a component contributing to sustainability. It is important that we keep this in mind during our discussions in the next few days.

Notes

Author's address: Deputy Director General for Partnerships, International Rice Research Institute, Los Baños, Laguna, Philippines.

Acknowledgments: Finally, let me acknowledge those who worked hard to make this symposium a success. These efforts were ably handled by Dr. Gurdev Khush and Dr. Darshan Brar. I should also like to acknowledge the excellent staff support provided by Mr. Ike Navarro, who led the team that worked and continued to work behind the scenes to make sure that all activities during this symposium ran smoothly. Also, we would like to acknowledge the financial assistance that the Rockefeller Foundation extended to this symposium.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Overview

Rice genetics from Mendel to functional genomics

G.S. Khush and D.S. Brar

From being a poor cousin to maize, wheat, and tomato for genetic knowledge as recently as the 1980s, rice has become a model plant for molecular genetic research. Numerous scientists in laboratories worldwide have helped make rice a favored higher plant for molecular and cellular genetic studies. Below are some of the major advances in this rapid progress in rice genetics:

- Van der Stok in 1908 for the first time reported Mendelian segregation in rice.
- Kuwada established the basic chromosome number of rice to be 24 in 1910.
- The first linkage in rice was reported by Parnell et al in 1917.
- Dr. K. Ramiah advocated the standardization of gene symbols in rice.
- Kadam and Ramiah published a review of gene symbols for the first time in 1943.
- Gene symbolization was discussed by the International Rice Commission working party on rice breeding at its Sixth Session in Penang, Malaysia, in 1955.
- Shastry et al numbered the chromosomes in decreasing order of length at the pachytene stage of meiosis in 1960.
- Rules for gene symbolization were reviewed during the symposium on rice genetics and cytogenetics held at IRRI in 1963.
- Nagao and Takahashi proposed the 12 linkage groups of rice in 1963.
- Regeneration of haploids from anther culture was reported by Niizeki and Oono in 1968.
- Independence of linkage groups was tested by Iwata and Omura in Japan and by G.S. Khush et al at IRRI in 1984 through trisomic analysis.
- Publication of the *Rice Genetics Newsletter* began in 1984 under the editorship of H.I. Oka and G.S. Khush. The First International Rice Genetics Symposium was held at IRRI in 1985 and the Rice Genetics Cooperative was established for international collaboration in rice genetics.

- The Rockefeller Foundation established an International Program on Rice Biotechnology in 1985.
- Yamada et al obtained regeneration from protoplasts in 1985.
- McCouch et al constructed the first molecular genetic linkage map in 1988.
- Transgenic rice plants were produced first by three groups: Toriyama et al, Zhang and Wu, and Zhang et al in 1988.
- The Second International Rice Genetics Symposium was held at IRRI in 1990 and a uniform chromosome numbering system was established.
- The Rice Genome Research Program (RGRP) began at Tsukuba in 1991.
- Ahn and Tanksley constructed comparative linkage maps of the rice and maize genomes in 1993.
- The yeast artificial chromosome (YAC) library in rice was established by the RGRP and a bacterial artificial chromosome (BAC) library by Wang et al in 1995.
- The first agronomically important gene in rice, *Xa21*, was cloned by Song et al through map-based cloning in 1995.
- The Third International Rice Genetics Symposium was held in Manila in 1995 and the correct orientation of morphological and molecular genetic maps was established.
- An international network on rice genome sequencing was established in 1998 under the leadership of the RGRP.
- Projects on functional genomics began in 1999.

Rice is now the model plant for genetic research among crop plants. However, this was not the case till about 15 years ago. In fact, the status of rice genetics was far behind that of maize, wheat, tomato, and barley. An agreed system of chromosome numbering in rice did not exist. Linkage groups were poorly known and their independence had not been tested. Nomenclature for gene symbolization was not uniformly followed and rice geneticists had no platform to discuss and enhance international collaboration in rice genetics. However, a series of events starting in 1985 have contributed much to elevate rice to its present state of preeminence. Here we review the major milestones in rice genetics during the 20th century.

Rice karyotype

Kuwada (1910) first reported the basic chromosome number (n=12) in rice from the study of microsporogenesis, megasporogenesis, and mitosis. Since then, many workers have confirmed this number. The somatic chromosomes of rice, however, are very small and difficult to distinguish from each other. The individual chromosomes are easy to identify at the pachytene stage of meiosis and Shastry et al (1960) described the pachytene chromosome complement for the first time. They numbered the chromosomes in decreasing order of length, with the longest as 1 and the shortest as 12. Kurata et al (1981) also analyzed the pachytene chromosome complement of Japa-

nese cultivar Nipponbare. The chromosome designations in two studies agreed remarkably well except for chromosomes 11 and 12, which were interchanged. The Rice Genetics Cooperative adopted the pachytene numbering system of Shastry et al (1960) in 1985.

Genome analysis

Various approaches involving morphological differentiation, meiotic chromosome pairing in F_1 hybrids, molecular divergence analysis, and fraction I protein have been used in genome analysis and in determining species relationships in *Oryza*. Morinaga (1939), based on chromosome pairing in F_1 hybrids, concluded that *O. glaberrima, O. breviligulata, O. perennis, O. cubensis, O. sativa* var. *fatua*, and *O. sativa* var. *spontanea* have the same genomic constitution designated as the AA genome. Nezu et al (1960) studied chromosome pairing in F_1 hybrids of *O. sativa* with related diploid species such as *O. perennis, O. glaberrima, O. stapfii*, and *O. breviligulata*. The hybrids invariably showed 12 bivalents at meiosis, indicating that these species also have the AA genome. Lu et al (1998) observed normal chromosome pairing in hybrids of four A genome species, *O. rufipogon, O. nivara, O. glumaepatula*, and *O. meridionalis*.

On the basis of meiotic chromosome pairing in F_1 hybrids, *O. sativa*, *O. officinalis, O. minuta*, and *O. latifolia* genomes were designated as AA (Morinaga 1942), CC (Morinaga and Kuriyama 1959), BBCC, and CCDD (Morinaga 1943), respectively. Li et al (1963) studied F_1 hybrids of *O. sativa* × *O. australiensis* and *O. minuta* × *O. australiensis* and suggested the E genome for *O. australiensis*. Katayama (1967) reported that diploid *O. punctata* has the BB genome. Li et al (1961) and Wuu et al (1963) designated the F genome for *O. brachyantha*. On the basis of chromosome pairing in F_1 hybrids, various authors have assigned the genome symbol AA for the *Sativa* complex; BB, CC, BBCC, CCDD, and EE for the *Officinalis* complex; and FF for *O. brachyantha*.

Because of strong crossability barriers between *Meyeriana* and *Ridleyi* complexes, hybrids are difficult to produce and assigning genomes based on meiotic pairing could not be carried out. Under such situations, an alternate approach based on total genomic DNA hybridization and molecular divergence analysis has been used; the genome GG has been proposed for the diploids of the *O. meyeriana* complex and HHJJ for the allotetraploids of the *O. ridleyi* complex (Aggarwal et al 1997). Ge et al (1999) based on sequence analysis of nuclear genes (*Adh1*, *Adh2*) and a chloroplast gene (*matK*) proposed the HHKK genome for *O. schlechteri* and *Porteresia coarctata*, further suggesting that *P. coarctata* should be treated as an *Oryza* species.

Results of random fragment length polymorphism (RFLP) analysis (Wang et al 1992), amplified fragment length polymorphism (AFLP) analysis (Aggarwal et al 1999), sequence analysis of genes (Ge et al 1999), and seed protein analysis (Sarkar and Raina 1992) support the genomic classification based on morphological and cytological data. Table 1 shows the genomic constitution of different species of *Oryza*.

Table 1. Chromosome number, genomic composition, and geographical distribution of *Oryza* species. Modified from Khush and Brar (2001).

Species	2n	Genome	Distribution
O. sativa complex			
O. sativa L.	24	AA	Worldwide
O. nivara Sharma et Shastry	24	AA	Tropical and subtropical Asia
O. rufipogon Griff.	24	AA	Tropical and subtropical Asia, tropical Australia
O. breviligulata A. Chev. et Roehr.	24	AA	Africa
O. glaberrima Steud.	24	AA	West Africa
O. longistaminata A. Chev. et Roehr.	24	AA	Africa
O. meridionalis Ng	24	AA	Tropical Australia
O. glumaepatula Steud.	24	AA	South and Central America
O. officinalis complex			
O. punctata Kotschy ex Steud.	24, 48	BB, BBCC	Africa
O. minuta J.S. Presl. ex C.B. Presl.	48	BBCC	Philippines and Papua New Guinea
O. officinalis Wall ex Watt	24	CC	Tropical and subtropical Asia, tropical Australia
O. rhizomatis Vaughan	24	CC	Sri Lanka
O. eichingeri A. Peter	24	CC	South Asia and East Africa
O. latifolia Desv.	48	CCDD	South and Central America
O. alta Swallen	48	CCDD	South and Central America
O. grandiglumis (Doell) Prod.	48	CCDD	South and Central America
O. australiensis Domin.	24	EE	Tropical Australia
O. meyeriana complex			
O. granulata Nees et Arn. ex Watt	24	GG	South and Southeast Asia
O. meyeriana (Zoll. et Mor. ex Steud.) Baill.	24	GG	Southeast Asia
O. ridleyi complex			
O. longiglumis Jansen	48	HHJJ	Irian Jaya (Indonesia) and Papua New Guinea
O. ridleyi Hook. F.	48	HHJJ	South Asia
Unclassified			
O. brachyantha A. Chev. et Roehr.	24	FF	Africa
O. schlechteri Pilger	48	HHKK	Papua New Guinea

Linkage groups

The first report of Mendelian segregation in rice was by a Dutch botanist, Van der Stok (1908), in Indonesia. The first report of linkage in rice was by Parnell et al (1917), who reported linkage between black hull and colored internode. Yamaguchi (1921, 1926) reported linkage between apiculus color and waxy endosperm and set up the so-called waxy linkage group. Nagamatsu (1942) found linkage among purple leaf, liguleless, and phenol-staining reaction and established the purple leaf linkage group. Jodon (1956) proposed seven linkage groups on the basis of information available up to that time. Nagao and Takahashi (1963) proposed 12 linkage groups (I–XII) for the first time. Iwata and Omura (1975, 1976) and Khush et al (1984) tested the independence of linkage groups through trisomic tests. Three linkage groups of Nagao and Takahashi (1963) were found to belong to one chromosome and the number of linkage groups was reduced to nine. Markers for the three remaining chromosomes were identified and 12 groups were established (Khush et al 1984).

Gene symbolization in rice

In the absence of any rules for assigning gene symbols, different symbols were assigned to the same genes. As an example, the gene symbols *m, U, am, gl, g,* and *wx* were assigned to the gene for glutinous endosperm (Kihara 1964). Likewise, the same gene symbols were assigned to designate entirely different genes. The gene symbol *gl* has been used to designate genes for glutinous endosperm and glabrous leaves. Dr. K. Ramiah of India was the first geneticist to advocate the standardization of gene symbols in rice. Kadam and Ramiah (1943) published a review of the existing positions on the use of gene symbols. They suggested rules for gene symbolization wherein many conventions used in maize, cotton, and *Drosophila* were adopted. Later, Nagao (1951) also proposed a partly modified system of gene symbolization.

Gene symbols were discussed by the International Rice Commission (IRC) Working Party on Rice Breeding at its Sixth Session in Penang, Malaysia, in 1955. The Working Party suggested that a unified system of gene nomenclature be evolved and appointed a committee for that purpose with N.E. Jodon of the United States as convenor. The report of that committee was accepted by the IRC Working Party on Rice Production and Protection in 1959 and published in the IRC Newsletter (IRC 1959). The rules for gene nomenclature and gene symbols were reviewed during the symposium on Rice Genetics and Cytogenetics held at the International Rice Research Institute (IRRI) in 1963 and were accepted by the participants. Unfortunately, however, no mechanism existed for monitoring the gene symbols till the establishment of the Rice Genetics Cooperative in 1985.

Associating linkage groups with chromosomes

Primary trisomics, monosomics, translocations, and chromosomal deficiencies are useful for associating linkage groups with respective chromosomes. Primary trisomics of rice were first produced by Ramanujam (1937). Since then, several workers have produced primary trisomics (Khush and Kinoshita 1991). However, only the primary trisomics produced by Iwata et al (1970, 1984) and Khush et al (1984) were used for associating linkage groups with the respective chromosomes through modified segregation ratios. The extra chromosomes of trisomics of Iwata et al (1984) were identified at somatic metaphase and those of Khush et al (1984) at the pachytene stage of meiosis. When the chromosome-linkage group associations determined by the two groups were compared, the results did not agree for some of the chromosome-linkage group associations. These differences were resolved through mutual discussions between two groups and a unified system of numbering chromosome and linkage groups was agreed upon and accepted by rice geneticists during the Second International Rice Genetics Symposium held at IRRI in 1990. The orientation of linkage groups and position of centromeres were determined by Singh et al (1996b) through the use of secondary and telotrisomics. Figure 1 shows the latest linkage map of rice.

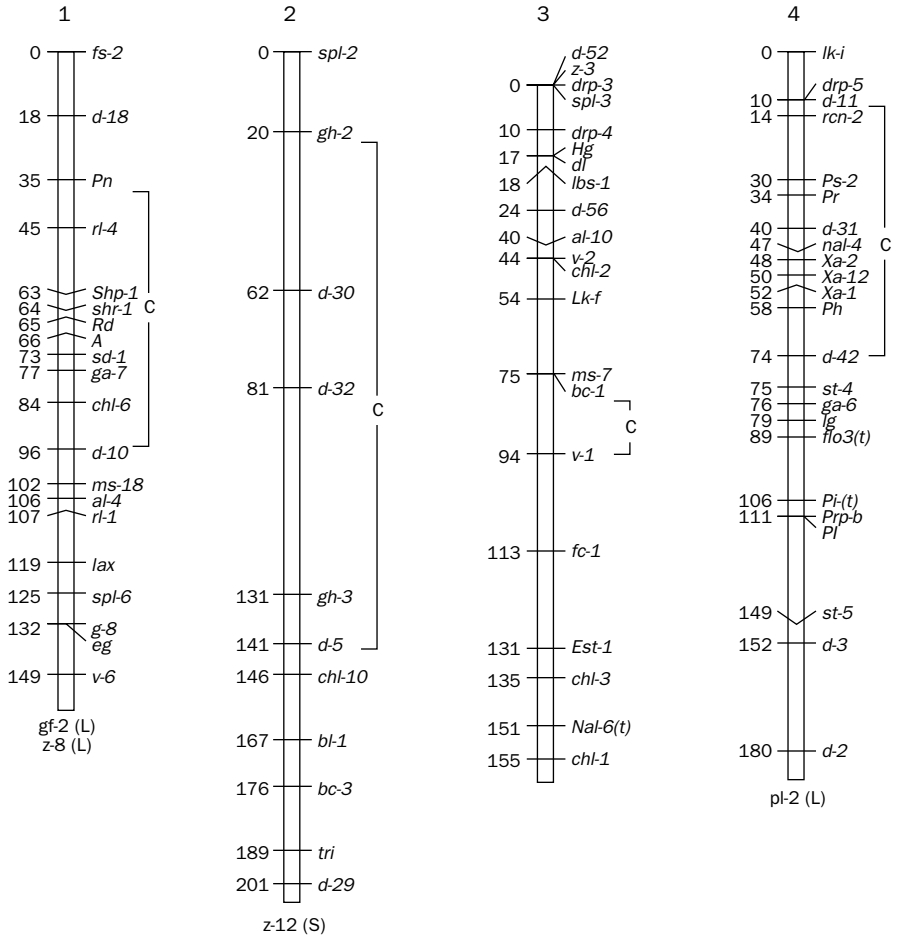
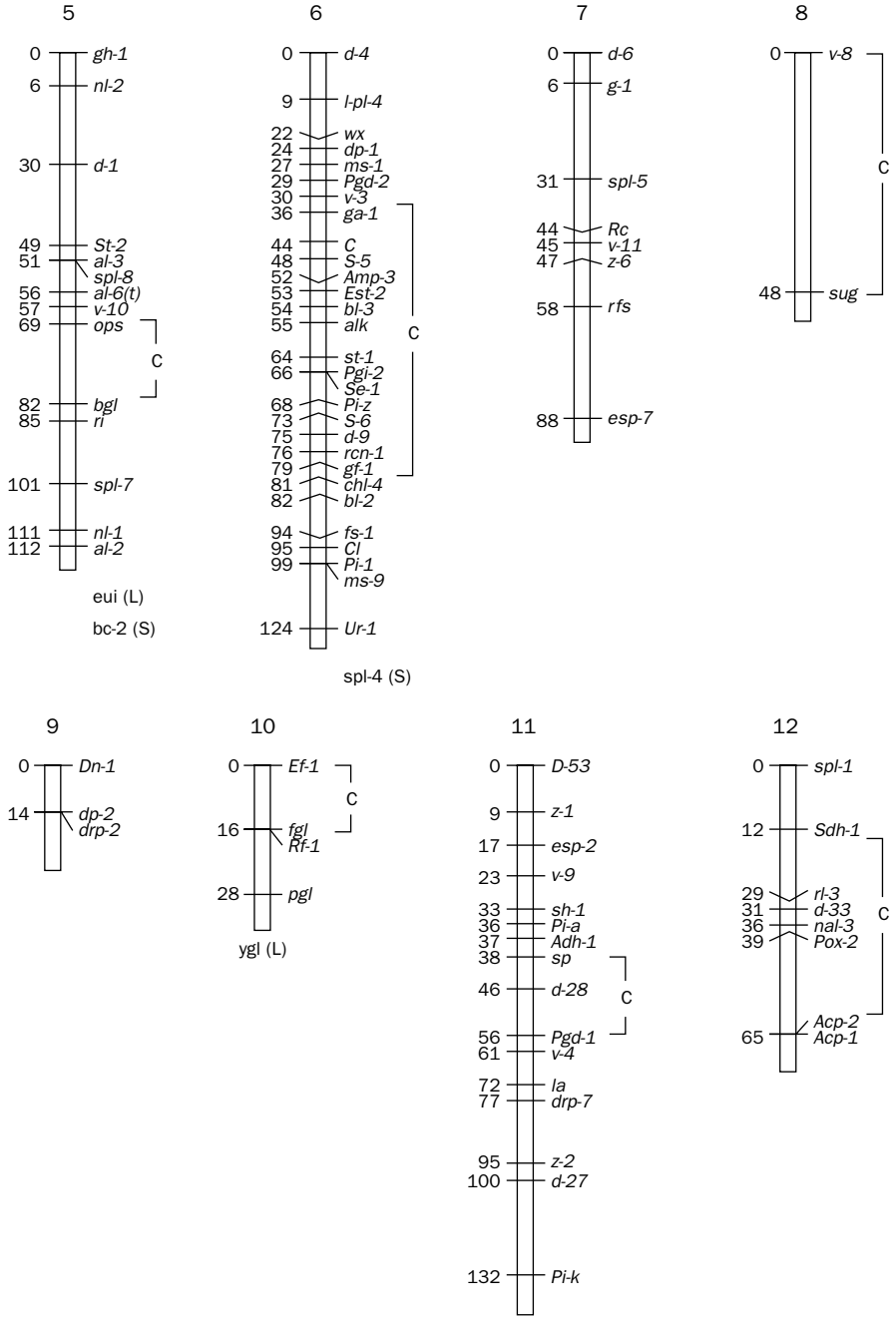


Fig. 1. Chromosome map of rice showing correct orientation of linkage groups. Markers below the map are located on respective chromosome but are not yet mapped. S and L in parentheses indicate their location on short and long arms, respectively. The centromere regions are indicated by a vertical line on the right side of each linkage group. (Adapted from Singh et al 1996b.)

International collaboration on rice research

As mentioned earlier, an international mechanism for coordination and collaboration in rice did not exist. On the basis of correspondence among IRRI scientists (Dr. M.S. Swaminathan and G.S. Khush) and officials of the Japanese Rice Genetics Information Committee (H.I. Oka, T. Kinoshita, and Y. Futsuhara), it was decided to publish an annual *Rice Genetics Newsletter* (*RGN*) to exchange information among rice scientists and to hold an international rice genetics symposium. The first volume of *RGN* was published in 1984 under the editorship of Drs. H.I. Oka and G.S. Khush. It contained the proposed rules for gene nomenclature. The First International Rice


5
0 gh-1
6 nl-2
30 d-1
49 St-2
51 al-3
spl-8
56 al-6(t)
57 v-10
69 ops
C
82 bgl
85 ri
101 spl-7
111 nl-1
112 al-2
eui (L)
bc-2 (S)
6
0 d-4
9 l-pl-4
22 wx
24 dp-1
27 ms-1
29 Pgd-2
30 v-3
36 ga-1
44 C
48 S-5
52 Amp-3
53 Est-2
54 bl-3
55 alk
C
64 st-1
66 Pgi-2
Se-1
68 Pi-z
73 S-6
75 d-9
76 rcn-1
79 gf-1
81 chl-4
82 bl-2
94 fs-1
95 Cl
99 Pi-1
ms-9
124 Ur-1
spl-4 (S)
7
0 d-6
6 g-1
31 spl-5
44 Rc
45 v-11
47 z-6
58 rfs
88 esp-7
8
0 v-8
C
48 sug
9
0 Dn-1
14 dp-2
drp-2
10
0 Ef-1
C
16 fgl
Rf-1
28 pgl
ygl (L)
11
0 D-53
9 z-1
17 esp-2
23 v-9
33 sh-1
36 Pi-a
37 Adh-1
38 sp
C
46 d-28
56 Pgd-1
61 v-4
72 la
77 drp-7
95 z-2
100 d-27
132 Pi-k
12
0 spl-1
12 Sdh-1
C
29 rl-3
31 d-33
36 nal-3
39 Pox-2
Acp-2
65 Acp-1

Genetics Symposium was held at IRRI in 1985. Participants decided to organize the Rice Genetics Cooperative (RGC) to promote international cooperation in rice genetics. One of the standing committees of the RGC monitors and coordinates gene symbols and another publishes the *RGN* once a year. The RGC also holds international rice genetics symposia at five-year intervals to discuss the latest advances in rice genetics. Four international rice genetics symposia have been held to date: in 1985, 1990, 1995, and 2000.

Another event that led to international collaboration and advances in rice genetics was the establishment of the International Program on Rice Biotechnology by the Rockefeller Foundation in 1985. The Foundation invested $100 million in research grants and human resource development over a 15-year period (1985-2000). Major advances were made in the molecular and cellular biology of rice by the Foundation grantees and several hundred scientists from developing countries received training. Biennial meetings of the rice scientists organized by the Foundation did much to develop camaraderie among rice geneticists.

The third development was the establishment of the Rice Genome Research Program (RGRP) at Tsukuba, Japan, in 1991. Advances in molecular marker research led to the construction of a densely populated molecular genetic map, preparation of BAC, YAC, and P1-derived artificial chromosome (PAC) libraries, physical map-facilitated gene cloning, and molecular marker-aided selection. In 1997, the Tsukuba program began the rice genome sequencing project, which eventually became the International Rice Genome Sequencing Project (IRGSP).

Genetics of disease and insect resistance

One of the major advances in rice genetics has been the identification of genes for major diseases and insects that have been employed for developing disease- and insect-resistant varieties. Sakaguchi (1967) identified *Xa1* and *Xa2* for resistance to bacterial blight. Twenty-two genes (*Xa1, Xa2, Xa3, Xa4, xa5, Xa7, xa8, Xa10, Xa11, Xa12, xa13, Xa14, xa15, Xa16, Xa17, Xa18, xa19, xa20, Xa21, Xa22, Xa23, Xa24)* for resistance to bacterial blight are now known. The first gene for resistance to blast, *Pia*, was identified by Shinoda et al (1971). Since then, 25 genes for resistance to blast have been identified. Athwal et al (1971) identified *Bph1* and *bph2* for resistance to brown planthopper and *Glh1, Glh2,* and *Glh3* for resistance to green leafhopper. Eleven genes (*Bph1, bph2, Bph3, bph4, bph5, Bph6, bph7, bph8, Bph9, Bph10, bph11*) for resistance to brown planthopper, nine (*Glh1, Glh2, Glh3, glh4, Glh5, Glh6, Glh7, glh8, Glh9*) for resistance to green leafhopper, six (*Wbph1, Wbph2, Wbph3, wbph4, Wbph5, Wbph6*) for resistance to whitebacked planthopper, three (*Zlh1, Zlh2, Zlh3*) for resistance to zigzag leafhopper, two (*Grh1, Grh2*) for green rice leafhopper, and six (*Gm1, Gm2, gm3, Gm4, Gm5, Gm6t*) for resistance to gall midge have been identified.

Tissue culture

As early as 1968, Niizeki and Oono reported the production of haploids from anther culture of rice. Since then, the anther culture technique has been greatly refined. It is now possible to produce haploids from the anther culture of many japonica and indica rice varieties although the frequency of regenerated plants is relatively lower in indicas. Several varieties and improved breeding lines have been developed through anther culture in China. The Republic of Korea has also released several rice varieties through anther culture. One anther culture–derived line, IR51500-AC-11-1, has been named as a variety (PSBRc50) in the Philippines. Heszky and Simon-Kiss (1992) released rice variety DAMA in Hungary through gametoclonal variation. Most of the anther culture–derived varieties are japonicas; however, indica varieties are generally recalcitrant. The doubled-haploid (DH) lines produced from indica × japonica (IR64 × Azucena, CT9993 × IR62226) are being used in molecular mapping of genes and quantitative trait loci (QTLs) governing agronomic traits.

Major progress has been made in plant regeneration from protoplasts of indica and japonica rice. Yamada et al (1985) were the first to regenerate plants from rice protoplasts. Since then, many laboratories have regenerated plants of several japonica and indica cultivars. Yang et al (1989) produced cybrid plants through the donor-recipient protoplast fusion method by electrofusing the gamma-irradiated protoplast of A-58 CMS and the iodoacetamine-treated protoplasts of the fertile cultivar Fujiminori. The donor-recipient method has been successfully used to transfer cytoplasmic male sterility (CMS) from indica variety Chinsurah Boro II into japonica cultivars. Hayashi et al (1988) produced somatic hybrids through protoplast fusion between rice and four wild species of *Oryza*. Ogura and Shimamoto (1991) identified useful somaclonal variants from protoplast-regenerated progenies of japonica variety Koshihikari, and a new cultivar, Hatsume, was released.

Transformation

Protocols for transformations of rice are well known. Biolistic and *Agrobacterium*-mediated transformation have greatly facilitated the production of transgenic rice. Transgenic plants were first produced by three groups: Toriya et al (1988), Zhang et al (1988), and Zhang and Wu (1988). Since then, transgenic plants have been produced in many laboratories in both indica and japonica rice carrying genes for herbicide tolerance, resistance to stem borer, virus tolerance, resistance to fungal and bacterial pathogens, and other agronomic traits.

Goto et al (1999) introduced the ferritin gene into rice. The transgenic plants showed increased accumulation of iron in the grain. Further studies are needed to determine the usefulness of engineered rice as a source of dietary iron. Several laboratories have produced transgenic rice, mainly through protoplast-mediated DNA transformation but also via microprojectile bombardment. Cheng et al (1998) produced more than 2,600 transgenic rice plants through *Agrobacterium*-mediated transformation. Chen et al (1998) produced transgenic rice carrying multiple transgenes after co-

bombarding embryogenic cells with a mixture of 14 different puc-based plasmids. Eighty-five percent of the R_0 plants contained more than two and another 17% had more than nine of the target genes. Plants containing multiple transgenes had normal morphology and 63% set viable seeds.

Tu et al (2000) evaluated transgenic elite commercial hybrid rice expressing the *Bacillus thuringiensis* (*Bt*) genes *cry1A*(b) and *cry1A*(c) under field conditions. The transgenic plants showed a high level of protection to both leaffolder and yellow stem borer. More recently, Ye et al (2000) produced transgenic rice ("golden rice") with the provitamin-A (β-carotene) biosynthetic pathway engineered into the rice endosperm. *Agrobacterium*-mediated cotransformation was used to introduce three genes: phytoene synthase (*psy*), phytoene desaturase (*crtl*), and lycopene β-carotene (*lcy*). The genes *psy* and *lcy* originated from daffodil (*Narcissus pseudonarcissus*) and *crtl* from the bacterium *Erwinia uredovora*. High-performance liquid chromatography (HPLC) analysis of transgenic rice seeds showed the presence of β-carotene. Major efforts are under way to introduce genes for improved yield potential, disease and insect resistance, and abiotic stress tolerance into rice.

Integration of transgenes and enhancement of transgene expression are being investigated. Takano et al (1997) found illegitimate recombination accompanying rearrangement in transgenic plants. Kohli et al (1998) reported that transgene integration in rice is a two-phase mechanism in which the original site of transgene integration acts as a hot spot, facilitating subsequent integration of successive transgenic molecules at the same locus. Transformation through particle bombardment generally results in a single transgenic locus as a result of this two-phase integration mechanism. Further analysis of transgenic rice lines carrying a range of transforming plasmid rearrangements revealed a recombination hot spot in the CaMV 35S promoter with predominance of microhomology-mediated recombination (Kohli et al 1999). Vain et al (1999) analyzed the role of matrix attachment regions (MAR) in gene expression and suggested the use of one or several MAR sequences to flank the genes of interest to maximize high-level expression of transgenes.

Molecular linkage maps

The construction of a comprehensive molecular genetic map of rice containing more than 2,200 DNA markers has been a major advance in rice genetics. A molecular genetic map of rice containing 135 markers based on RFLPs was developed at Cornell University (Ithaca, New York, USA) in collaboration with IRRI (McCouch et al 1988). The map was generated from an indica × tropical japonica F_2 population. Primary trisomics were used to assign linkage groups to each of the 12 chromosomes. A second RFLP map containing 322 markers based on an indica × japonica cross was prepared by Saito et al (1991). Causse et al (1994) developed the map, which had 726 markers. The mapping population was derived from the cross of cultivated rice (*O. sativa*) and wild species *O. longistaminata*. Kurata et al (1994b) developed a map consisting of 1,383 DNA markers under the Rice Genome Research Program in Ja-

pan. The markers, distributed along 1,575 cM on 12 linkage groups, had 883 cDNAs, 265 genomic DNAs, 147 randomly amplified polymorphic DNAs (RAPDs), and 88 other DNA markers. Harushima et al (1998) constructed a comprehensive map consisting of 2,275 markers using a Nipponbare (japonica) × Kasalath (indica) F_2 population. Singh et al (1996a) mapped centromeres on the molecular genetic map of rice and determined the correct orientation of linkage groups. More than 170 RFLP markers were assigned to specific chromosome arms through gene dosage analysis using secondary and telotrisomics and positions of centromeres were mapped on all 12 linkage groups.

Gene tagging and QTL mapping

The availability of comprehensive molecular maps in rice has opened new avenues to tag genes governing agronomic traits with molecular markers. This has led to major advances in marker-assisted selection and pyramiding of useful genes. McCouch et al (1991) and Yu et al (1991) were the first to tag genes for bacterial blight and blast resistance with molecular markers. Since then, many genes for disease and insect resistance and for other agronomic traits have been tagged with molecular markers (Table 2) (Mohan et al 1997, Khush and Brar 1998). Some examples include genes for blast resistance—*Pi1, Pi2(t), Pita, Pi5t, Pi7(t), Pi9(t), Pi10(t), Pi11(t),* and *Pib*; bacterial blight resistance—*Xa1, Xa2, Xa3, Xa4, xa5, Xa10, xa13,* and *Xa21*; brown planthopper resistance—*Bph1* and *Bph10*; gall midge resistance—*Gm1, Gm2, gm3, Gm4, Gm5,* and *Gm6(t)*; striped virus resistance (*Stvb1*); yellow mottle virus resistance; submergence tolerance (*sub1*); thermosensitive male sterility (*tms2, tms3*); photoperiod sensitivity (*Se1, Se3*); wide compatibility (*WC*); and fragrance.

Although several important characters are controlled by loci having a major effect on phenotype, several agronomically important traits such as yield, quality, and tolerance for abiotic stresses (drought, salinity, submergence, etc.) are quantitative in nature. The genes governing such traits, called polygenes or minor genes, also show Mendelian inheritance but are greatly influenced by the environment. The advent of molecular markers has made it possible to map such quantitative trait loci (QTLs). QTLs have been mapped for blast resistance, submergence tolerance, drought-related traits, and for several other agronomic traits such as days to heading, days to maturity, panicle length, spikelets per panicle, and grains per plant. Recently, Courtois et al (2000) identified QTLs for drought-related traits: 11 for leaf rolling, 10 for leaf drying, 11 for relative water content, and 10 for relative growth rate under stress.

Another major development involves the use of molecular markers to identify QTLs from unadapted germplasm or wild species that can enhance the grain yield of rice. Two yield-enhancing loci (*yld1, yld2*) located on chromosome 1 and 2 of *O. rufipogon* have been identified (Xiao et al 1996). In a similar experiment, Xiao et al (1998) identified 68 QTLs, of which 35 had beneficial alleles derived from a phenotypically inferior *O. rufipogon* parent.

Table 2. Some examples of mapping genes of agronomic importance with molecular markers in rice. Modified from Khush and Brar (1998).

Gene	Trait	Chromosome	Linked marker
Pi1	Blast resistance	11	Npb181
Pi2(t)	Blast resistance	6	RG64
Pi4	Blast resistance	12	RG869
Pita	Blast resistance	12	RZ397
Pi5(t)	Blast resistance	4	RG498, RG788
Pi6(t)	Blast resistance	12	RG869
Pi7(t)	Blast resistance	11	RG103
Pi9(t)	Blast resistance	6	RG16
Pi10(t)	Blast resistance	5	RRF6, RRH18
Pi11(t)	Blast resistance	8	BP127
Pib	Blast resistance	2	RZ123
Pi20	Blast resistance	12	XNbp88
Pik^{m}	Blast resistance	11	R1506
Xa1	Bacterial blight resistance	4	Npb235, Npb197
Xa2	Bacterial blight resistance	4	Npb235, Npb197
Xa3	Bacterial blight resistance	11	Npb181, Npb78
Xa4	Bacterial blight resistance	11	Npb181, Npb78
xa5	Bacterial blight resistance	5	RG556
Xa10	Bacterial blight resistance	11	$OP07_{2000}$
xa13	Bacterial blight resistance	8	RZ390, RG136
Xa21	Bacterial blight resistance	11	RG103
Xa22(t)	Bacterial blight resistance	–	RZ536
RTSV	Rice tungro spherical virus resistance	4	RZ262
RYMV	Rice yellow mottle virus resistance	12	RG341
Stvb1	Stripe virus resistance	12	XNpb220
Bph1	Brown planthopper resistance	12	XNpb248
Bph10	Brown planthopper resistance	12	RG457
ef	Early flowering	10	CD098
fgr	Fragrance	8	RG28, RM223
Wph1	Whitebacked planthopper resistance	7	–
WBPH	Whitebacked planthopper resistance	11	RG103
Gm1	Gall midge resistance	–	OPK7
Gm2	Gall midge resistance	4	BG329
gm3	Gall midge resistance	4	OPQ12
Gm4	Gall midge resistance	8	OPM12, RG476
Gm5	Gall midge resistance	12	OPB14
Gm6(t)	Gall midge resistance	4	RG214
Rf1	Fertility restoration	10	OSRRF
Rf2	Fertility restoration	1	CD0686/RZ58
Rf5	Fertility restoration	1	RG374
Rf3	Fertility restoration	1	RG532
S5	Wide compatibility	6	RG213
Se1	Photoperiod sensitivity	6	RG64
Se3	Photoperiod sensitivity	5	A19
Sdg(t)	Semidwarf	5	RZ182
Sd1	Semidwarf	1	RG109
tms1	Thermosensitive male sterility	8	–
tms3(t)	Thermosensitive male sterility	6	$OPAC3_{640}$
tms4(t)	Thermosensitive male sterility	2	RM27
pms1	Photoperiod-sensitive male sterility	7	RG477

continued on next page

Table 2 continued.

Gene	Trait	Chromosome	Linked marker
pms2	Photoperiod-sensitive male sterility	3	RG191
pms3	Photoperiod-sensitive male sterility	12	C751/RZ261
Sub1(t)	Submergence tolerance	9	RZ698
Salt	Salt tolerance	7	RG64
OSA3	Salt tolerance	12	RG457

Marker-assisted selection and gene pyramiding

The availability of comprehensive molecular linkage maps, tight linkage of target genes with molecular markers, and rapid development of polymerase chain reaction (PCR)-based DNA markers have facilitated the employment of marker-assisted selection (MAS) in rice breeding. In MAS, individuals carrying target genes are selected in a segregating population based on linked markers rather than on their phenotype. Thus, the population can be screened at any stage of growth and in various environments. MAS increases the efficiency of a breeding program by selecting for markers linked to target traits or QTLs. As mentioned, several genes for resistance to bacterial blight, blast, and gall midge have been tagged with molecular markers. Protocols for PCR-based MAS have been developed (Zheng et al 1995).

Yoshimura et al (1995) selected lines carrying *Xa4* + *xa5* and *Xa4* + *Xa10* using RFLP and RAPD markers linked to bacterial blight resistance genes. Lines carrying *Xa4* + *xa5* were more resistant to isolates of race 4 than were either of the parental lines. Huang et al (1997) used MAS to pyramid four genes for bacterial blight resistance—*Xa4*, *xa5*, *xa13*, and *Xa21*—into the background of IR24. Sanchez et al (2000) used sequence-tagged site (STS) markers to pyramid these three genes in an elite breeding line of new plant type rice. The pyramided lines having three or four genes in combination showed an increased and wider spectrum of resistance to bacterial blight than those having a single resistance gene. Such pyramided lines with different gene combinations are useful for developing varieties with durable resistance. Singh et al (2001) also used MAS to pyramid genes for bacterial blight resistance into a high-yielding indica rice cultivar, PR106, that is susceptible to bacterial blight. MAS has been useful in pyramiding recessive genes for resistance to bacterial blight such as *xa5* and *xa13* with a dominant gene *Xa21*, which confers resistance to many races and thus masks the resistance conferred by recessive genes.

MAS was also employed to pyramid genes for resistance to blast (Hittalmani et al 2000) and gall midge (Katiyar et al 2001). MAS also provides new opportunities to transfer and combine QTLs into agronomically desirable genotypes.

Physical maps

BAC and YAC libraries have facilitated the construction of physical maps of the rice genome. Wang et al (1995) and Umehara et al (1995) were the first to develop BAC

and YAC libraries, respectively. The BAC library consisted of 11,000 clones with an average DNA insert size of 125 kb. Twelve clones hybridized with three DNA markers closely linked to the *Xa21* locus. Yang et al (1997) developed a BAC library in the high-yielding and widely grown indica cultivar IR64. The library contained 18,432 clones with an average size of 107 kb. Several overlapping BAC clones were identified via colony hybridization with RFLP markers on chromosome 4. Zhang et al (1996) constructed two rice BAC libraries containing 22,000 clones with an average insert size of 130 and 150 kb, respectively. DNA markers and BAC-FISH (fluorescence *in situ* hybridization) technologies are becoming important for facilitating the generation and verification of the physical maps of rice.

YAC clones carrying several hundred kb to 1 Mb of rice genomic DNA have become important in physical mapping. The YAC library has an average insert size of 350 kb, ranging from 40 to more than 1,000 kb. Umehara et al (1996) constructed YAC libraries in japonica rice variety Nipponbare consisting of inserts covering about 6 times the genomic length. These two YAC libraries represented a *Not*1 fragment library and the *Eco*R1 partially digested fragment library. The 1,883 rice DNA markers were used to anchor YAC clones and ordered YAC libraries covering regions of all 12 rice chromosomes were constructed. YAC screening resulted in the identification of 5,701 YAC clones, of which 2,443 YACs have been arranged on distinct positions of the rice chromosomes.

Kurata et al (1997) described the arrangement of minimal overlapping YAC contigs and 188 YAC islands with multiple YACs on single marker positions were generated. The total coverage of all anchored YACs is estimated to be 215.8 Mb, corresponding to half the total length of the rice genome. YAC ordering showed that more than 30 DNA markers were distributed in the same order on both YAC clone arrays arranged on chromosomes 11 and 12 and that the two regions cover almost identical physical lengths on each chromosome. This is considered to be one of the longest duplications of the chromosome segment in the rice genome. Small segment duplications were also detected during the course of YAC selection with multiple-copy DNA sequences.

Recently, Saji et al (2001) constructed a physical map of rice with YAC clones covering 63% of the 12 rice chromosomes.

Map-based cloning of genes

The high-density genetic map coupled with the development of BAC and YAC libraries have been important discoveries leading to the isolation of rice genes (*Xa1, Xa21, Pib*). Song et al (1995) isolated the *Xa21* gene by positional cloning. The sequence of the predicted protein, which carries both a leucine-rich repeat (LRR) motif and a serine-threonine kinase (STK)-like domain, suggests a role in cell surface recognition of a pathogen ligand and subsequent activation of an intracellular defense response. The STK domain of *Xa21* is most similar to the tomato *Pto* resistance-gene product. Yoshimura et al (1998) cloned *Xa1* conferring resistance to bacterial blight. The deduced amino acid sequence of the *Xa1* gene product contains nucleotide-binding sites (NBS) and a new type of LRR. *Xa1* is a member of the NBS-LRR class of

plant resistance genes but is quite different from *Xa21*. It is interesting to note that *Xa1* gene expression, unlike other isolated resistance genes that show constitutive expression, is induced by pathogen infection and wound. Wang et al (1999) cloned the *Pib* blast resistance gene. The deduced amino acid sequence of the *Pib* gene product contains NBS and LRR. A duplication of the kinase 1a, 2, and 3a motifs of the NBS region was found in the N-terminal half of the *Pib* protein. Eight cysteine residues are clustered in the middle of the LRR, which has not been reported for other R genes. *Pib* gene expression was induced upon altered environmental conditions such as altered temperature and darkness. Sanchez et al (1999) identified BAC contigs flanking the *xa13* locus for bacterial blight resistance. Hayano-Saito (2000) identified overlapping BAC clones flanking *Stvb1*, the gene for stripe virus resistance.

Synteny relationships

The development of molecular genetic maps has been of great value in understanding the homoeologous relationships between the genomes of various crop plants. Ahn et al (1993) found extensive homoeologies in several regions of the genomes of wheat, rice, and maize. Kurata et al (1994a) analyzed synteny between rice and wheat and found that many wheat chromosomes contained homoeologous genes and genomic DNA fragments in an order similar to that found in rice. Comparative genome mapping in rice, maize, wheat, barley, sorghum, foxtail millet, and sugarcane into a single synthesis demonstrates that gene content and order are highly conserved at both the map and megabase level between different species within the grass family, but the amount and organization of repetitive sequences have diverged considerably (Devos and Gale 1997). Microsynteny analysis using the rice YAC clones of several hundred kilobases has also revealed remarkable similarities in marker orders between rice and barley or wheat.

Comparative genomics reveals that cereal genomes are composed of similar genomic building stocks (linkage blocks). The genomes of major cereals have been aligned by dissecting the individual chromosome into segments and rearranging these blocks into highly similar structures. Based on comparative mapping, gene location in one species can be used to predict the presence and location of orthologous loci in other species. Comparative mapping is accelerating map-based cloning of orthologous genes. The synteny relationships among cereals have resulted in the discovery of common genes such as the dwarf phenotype in maize (*D8*) and wheat *(Rht1*) based on genomic information derived from rice.

Alien introgression

Wild species of *Oryza* are an important reservoir of useful genes for resistance to major diseases and insects, tolerance for abiotic stresses, and a new source of cytoplasmic male sterility (CMS). A series of interspecific hybrids between rice and various wild species, monosomic alien addition lines (2n=25, MAAL), and introgression lines (2n=24) have been produced. The first two examples of the transfer of a useful

gene from wild species is the introgression of a gene for grassy stunt virus resistance from *O. nivara* into cultivated rice varieties (Khush et al 1977) and the transfer of cytoplasmic male sterility from wild rice, *O. sativa* f. *spontanea*, to develop CMS lines for commercial hybrid rice production (Lin and Yuan 1980). Since then, several useful genes for resistance to brown planthopper, whitebacked planthopper, bacterial blight, blast, and tungro have been introgressed from various wild species representing AA, BBCC, CC, CCDD, EE, and FF genomes into cultivated rice (Brar and Khush 1997). Some of the introgressed alien genes have been tagged with molecular markers. Molecular analysis reveals the introgression of small alien chromosome segments into the rice genome (Jena et al 1992, Brar et al 1996). FISH techniques have been employed to characterize parental genomes in interspecific hybrids and to detect homoeologous pairing and introgression of alien chromosome segments (Abbasi et al 1999). Wild species of *Oryza* are being explored for introgression of yield-enhancing loci/QTLs into rice (Xiao et al 1996, 1998, Khush and Brar, unpublished).

Structural genomics

Rice has become a model system for genomics research. Factors contributing to this situation include the comparatively smaller size of the rice genome, the synteny of its genome with those of other cereals, the availability of densely populated molecular maps containing more than 2,300 DNA markers, well-characterized YAC and BAC libraries, large-scale analysis of expressed sequence tags (ESTs), the vast amount of genetic resources (mutant markers, genetic stocks, wild species, mapping populations, introgression lines), and the comparative ease of transformation.

Major advances have been made in sequencing the rice genome under the International Rice Genome Sequencing Project (IRGSP) in Tsukuba, Japan, which began in 1998. This collaborative effort among 15 laboratories in 10 countries aims to produce publicly available sequence data for the complete rice genome. Each country has been assigned responsibility to sequence a specific rice chromosome (number 1—Japan, Korea; 2—UK; 3—USA; 4—China; 5—Taiwan; 6, 7, 8—Japan; 9—Thailand, Canada; 10—USA; 11—USA, India; 12—France, Brazil) and share the sequence data. As of April 2001, the public effort has generated 42 Mb of sequence data (10% of the genome). Two private sequencing efforts by Monsanto and Syngenta have resulted in working draft sequences for the complete rice genome. In August 2000, Monsanto transferred its rice genome draft sequence data to IRGSP. In IRGSP, a PAC library used as a vector to establish a genomics library for sequencing contained 70,000 clones of *Sau*3 AI fragments of DNA with an average insert size of 112 kb (Baba et al 2000). A total of 105,000 sequence-tagged connectors (STC) from Nipponbare BAC libraries (*Hin*dIII and *Eco*RI fragments) have been constructed by Clemson University and 55,000 STC from the IRGSP. At Tsukuba, 48 PAC on chromosome 1, 12 PAC on chromosome 6, two PAC on chromosome 2, one PAC on chromosome 3, and one PAC on chromosome 8 had been completely sequenced and annotated (Sasaki et al, this volume). A total of 102 PAC/BAC had been sequenced, covering 15.0 Mb of the rice genome. These sequences have been registered in the DNA data bank of Japan

and the information is freely available on the World Wide Web. An additional 100 PAC/BAC covering another 15 Mb have also been sequenced although several gap regions remain to be filled. These sequencing efforts by the IRGSP and other public and private organizations will provide sequences for the estimated 30,000 rice genes and for the intergenic DNA that plays an important but poorly understood role in gene expression, DNA replication, chromosome organization, recombination, specialization, and evolution.

Functional genomics

The availability of rice sequence data has ushered in the era of functional genomics. The *AC-DS* maize transposable elements, retrotransposons, miniature inverted repeat transposable elements (MITEs), and T-DNA insertions have provided a wealth of genetic resources for functional genomics. Some notable examples are T-DNA-tagged insertional mutants with 30,000 lines carrying 42,000 T-DNA inserts (An et al, this volume), Tos17 retrotransposon insertional mutants with about 30,000 lines carrying more than 250,000 Tos17s (Hirochika et al, this volume), and more than 40,000 deletion mutants produced by fast-neutron, gamma radiation, and chemical mutagenesis (Leung et al, this volume). These plant materials have provided the necessary link between structural or sequence data and gene function through forward and reverse genetics. Furthermore, a large number of ESTs from various cDNA libraries have been produced for microarray expression experiments. Maize transposon constructs have been used in the transformation of japonica and indica cultivars for knockout and gene detection insertion. A green fluorescent protein (GFP) excision assay is used to determine transposon excision in a variety of tissues. The functional genomic methodology is changing from forward genetics to reverse genetics. One of the major challenges is to determine the function of previously unknown rice genes revealed by sequencing. The second is to understand the functions of apparently redundant rice genes that may have different roles in different tissues or in response to different environments.

New high-throughput methods are being developed for expression analysis. Biochips are being used to follow changes in gene expression in response to abiotic stresses. Using gene chips or microarrays, the representative genes of rice can be analyzed on a glass slide and used in RNA hybridization to reveal gene expression patterns and identify pathways by association. A combination of mRNA and proteomics will precisely reveal the function of rice genes involved in the phenotypic expression of different agronomic traits.

References

Abbasi FM, Brar DS, Carpena AL, Fukui K, Khush GS. 1999. Detection of autosyndetic and allosyndetic pairing among A and E genomes of *Oryza* through genomic *in situ* hybridization. Rice Genet. Newsl. 16:24-25.

Aggarwal RK, Brar DS, Khush GS. 1997. Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. Mol. Gen. Genet. 254:1-12.

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS. 1999. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor. Appl. Genet. 98:1320-1328.

Ahn S, Anderson JA, Sorrells ME, Tanksley SD. 1993. Homoeologous relationships of rice, wheat and maize chromosomes. Mol. Gen. Genet. 241:483-490.

Ahn S, Tanksley SD. 1993. Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA 90:7980-7984.

Athwal DS, Pathak MD, Bacalangco EH, Pura CD. 1971. Genetics of resistance to brown planthopper and green leafhopper in *Oryza sativa* L. Crop Sci. 11:747-750.

Baba T, Katagiri S, Tanoue H, Tanaka R, Chiden Y, Saji S, Hamada M, Nakashima M, Okamoto M, Hayashi M, Yoshii S, Karasawa W, Honda M, Ichikawa Y, Arita K, Ikeno M, Ohta T, Umehara Y, Matsumoto T, de Jong PJ, Sasaki T. 2000. Construction and characterization of rice genomic libraries: PAC library of japonica variety Nipponbare and BAC library of indica variety Kasalath. Bull. NIAR 14:24-36.

Brar DS, Dalmacio R, Elloran R, Aggarwal R, Angeles R, Khush GS. 1996. Gene transfer and molecular characterization of introgression from wild *Oryza* species into rice. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Los Baños (Philippines): International Rice Research Institute. p 477-486.

Brar DS, Khush GS. 1997. Alien introgression in rice. Plant Mol. Biol. 35:35-47.

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD. 1994. Saturated molecular map of the rice genome based on interspecific backcross population. Genetics 138:1251-1274.

Chen L, Marmey P, Taylor NJ, Brizard J, Espinoza C, D'Cruz P, Huet H, Zhang S, de Kochko A, Beachy RN, Fauquet CM. 1998. Expression and inheritance of multiple transgenes in rice plants. Nat. Biotechnol. 16:1060-1064.

Cheng X, Sardana R, Kaplan H, Altosaar I. 1998. *Agrobacterium*-transformed rice plants expressing synthetic *cry1A(b)* and *cry1A(c)* genes are highly toxic to striped stemborer and yellow stemborer. Proc. Natl. Acad. Sci. USA 95:2767-2772.

Courtois B, McLaren G, Sinha PK, Prasad K, Yadav R, Shen L. 2000. Mapping QTLs associated with drought avoidance in upland rice. Mol. Breed. 6:55-66.

Devos KM, Gale MD. 1997. Comparative genetics in the grasses. Plant Mol. Biol. 35:3-15.

Ge S, Sang T, Lu B-R, Hong D-U. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA 96:14400-14405.

Goto F, Yoshihara T, Shigemoto N, Toki S, Takaiwa F. 1999. Iron fortification of rice seed by the soybean ferritin gene. Nat. Biotechnol. 17:282-286.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Liu S-Y, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

Hayano-Saito Y, Saito K, Nakamura S, Kawasaki S, Iwasaki M. 2000. Fine physical mapping of the rice stripe resistance gene locus, *Stvb-1*. Theor. Appl. Genet. 101:59-63.

Hayashi Y, Kyozuka J, Shimamoto K. 1988. Hybrids of rice (*Oryza sativa* L.) and wild *Oryza* species obtained by cell fusion. Mol. Gen. Genet. 214:6-10.

Heszky LE, Simon-Kiss I. 1992. 'DAMA,' the first plant variety of biotechnology origin in Hungary, registered in 1992. Hung. Agric. Res. 1:30-32.

Hittalmani S, Parco A, Mew TW, Zeigler RS, Huang N. 2000. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor. Appl. Genet. 100:1121-1128.

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95:313-320.

IRC (International Rice Commission). 1959. Genetic symbols for rice recommended by the International Rice Commission. Int. Rice Comm. Newsl. 8:1-7.

Iwata N, Omura T. 1975. Studies on the trisomics in rice plants (*Oryza sativa* L.). III. Relation between trisomics and genetic linkage groups. Jpn. J. Breed. 25:363-368.

Iwata N, Omura T. 1976. Studies on the trisomics in rice plants (*Oryza sativa* L.). IV. On the possibility of association of three linkage groups with one chromosome. Jpn. J. Genet. 51:135-137.

Iwata N, Omura T, Nakagahra M. 1970. Studies on the trisomics in rice plants (*Oryza sativa* L.). I. Morphological classification of trisomics. Jpn. J. Breed. 20:230-236.

Iwata N, Satoh H, Omura T. 1984. Relationship between the twelve chromosomes and the linkage groups. (Studies on the trisomics in rice plants *Oryza sativa* L.) V. Jpn. J. Breed. 34:314-321.

Jena KK, Khush GS, Kochert G. 1992. RFLP analysis of rice (*Oryza sativa* L.) introgression lines. Theor. Appl. Genet. 84:608-616.

Jodon NE. 1956. Revision of Figs. 3 to 6, "present status of rice genetics." J. Agric. Assoc. China 14(N.S.):69-73.

Kadam GS, Ramiah K. 1943. Symbolization of genes in rice. Indian J. Genet. Plant Breed. 3:7-27.

Katayama T. 1967. Cytogenetical studies on the genus *Oryza*: the F_1 hybrids of the crosses BBCC × CC, BBCC × a diploid strain of *O. punctata* and CC × a diploid strain of *O. punctata*. Proc. Jpn. Acad. Sci. 43:327-331.

Katiyar SK, Tan Y, Huang B, Chandel G, Xu Y, Bennett J. 2001. Molecular mapping of gene *Gm6*(t) which confers resistance against four biotypes of Asian rice gall midge in China. Theor. Appl. Genet. (In press.)

Khush GS. 1977. Disease and insect resistance in rice. Adv. Agron. 29:265-341.

Khush GS, Brar DS. 1998. The application of biotechnology to rice. In: Ives CL, Bedford BM, editors. Agricultural biotechnology in international development. Wallingford (UK): CAB International. p 97-121.

Khush GS, Brar DS. 2001. Rice. In: Chopra VL, editor. Evolution and adaptation of crops. I. Cereals. Oxford and IBH Publishing Co., New Delhi, Science Publishers, New Jersey, USA. (In press.)

Khush GS, Kinoshita T. 1991. Rice karyotype, marker genes and linkage groups. In: Khush GS, Toenniessen GH, editors. Rice biotechnology. Oxford (UK): CAB International. p 83-108.

Khush GS, Singh RJ, Sur SC, Librojo AL. 1984. Primary trisomics of rice: origin, morphology, cytology and use in linkage mapping. Genetics 107:141-163.

Kihara H. 1964. Need for standardization of genetic symbols and nomenclature in rice. In: Rice genetics and cytogenetics. Proceedings of the Symposium on Rice Genetics and Cytogenetics. Amsterdam (Netherlands): Elsevier. p 3-11.

Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P. 1999. Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hot spot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. Plant J. 17:591-601.

Kohli A, Leech M, Vain P, Laurie DA, Christou P. 1998. Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. Proc. Natl. Acad. Sci. USA 95:7203-7208.

Kurata N, Moore G, Nagamura Y, Footo T, Yano M, Minobe Y, Gale M. 1994a. Conservation of genome structure between rice and wheat. Bio/Technology 12:276-278.

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Nonna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y. 1994b. A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nat. Genet. 8:365-372.

Kurata N, Omura T, Iwata N. 1981. Studies on centromere, chromomere and nucleolus in pachytene nuclei of rice, *Oryza sativa* L., microsporocytes. Cytologia 46:791-800.

Kurata N, Umehara Y, Tanoue H, Sasaki T. 1997. Physical mapping of the rice genome with YAC clones. Plant Mol. Biol. 35:101-113.

Kuwada Y. 1910. A cytological study of *Oryza sativa* L. Shokubutsugaku Zasshi 24:267-281.

Li HW, Chen CC, Weng TW, Wuu KD. 1963. Cytogenetical studies of *Oryza sativa* L. and its related species. 4. Interspecific crosses involving *O. australiensis* with *O. sativa* and *O. minuta*. Bot. Bull. Acad. Sin. 4(2):65-74.

Li HW, Wang TS, Chen CC, Wand WH. 1961. Cytogenetical studies of *Oryza sativa* L. and its related species. I. Hybrids *O. paraguaiensis* Wedd. × *O. brachyantha* Chev. et Roehr, *O. paraguaiensis* Wedd × *O. australiensis* Domin. and *O. australiensis* Domin. × *O. alta* Swallen. Bot. Bull. Acad. Sin. 2(2):79-86.

Lin SC, Yuan LP. 1980. Hybrid rice breeding in China. In: Innovative approaches to rice breeding. Manila (Philippines): International Rice Research Institute. p 35-51.

Lu BR, Naredo EB, Juliano AB, Jackson MT. 1998. Taxonomic status of *Oryza glumaepatula* Steud. III. Assessment of genomic affinity among AA genome species from the New World, Asia and Australia. Gen. Res. Crop Evol. 45:215-223.

McCouch SR, Abenes ML, Angeles R, Khush GS, Tanksley SD. 1991. Molecular tagging of a recessive gene *xa-5* for resistance to bacterial blight of rice. Rice Genet. Newsl. 8:143-145.

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed. 3:87-103.

Morinaga T. 1939. Cytogenetics of rice. Bot. Zool. 7:197-201.

Morinaga T. 1942. On the rice plant. Rep. Agric. Assoc. Korea 1:201-208.

Morinaga T. 1943. Cytogenetical studies on *Oryza sativa* L. VI. The cytogenetics of F_1 hybrid of *O. minuta* Presl. and *O. latifolia* Desv. Jpn. J. Bot. 11:1-16.

Morinaga T, Kuriyama H. 1959. Genomic constitution of *Oryza officinalis*. Jpn. J. Breed. 8:259.

Nagamatsu T. 1942. Studies of the geographical differentiation of characters in the cultivated rice. 3. Proc. Crop Sci. Soc. Jpn. 14:132-134.

Nagao S. 1951. Genic analysis and linkage relationship of characters in rice. Adv. Genet. 4:181-212.
Nagao S, Takahashi M. 1963. Genetical studies on rice plant. XXVII. Trial construction of twelve linkage groups of Japanese rice. J. Fac. Agric. Hokkaido Univ. 53:72-130.
Nezu M, Katayama TC, Kihara H. 1960. Genetic study of the genus *Oryza*. I. Crossability and chromosomal affinity among 17 species. Seiken Ziho 11:1-11.
Niizeki H, Oono K. 1968. Induction of haploid rice plants from anther culture. Proc. Jpn. Acad. Sci. 44:554-557.
Ogura H, Shimamoto K. 1991. Field performance of protoplast derived rice plants and the release of a new variety. In: Bajaj YPS, editor. Biotechnology in agriculture and forestry. Vol. 14. Berlin (Germany): Springer-Verlag. p 269-282.
Parnell FR, Ayyangar GNR, Ramiah K. 1917. Inheritance of characters in rice. I. Mem. Dep. Agric. India Bot. Ser. 9:75-105.
Ramanujam S. 1937. Cytogenetical studies in the Oryzae. II. Cytogenetical behavior of an autotriploid in rice (*Oryza sativa* L.). J. Genet. 35:183-221.
Saito A, Yano M, Kishimoto N, Nakagahra M, Yoshimura A, Saito K, Kuhara S, Ukai Y, Kawase M, Nagamine T, Yoshimura S, Ideta O, Ohsawa R, Hayano Y, Iwata N, Sugiura M. 1991. Linkage map of restriction fragment length polymorphism loci in rice. Jpn. J. Breed. 41:665-670.
Saji S, Umehara Y, Antonio BA, Yamane H, Tanoue H, Baba T, Aoki H, Ishige N, Wu J, Koike K, Matsumoto T, Sasaki T. 2001. A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. Genome 44:32-37.
Sakaguchi S. 1967. Linkage studies on the resistance to bacterial leaf blight, *Xanthomonas oryzae* (Uyed et Ishiyama) Dowson, in rice. Bull. Natl. Inst. Agric. Sci. Ser. D 16:1-18.
Sanchez AC, Brar DS, Huang N, Li Z, Khush GS. 2000. Sequenced tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci. 40:792-797.
Sanchez AC, Ilag LL, Yang D, Brar DS, Ausubel F, Khush GS, Yano M, Sasaki T, Li Z, Huang N. 1999. Genetic and physical mapping of *xa13*, a recessive bacterial blight resistance gene in rice. Theor. Appl. Genet. 98:1022-1028.
Shastry SVS, Rangao Rao DR, Misra RN. 1960. Pachytene analysis in *Oryza*. I. Chromosome morphology in *Oryza sativa* L. Indian J. Genet. Plant Breed. 20:15-21.
Shinoda H, Toriyama K, Yunoki T, Ezuka A, Sakurai Y. 1971. Studies on the variety resistance of rice to blast. 6. Linkage relationships of blast resistance genes. Bull. Chugoku Agric. Exp. Stn. Ser. A 20:1-25.
Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS. 1996a. Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). Proc. Natl. Acad. Sci. USA 93:6163-6168.
Singh K, Multani DS, Khush GS. 1996b. Secondary trisomics and telotrisomics of rice: origin, characterization, and use in determining the orientation of chromosome map. Genetics 143:517-529.
Singh S, Sidhu JS, Huang N, Vikal Y, Li Z, Brar DS, Dhaliwal HS, Khush GS. 2001. Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor. Appl. Genet. 102:1011-1015.
Song W-H, Wang G-L, Chen L-L, Kim H-S, Pi L-Y, Holsten T, Gardner J, Wang B, Zhai W-X, Zhu L-H, Fauquet C, Ronald P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa-21*. Science 270:1804-1806.
Takano M, Egawa H, Ikeda J-E, Wakasa K. 1997. The structures of integration sites in transgenic rice. Plant J. 11:353-361.

Toriyama K, Arimoto Y, Uchimiya H, Hinata K. 1988. Transgenic rice plants after direct gene transfer into protoplasts. Bio/Technology 6:1072-1074.

Tu J, Zhang G, Datta K, Xu C, Ye Y, Zhang Z, Khush GS, Datta SK. 2000. Field performance of transgenic elite commercial hybrid rice expressing *Bacillus thuringiensis* δ-endotoxin. Nat. Biotechnol. 18:1101-1104.

Umehara Y, Inagaki A, Tanoue H, Yasukouchi Y, Nagamura Y, Saji S, Otsuki Y, Fujimura N, Kurata N, Minobe Y. 1995. Construction and characterization of rice YAC library for physical mapping. Mol. Breed. 1:79-89.

Umehara Y, Tanoue H, Kurata N, Ashikawa I, Minobe Y, Sasaki T. 1996. An ordered yeast artificial chromosome library covering one half of rice chromosome 6. Genome Res. 6:935-942.

Vain P, Worland B, Kohli A, Snape JW, Christou P, Allen GC, Thompson WF. 1999. Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J. 18:233-242.

Van der Stok JE. 1908. On the inheritance of grain color in rice. Teysmannia 65:5.

Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC. 1995. Construction of a rice bacterial artificial chromosome library and identification of clones linked to *Xa-21* disease resistance locus. Plant J. 7:525-533.

Wang Z-X, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T. 1999. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J. 19:55-64.

Wang ZY, Second G, Tanksley SD. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. 83:565-581.

Wuu KD, Jui Y, Lu KCL, Chou C, Li HW. 1963. Cytogenetical studies of *Oryza sativa* L. and its related species. 3. Two intersectional hybrids, *O. sativa* Linn × *O. brachyantha* A. Chev. et Roehr and *O. minuta* Presl. × *O. brachyantha* A. Chev. et Roehr. Bot. Bull. Acad. Sin. 4:51-59.

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L. 1996. Genes from wild rice improve yield. Nature 384:223-224.

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899-909.

Yamada Y, Yang Z-Q, Tang D-T. 1985. Regeneration of rice plants from protoplasts. Rice Genet. Newsl. 2:94-95.

Yamaguchi Y. 1921. Etudes d'heredité sur la couleur de glumes chez la riz. Bot. Mag. Tokyo 35:106-112.

Yamaguchi Y. 1926. Kreuzungsuntersuchungen an Reispflanzen. I. Genetische Analyse der Spelzenfarbe, und der Endospermbeschaffenheit bei einigen Sorten des Reises. Ber. Ohara Inst. Landwirtsch. Forsch. 3:1-126.

Yang D, Parco A, Nandi S, Subudhi P, Zhu Y, Wang G, Huang N. 1997. Construction of a bacterial artificial chromosome (BAC) library and identification of overlapping BAC clones with chromosome 4 specific RFLP markers in rice. Theor. Appl. Genet. 95:1147-1154.

Yang Z-Q, Shikanai T, Mori M, Yamada Y. 1989. Plant regeneration from cytoplasmic hybrids of rice (*Oryza sativa* L.). Theor. Appl. Genet. 77:305-310.

Ye X, He-Babili S, Kloti A, Zhang A, Lucca P, Beyer P, Potrykus I. 2000. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303-305.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T. 1998. Expression of *Xa1*, a bacterial blight resistance gene in rice, is induced by bacterial inoculation. Proc. Natl. Acad. Sci. USA 95:1663-1668.

Yoshimura S, Yoshimura A, Iwata N, McCouch SR, Abenes ML, Baraoidan MR, Mew TW, Nelson RJ. 1995. Tagging and combining bacterial blight resistance genes in rice using RAPD and RFLP markers. Mol. Breed. 1:375-387.

Yu ZH, Mackill DJ, Bonman JM, Tanksley SD. 1991. Tagging genes for blast resistance in rice via linkage to RFLP markers. Theor. Appl. Genet. 81:471-476.

Zhang H-B, Choi S, Woo S-S, Li Z, Wing RA. 1996. Construction of a bacterial artificial chromosome library and identification of clones linked to *Xa21* disease resistance locus. Plant J. 7:525-533.

Zhang HM, Yang H, Rech EL, Golds TJ, Davis AS, Mulligan BJ, Cocking EC, Davey MR. 1988. Transgenic rice plants produced by electroporation-mediated plasmid uptake into protoplasts. Plant Cell Rep. 7:379-384.

Zhang W, Wu R. 1988. Efficient regeneration of transgenic plants from rice protoplasts and correctly regulated expression of the foreign gene in the plants. Theor. Appl. Genet. 76:835-840.

Zheng K, Huang N, Bennett J, Khush GS. 1995. PCR-based marker-assisted selection in rice breeding. IRRI Discussion Paper Series No. 12. Manila (Philippines): International Rice Research Institute. 24 p.

Notes

Authors' address: Plant Breeding, Genetics, and Biochemistry Division, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Application of Mendelian genetics in rice breeding

J.N. Rutger and D.J. Mackill

The application of Mendelian genetics has clearly led to many breeding advances in rice as well as in other crops. In this chapter, we emphasize economically important traits for which segregation ratios can be distinguished without elaborate progeny testing or molecular markers. Four general groups of traits are reviewed: agronomic and physiological traits, grain quality, pest resistance, and resistance to abiotic stresses. The single most important trait has been semidwarfism, conferred by the *sd1* gene. Other important agronomic and physiological traits are photoperiod sensitivity, glabrous hulls, gold hull color, and purple leaf. Among grain quality characters, amylose content is the most important, ranging from waxy types with essentially no amylose to temperate japonica short- and medium-grain types with 16–18% amylose, to tropical japonica long grains with 21–24% amylose, and to many indica types with up to 28% amylose. Another significant grain quality trait is aroma, which often appears to be under simple genetic control but which is difficult to recover in high-yielding backgrounds. Breeding for pest resistance, including both diseases and insects, has been one of the most successful examples of the use of major genes in crops, and yet it is a recurring challenge. The most important disease example is blast resistance, which has been a focus of breeding efforts for decades. The exploitation of major genes for bacterial blight resistance has been more successful than for blast. Many successful cases of major gene resistance for brown planthopper, green leafhopper, and gall midge have been reported in the past three decades. In general, resistance to abiotic stresses has been polygenic. Exceptions have been low-temperature-induced chlorosis at the seedling stage and submergence tolerance.

King and Stansfield (1985) define Mendelian genetics as referring to the inheritance of chromosomal genes following the laws governing the transmission of chromosomes to subsequent generations. This is also called Mendelism. For the purposes of this chapter, this all-inclusive definition will be modified to emphasize economically

important traits for which segregation ratios can be distinguished without elaborate progeny testing or molecular markers.

Agronomic and physiological traits

Semidwarfism

Clearly, the most important gene in worldwide rice breeding is *sd1*. In fact, it can be argued that the *sd1* gene in rice and the *Rht* genes in wheat *are* the Green Revolution. An overwhelming preponderance of the world's semidwarf rice varieties share the *sd1* semidwarfing gene locus. Most tropical semidwarfs outside of China received *sd1* by descent from Deo Geo Woo Gen (DGWG), Taichung Native 1 (TN1), or I-Geo-Tze (IGT), hereafter referred to as the DGWG source, through the International Rice Research Institute (IRRI) breeding program. In China, many semidwarfs were derived from variety Ai-Jio-Nan-Te, which also has *sd1* (Oba et al 1990). Two mutant semidwarfs in japonica rice, Reimei in Japan and Calrose 76 in the United States, also have a semidwarfing gene at the *sd1* locus (Kikuchi et al 1985, Rutger 1992a). Considerable use has been made of the induced *sd1* mutants in japonica breeding programs in Japan and California. In the U.S. as a whole, greater use has been made of semidwarfism from DGWG derivatives. Although numerous additional semidwarfs have been identified, both through induced mutation and through naturally occurring sources, none has been as successful as *sd1*. Some of the nonallelic sources have obvious pleiotropic effects, such as small kernel size (Mackill and Rutger 1979) or other problems, whereas others have no obvious phenotypic problems but are just not as high-yielding as the tall check varieties (Rutger et al 1982, Rutger 1998). Rutger (1992a) has termed this phenomenon "the *sd1* mystique." A somewhat analogous situation exists for semidwarfism in wheat, in which the *Rht* semidwarfing loci predominate in world semidwarf varieties.

Photoperiod sensitivity

Another important agronomic trait in worldwide breeding efforts has been photoperiod sensitivity. The development of high-yielding varieties with broad adaptation for irrigated areas emphasized insensitivity to photoperiod. The control of photoperiod sensitivity by a dominant major gene (*Se1*) in traditional indica rice cultivars has long been known (Chandraratna 1955, Yokoo and Kikuchi 1978). There is evidence that different alleles of this locus may result in different flowering dates of photoperiod-sensitive cultivars (Yokoo and Kikuchi 1977). Poonyarit et al (1989) identified a locus designated *Se3* that interacted with the *Se1* locus. Varieties with the dominant allele at both loci flowered very late in the season, indicating that they had a very short critical photoperiod. In the 1980s, breeding for rainfed lowland areas emphasized photoperiod sensitivity as a valuable trait (Pushpavesa and Jackson 1979). Breeding lines combining the sensitive allele with semidwarfism (*sd1*) were developed for these areas. While it would appear that developing this combination of the sensitivity and semidwarfing genes would be straightforward, in fact, the number of sensitive lines that are truly productive under more favorable conditions has been very limited.

In temperate regions, breeders have used naturally occurring early varieties as sources of early maturity or relied upon transgressive segregation for earliness in crosses. Early maturity mutants have often been induced in otherwise well-adapted varieties (Carnahan et al 1989, Rutger et al 1987). The ease with which earliness is acquired is an indication that early maturity usually is rather simply inherited. For example, in California, an early maturity mutant was found to be controlled by a single partially dominant gene (McKenzie et al 1978).

Glabrous hull

The recessive gene for glabrous hull (*gl*) has been very important for the U.S. rice industry. Initially, the glabrous hull characteristic, which also confers glabrous leaves, became valued for hand harvesting and threshing of rice in the humid southern U.S. environment. Less itching occurred because the glabrous varieties lacked the small leaf barbs (trichomes) of pubescent varieties. Subsequently, it was observed that the bulk density of paddy rice of the glabrous varieties was greater because the glabrous kernels packed closer together. Also, less dust was created in the dehulling of paddy rice at mills, which are often in urban areas and are subject to clean-air acts. The first glabrous-hull varieties in the U.S. were Rexoro, selected in Louisiana in 1926 from the variety Marong-paroc introduced in 1911 from the Philippines, and Nira, selected in Louisiana in 1928 from an unnamed variety introduced in 1916 from the Philippines (Jones et al 1941). By the 1950s, most southern U.S. varieties were glabrous, having descended from either Rexoro or Nira. Glabrous varieties were developed in California at later dates. The first was CS-M3, which received its glabrous gene from Smooth No. 4 of unrecorded origin (Mastenbroek and Adair 1970). The second was CS-S4, which received its glabrous gene from Smooth No. 3, also of unrecorded origin (Adair et al 1972). Virtually all U.S. varieties today are glabrous, as are many Australian varieties that share U.S. parentage. To date, the glabrous gene has been "neutral" with no selective advantage or disadvantage. Most of the world's rice varieties are pubescent, except for tropical japonica varieties including upland and bulu types, which are characteristically glabrous.

Gold hull

Gold hull (*gh*) color was previously popular in U.S. varieties, beginning with the landmark variety Carolina Gold, which apparently arrived as a mixture in Carolina White, a line introduced into South Carolina in 1694 from Madagascar (Jones 1936). These two varieties became famous in the clean-rice markets of Europe and were grown in the U.S. for nearly two centuries (Jones 1936). Later varieties, including Rexoro and Delitus, also had gold hulls (Jones et al 1941). Rexoro was the gold-hull source for several varieties released in the southern U.S. between 1930 and 1965. The last widely grown gold-hull variety was Bluebelle (Bollich et al 1968). By the mid-1970s, the gold-hull varieties lost their popularity because of the increased use of parboiling, in which penetration of the gold color into milled rice produced undesired off-color rice. Allelic relationships of gold hull color in Carolina Gold and the Rexoro source have not been determined. It is well known that the gold hull color from the Rexoro source is con-

trolled by a single recessive gene, *gh*, and that mutants with gold hull color frequently recur in mutagenized populations, as in the M-101 gold hull reported by Rutger et al (1987) and in several other varieties (Rutger, unpublished). Since gold hull color seems to be another gene that is agronomically neutral, it is being incorporated into low-phytic-acid germplasm for phenotypic identification in the marketplace (Rutger, unpublished).

Purple leaf

Rice varieties possessing the purple leaf gene, *pl* (Kinoshita and Maekawa 1986), were long regarded as a curiosity. Breeders transferred this recessive allele into indica semi-dwarf varieties for use as an attractive border or "replacement" plots for their nurseries (e.g., IR1552). However, these purple-leaf varieties serve a practical purpose and are planted in India to facilitate removal of green weedy rice in direct-seeded areas. In Madhya Pradesh, some improved purple varieties have been released for this purpose.

Hybrid rice

Hybrid rice, grown extensively in China, resulted from the discovery and application of cytoplasmic male sterility and restorer systems. These mechanisms, called 3-line hybrid rice, have been reviewed by Yuan (1994) and Virmani (1994). Further hybrid rice refinements, including the 2-line system using environmentally sensitive genetic male sterility, have also been reviewed by these authors. Current applications of the 2-line system involve wide compatibility gene(s) for making indica-japonica hybrids (Ikehashi et al 1994).

Genetic male sterility

Genetic male sterility as an aid for making composite crosses and doing evolutionary breeding was first proposed by Suneson (1956) in barley. Sorghum and soybean researchers (Doggett 1972, Brim and Stuber 1973) have proposed the use of male sterility for further population improvement applications in their respective crops. Among the first to propose breeding applications of genetic male sterility in rice were Fujimaki et al (1977), Trees and Rutger (1978), and Singh and Ikehashi (1979). Studies on population improvement using genetic male sterility were reported by Choi et al (1988), Veillet et al (1996), and Watanesk and Mackill (1991). As a breeding tool, Rutger (1992b) used the genetic male sterile mutant M-101 ms #2 in a search for apomixis in rice. The male sterile line was used to produce 3,728 F_1 plants by natural crossing. The female parent was homozygous-recessive for three marker genes: *sd*, *gh*, and *ms*. The pollen parents were a collection of about 400 japonicas chosen for maturity similar to that of the male sterile line. All of the pollen parents were homozygous-dominant for the three marker genes. The F_1s were advanced to the F_2 and scrutinized for excess segregation of F_1-parent types, which would be an indicator of apomixis. Although some excess F_1-type segregation was observed, it was eventually concluded from F_3 and F_4 generations that the initial abnormal segregation was due to the small sample size, not apomixis. Dominant genetic male sterile lines, which are more efficient than recessive ones for population improvement (Sorrells and Fritz 1982), have been reported in recent years (Yan et al 1989, Zhu and Rutger 1999).

Elongated uppermost internode

A gene for elongated uppermost internode (*eui*), which produces a doubling in length of the uppermost internode, was described by Rutger and Carnahan (1981). The *eui* was proposed by the authors for use in the pollen parent in hybrid rice seed production. There would be obvious advantages to having a taller male parent in crossing fields and the resulting F_1, being recessive, is semidwarf like the female parent. Although patented, the *eui* stocks were distributed freely to requesting breeders. A use unforeseen by Rutger and Carnahan (1981) was that of putting the *eui* gene into the female parent in order to get the panicle above the leaf sheath (Zongtan and Zuhua 1989). This would then eliminate the need for gibberellin treatment of females in seed production fields. As far as is known, use of *eui* in either male or female seed parents has not been practiced on any significant scale.

Grain quality

Amylose content

Amylose content is the single most important criterion of cooking quality in rice. Waxy *(wx)* rice varieties have near zero amylose in endosperm starch. Such waxy or glutinous rice varieties are prized for pastries and ceremonial foods. Many waxy varieties are known in world collections, especially in short-grain temperate japonica rice. The waxy varieties are usually lower yielding than standard varieties, but this is offset by the higher prices for waxy grain. Typical was the experience with Mochi Gomi in the U.S., which was valued for its waxy grain even though it was tall and low-yielding. A clever use was made of induced mutation by Carnahan et al (1979) to develop variety Calmochi-201 in the otherwise well-adapted, high-yielding variety S6. Many waxy varieties are known in world collections, especially in short-grain temperate japonica rice and long-grain indica rice from Southeast Asia. RD6 is a popular rice variety from Thailand that is a waxy mutant of Khao Dawk Mali 105. Niaw San Patang used to be the most popular waxy variety in Thailand and is reportedly a waxy mutant of Leuang Yai 148.

Since the waxy rice markets are somewhat limited, greater emphasis has been given to inheritance of other amylose levels, including low amylose (16–18%) content as in temperate japonica short- and medium-grain rice, intermediate amylose (21–24%) as in the tropical japonica long grains of the southern U.S., and high amylose (26–28%) of older indicas. In crosses between low amylose and intermediate amylose parents, amylose content was usually controlled by one or two genes (Stansel 1965, Bollich and Webb 1973, McKenzie and Rutger 1983, Kumar and Khush 1987). In crosses with high-amylose parents, genes for high amylose were dominant over those for low or intermediate levels (Kumar and Khush 1987). In addition to the major genes, minor genes are usually present.

Additional genes affecting starch composition have been identified in japonica germplasm (Yano et al 1988) and transferred to an indica background by backcrossing (Kaushik and Khush 1991). As far as can be determined, these genes—for sugary, shrunken, floury, white core, amylose extender, and dull endosperm—have not yet been used in variety development.

Aroma

Aromatic rice varieties often receive a premium in many markets. The most prominent examples are basmati varieties grown in the Punjab region of Pakistan and India and the jasmine rice grown in Thailand. The first aromatic variety in the U.S. was Delitus, selected in 1911 in Louisiana from the Bertone variety introduced in 1904 from France (Jones et al 1941). The aroma was transferred to variety Della (Jodon and Sonnier 1973), which was grown on a modest area for some 20 years. Aroma in Della was controlled by a single recessive gene (Berner and Hoff 1986) that was subsequently mapped to chromosome 8 by Ahn et al (1992). Pinson (1994) investigated several aromatic varieties, finding that Jasmine 85 and PI 457917 (a semidwarf mutant from Basmati 370) both possess a single recessive gene for aroma and both are allelic with each other and with the aroma genes in A-301 and Della-X_2 (a short-stature mutant of Della). Two additional varieties, Amber and Dragon Eyeball 100, both possess two genes for aroma, one of which is allelic to the aroma gene in the other four varieties (Pinson 1994). Several U.S. breeding programs have released aromatic varieties in recent years, but various ethnic groups in the U.S. prefer imported jasmine and basmati rice. More than 12% of rice eaten in the U.S. is imported rice (Childs 1999), mostly jasmine and basmati.

Pest resistance

The breeding of disease and insect resistance into rice varieties has been one of the most successful examples of the use of major genes in crop species. These genes have also become major targets for gene tagging through molecular markers, marker-assisted selection and pyramiding, and positional cloning programs. The development of disease- and insect-resistant rice has been reviewed elsewhere (Khush 1984, Khush and Virmani 1985). A brief summary of the more important resistances is provided here.

Diseases

Blast. Rice blast (caused by *Magnaporthe grisea*) is the most widespread rice disease and has been a focus of breeding efforts for decades. The elegant work of Kiyosawa (1981) established the basis for genetic studies of blast resistance genes in japonica rice using Japanese blast strains. Work in the tropics was hampered by the diversity of the pathogen races, the apparently high number of resistance genes in indica rice, and inoculation difficulties (Bonman and Mackill 1988). Development of near-isogenic lines improved the situation (Mackill and Bonman 1992). Despite the identification of many major genes for blast resistance, use of this information has only been common in temperate areas such as the United States. Katy, for example, which was initially released in 1989 (Moldenhauer et al 1990), was the only variety with resistance to all common blast races in the southern U.S. It was initially believed that there was a single gene, *Pi-ta*^{2}, or a tightly linked group of genes from Tetep controlling this resistance (Moldenhauer et al 1992). Further research has shown that the resistance in Katy is due to a tightly linked cluster of at least seven genes (Chao et al 1999). In

recent years, Katy has been used as a source of blast resistance in the southern U.S. In blast-prone areas, major-gene resistance has been ineffective due to the rapid development of new races of the pathogen. In these areas, partial resistance (along with chemical control) has been used (Ezuka 1979). With a better understanding of the relationship between resistance genes and pathogen lineages, gene pyramiding using molecular markers is now a viable approach.

Bacterial blight. The exploitation of major genes for bacterial blight (caused by *Xanthomonas campestris*) resistance has been more successful than that for blast (Mew 1987, Khush et al 1989). A large number of genes for resistance have been identified and several have been transferred to high-yielding varieties (Ogawa and Khush 1989). *Xa4* has been a particularly successful gene, showing some residual resistance even after compatible races of the pathogen became common. In the presence of virulent pathogen races, genes may differ in their level of resistance (Ona et al 1998). Again, gene pyramiding is currently being pursued in an effort to develop more durable and broader resistance (Huang et al 1997), although varietal mixtures have also been advocated (Ahmed et al 1997).

Insects

Brown planthopper (BPH). This insect devastated high-yielding rice crops in the 1970s until resistant varieties were developed. Variety IR26, possessing the resistance gene *Bph1*, was developed at IRRI (Khush 1989). When this resistance broke down in 1976, the recessive *bph2* was used. In general, BPH resistance was believed to be successful; however, some workers contend that this pest is caused by mismanagement of pesticides and continuous rice planting, and that the biotypes are not fixed but highly fluid. Most likely, resistance has had some success in reducing losses from this pest (Gallagher et al 1994).

Green leafhopper (GLH). This insect rarely causes direct damage but is more a problem as the vector of rice tungro viruses. A large number of resistance genes have been identified (Ghani and Khush 1988). Resistance to the vector also confers resistance to the viruses; however, new GLH biotypes can develop and require the incorporation of new resistance genes.

Gall midge. Several genes conferring resistance to gall midge have been identified and transferred into high-yielding cultivars (Tomar and Prasad 1992, Pani and Sahu 2000). Resistance has been very effective in reducing damage from the disease. Although several biotypes exist, resistance has been relatively durable.

Resistance to abiotic stresses

In general, tolerance of or resistance to abiotic stresses is a polygenic trait. One exception is low-temperature-induced chlorosis or wilting at the seedling stage. Several genes controlling this trait were identified and single-gene segregation ratios are usually observed in indica × japonica crosses (Kwak et al 1984, Nagamine 1991). However, it appears that this gene has not been exploited in breeding for cold tolerance.

Submergence tolerance is an interesting example in which most, but not all, of the resistance is controlled by a major gene. In many genetic studies, this trait appears quantitative, but more careful testing reveals the effect of the major gene, designated *Sub1* (Mishra et al 1996, Xu and Mackill 1996). This gene has been transferred from the traditional cultivar FR13A into high-yielding types (Mackill et al 1993). These are being extensively evaluated in rainfed lowland areas where short-term flooding is a problem.

Conclusions

Spectacular advances have been achieved in rice breeding through manipulations of rather simply inherited characters. These include semidwarfism, the cornerstone of the Green Revolution; photoperiod insensitivity and sensitivity for adjusting maturity ranges; kernel starch variation for modifying cooking quality; resistance to ever-changing spectra of diseases and insects; and stabilization of crop performance through resistance to abiotic stresses. Although many researchers would like to find similar big-ticket items, the easy achievements are behind us. Fortunately, ever-more-sophisticated genetic techniques are becoming available for rice improvement, as reviewed by other chapters in this book.

References

Adair CR, Thysell JR, Mastenbroek JJ, Carnahan HL. 1972. Registration of CS-S4 rice. Crop Sci. 12:256-257.

Ahmed HU, Finckh MR, Alfonso RF, Mundt CC. 1997. Epidemiological effect of gene deployment strategies on bacterial blight of rice. Phytopathology 87:66-70.

Ahn SN, Bollich CN, Tanksley SD. 1992. RFLP tagging of a gene for aroma in rice. Theor. Appl. Genet. 84:825-828.

Berner DK, Hoff BJ. 1986. Inheritance of scent in American long grain rice. Crop Sci. 26:876-878.

Bollich CN, Scott JE, Webb BD, Atkins JG. 1968. Registration of Bluebelle rice. Crop Sci. 8:399-400.

Bollich CN, Webb BD. 1973. Inheritance of amylose in two hybrid populations of rice. Cereal Chem. 50:631-636.

Bonman JM, Mackill DJ. 1988. Durable resistance to rice blast disease. Oryza 25:103-110.

Brim CA, Stuber CW. 1973. Application of genetic male sterility to recurrent selection schemes in soybeans. Crop Sci. 13:528-530.

Carnahan HL, Johnson CW, Tseng ST, Brandon DM. 1979. Registration of "Calmochi-201" rice. Crop Sci. 19:746.

Carnahan HL, Johnson CW, Tseng ST, Oster JJ, Hill JE, Brandon DM. 1989. Registration of "M-203" rice. Crop Sci. 29:1089-1090.

Chandraratna MF. 1955. Genetics of photoperiod sensitivity in rice. J. Genet. 53:215-223.

Chao CT, Moldenhauer KAK, Ellingboe AH. 1999. Genetic analysis of resistance/susceptibility in individual F_3 families of rice against strains of *Magnaporthe grisea* containing different genes for avirulence. Euphytica 109:183-190.

Childs N. 1999. Rice situation and outlook yearbook. Washington, D.C. (USA): Economic Research Service, U.S. Department of Agriculture, RCS-1999. 80 p.

Choi HC, Cho SY, Park RK. 1988. Studies on population improvement breeding in rice using genic male-sterile factor. II. Variation of some quantitative characters in process of population improvement by recurrent random mating and new germplasm introduction using male-sterile factor. Korean J. Breed. 20:89-103. (In Korean with English summary.)

Doggett H. 1972. Recurrent selection in sorghum populations. Heredity 28:9-29.

Ezuka A. 1979. Breeding for and genetics of blast resistance in Japan. In: Proceedings of the rice blast workshop. Los Baños (Philippines): International Rice Research Institute. p 27-48.

Fujimaki H, Hiraiwa S, Kushibuchi K, Tanaka S. 1977. Artificially induced male-sterile mutants and their usages in rice breeding. Jpn. J. Breed. 27:70-77.

Gallagher KD, Kenmore PE, Sogawa K. 1994. Judicial use of insecticides deters planthopper outbreaks and extends the life of resistant varieties in Southeast Asian rice. In: Denno RF, Perfect TJ, editors. Planthoppers: their ecology and management. New York, NY (USA): Routledge Chapman & Hall Inc. p 599-614.

Ghani MU, Khush GS. 1988. A new gene for resistance to green leafhopper *Nephotettix virescens* (Distant) in rice. J. Genet. 67:151-159.

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95:313-320.

Ikehashi H, Zou JS, Huhn PM, Maruyama K. 1994. Wide compatibility gene(s) and indica-japonica heterosis in rice for temperate countries. In: Virmani SS, editor. Hybrid rice technology: new developments and future prospects. Selected papers from the International Rice Research Conference. Manila (Philippines): International Rice Research Institute. p 21-31.

Jodon NE, Sonnier EA. 1973. Registration of Della rice. Crop Sci. 13:773.

Jones JW. 1936. Improvement in rice. In: USDA Yearbook. Washington, D.C. (USA): United States Department of Agriculture. p 415-454.

Jones JW, Jenkins JM, Nelson M, Carter LC, Adair CR, Wyche RH, Beachell HM, Davis LL, King BM. 1941. Rice varieties and their comparative yields in the United States. Circular No. 612. Washington, D.C. (USA):. United States Department of Agriculture. 34 p.

Kaushik RP, Khush GS. 1991. Genetic analysis of endosperm mutants in rice *Oryza sativa* L. Theor. Appl. Genet. 83:146-152.

Khush GS. 1984. Breeding rice for resistance to insects. Prot. Ecol. 7:147-165.

Khush GS. 1989. Multiple disease and insect resistance for increased yield stability in rice. In: Progress in irrigated rice research. Los Baños (Philippines): International Rice Research Institute. p 79-92.

Khush GS, Mackill DJ, Sidhu GS. 1989. Breeding for resistance to bacterial blight. In: Bacterial blight of rice. Los Baños (Philippines): International Rice Research Institute. p 207-211.

Khush GS, Virmani SS. 1985. Breeding rice for disease resistance. In: Russel GE, editor. Progress in plant breeding. 1. London (UK): Butterworths. p 239-279.

Kikuchi F, Itakura N, Ikehashi H, Yokoo M, Nakane A, Maruyama K. 1985. Genetic analysis of semidwarfism in high-yielding rice varieties in Japan. Bull. Natl. Inst. Agric. Sci. D36:125-145.

King RC, Stansfield WD. 1985. A dictionary of genetics. New York (USA): Oxford University Press. 480 p.

Kinoshita T, Maekawa M. 1986. Genetical studies on rice plant. XCIV. Inheritance of purple leaf color found in indica rice. J. Fac. Agric. Hokkaido Univ. 62:453-466.
Kiyosawa S. 1981. Gene analysis for blast resistance. Oryza 18:196-203.
Kumar I, Khush GS. 1987. Genetic analysis of different amylose levels in rice. Crop Sci. 27:1167-1172.
Kwak TS, Vergara BS, Nanda JS, Coffman WR. 1984. Inheritance of seedling cold tolerance in rice. SABRAO J. 16:83-86.
Mackill DJ, Rutger JN. 1979. Inheritance of induced-mutant semidwarfing genes in rice. J. Hered. 70:335-341.
Mackill DJ, Amante MM, Vergara BS, Sarkarung S. 1993. Improved semidwarf rice lines with tolerance to submergence of seedlings. Crop Sci. 33:749-753.
Mackill DJ, Bonman JM. 1992. Inheritance of blast resistance in near-isogenic lines of rice. Phytopathology 82:746-749.
Mastenbroek JJ, Adair CR. 1970. Registration of "CS-M3" rice. Crop Sci. 10:728.
McKenzie KS, Board JE, Foster KW, Rutger JN. 1978. Inheritance of heading date of an induced mutant for early maturity in rice (*Oryza sativa* L.). SABRAO J. 10:96-102.
McKenzie KS, Rutger JN. 1983. Genetic analysis of amylose content, alkali spreading score, and grain dimensions in rice. Crop Sci. 23:306-313.
Mew TW. 1987. Current status and future prospects of research on bacterial blight of rice. Annu. Rev. Phytopathol. 25:359-382.
Mishra SB, Senadhira D, Manigbas NL. 1996. Genetics of submergence tolerance in rice (*Oryza sativa* L.). Field Crops Res. 46:177-181.
Moldenhauer KAK, Bastawisi AO, Lee FN. 1992. Inheritance of resistance in rice to races IB-49 and IC-17 of *Pyricularia grisea* rice blast. Crop Sci. 32:584-588.
Moldenhauer KAK, Lee FN, Norman RJ, Helms RS, Wells RB, Dilday RH, Rohman PC, Marchetti MA. 1990. Registration of "Katy" rice. Crop Sci. 30:747-748.
Nagamine T. 1991. Genic control of tolerance to chilling injury at seedling stage in rice. Jpn. J. Breed. 41:35-40.
Oba S, Kikuchi F, Maruyama K. 1990. Genetic analysis of semidwarfness and grain shattering of Chinese rice variety "Ai-Jio-Nan-Te". Jpn. J. Breed. 40:13-20.
Ogawa T, Khush GS. 1989. Major genes for resistance to bacterial blight in rice. In: Bacterial blight of rice. Los Baños (Philippines): International Rice Research Institute. p 177-192.
Ona I, Cruz CMV, Nelson RJ, Leach JE, Mew TW. 1998. Epidemic development of bacterial blight on rice carrying resistance genes *Xa-4*, *Xa-7*, and *Xa-10*. Plant Dis. 82:1337-1340.
Pani J, Sahu SC. 2000. Inheritance of resistance against biotype 2 of the Asian rice gall midge, *Orseolia oryzae*. Entomol. Exp. Appl. 95:15-19.
Pinson SRM. 1994. Inheritance of aroma in six rice cultivars. Crop Sci. 34:1151-1157.
Poonyarit M, Mackill DJ, Vergara BS. 1989. Genetics of photoperiod sensitivity and critical daylength in rice. Crop Sci. 29:647-652.
Pushpavesa S, Jackson BR. 1979. Photoperiod sensitivity in rainfed rice. In: Rainfed lowland rice: selected papers from the 1978 International Rice Research Conference. Manila (Philippines): International Rice Research Institute. p 139-147.
Rutger JN. 1992a. Impact of mutation breeding in rice: a review. Mutation Breeding Review No. 8. Vienna (Austria): FAO/IAEA. p 1-23.
Rutger JN. 1992b. Searching for apomixis in rice. In: Proceedings of the Apomixis Workshop, USDA-ARS-104, 11-12 February 1992, Atlanta, Georgia. p 36-39.
Rutger JN. 1998. Rice genetics and germplasm enhancement. In: Norman RJ, Johnston TH, editors. Rice research studies. Arkansas Exp. Stn. Res. Ser. 460:12-15.

Rutger JN, Carnahan HL. 1981. A fourth genetic element for facilitating hybrid seed production in cereals: a recessive tall in rice. Crop Sci. 21:373-376.
Rutger JN, Carnahan HL, Johnson CW. 1982. Registration of 10 germplasm lines of rice. Crop Sci. 22:164-165.
Rutger JN, Figoni RA, Webster RK, Oster JJ, McKenzie KS. 1987. Registration of early maturing, marker gene, and stem rot resistant germplasm lines of rice. Crop Sci. 27:1319-1320.
Singh RJ, Ikehashi H. 1979. Induction of monogenic recessive male sterility in IR36 by ethyleneimine treatment. Int. Rice Res. Newsl. 4(3):3-4.
Sorrells ME, Fritz SE. 1982. Application of a dominant male-sterile allele to the improvement of self-pollinated crops. Crop Sci. 22:1033-1035.
Stansel JW. 1965. Influence of heredity and environment on endosperm characteristics of rice (*Oryza sativa* L.). PhD thesis. Purdue University. University Microfilms, Ann Arbor, Michigan. (Diss. Abstr. 27:48B.)
Suneson CA. 1956. An evolutionary plant breeding method. Agron. J. 48:188-191.
Tomar JB, Prasad SC. 1992. Genetic analysis of resistance to gall midge (*Orseolia oryzae* Wood Mason) in rice. Plant Breed. 109:159-167.
Trees SC, Rutger JN. 1978. Inheritance of four genetic male steriles in rice. J. Hered. 69:270-272.
Veillet S, Filippi MC, Gallais A. 1996. Combined genetic analysis of partial blast resistance in an upland rice population and recurrent selection for line and hybrid values. Theor. Appl. Genet. 92:644-653.
Virmani SS. 1994. Prospects of hybrid rice in the tropics and subtropics. In: Virmani SS, editor. Hybrid rice technology: new developments and future prospects. Selected papers from the International Rice Research Conference. Manila (Philippines): International Rice Research Institute. p 7-19.
Watanesk O, Mackill DJ. 1991. Evaluating S1 family recurrent selection in a rice population. Int. Rice Res. Newsl. 16(3):8-9.
Xu K, Mackill DJ. 1996. A major locus for submergence tolerance mapped on rice chromosome 9. Mol. Breed. 2:219-224.
Yan L, Zhang J, Zhu C, Ou Y, Li J, Cai Y. 1989. The preliminary evaluation of a dominant male-sterile gene in rice. Acta Agron. Sin. 15:174-180. (In Chinese, English summary.)
Yano M, Okuno K, Satoh H, Omura T. 1988. Chromosomal location of genes conditioning low amylose content of endosperm starches in rice, *Oryza sativa* L. Theor. Appl. Genet. 76:183-189.
Yokoo M, Kikuchi F. 1977. Multiple allelism of the locus controlling heading time of rice, detected using the close linkage with blast-resistance. Jpn. J. Breed. 27:123-130. (In Japanese with English summary.)
Yokoo M, Kikuchi F. 1978. Genic analysis for heading time in rice by the aid of the linkage with blast resistance. Jpn. J. Breed. 28:49-55. (In Japanese with English summary.)
Yuan LP. 1994. Increasing yield potential in rice by exploitation of heterosis. In: Virmani SS, editor. Hybrid rice technology: new developments and future prospects. Selected papers from the International Rice Research Conference. Manila (Philippines): International Rice Research Institute. p 1-6.
Zhu X, Rutger JN. 1999. Inheritance of induced dominant and recessive genetic male-sterile mutants in rice (*Oryza sativa* L.). SABRAO J. Breed. Genet. 31(1):17-22.
Zongtan S, Zuhua H. 1989. Transfer of *eui* gene to WA-MS line Zhen-Shan 91A (*Oryza* ssp. *indica*) and eliminating its panicle enclosure. Int. Rice Res. Newsl. 14(4):8-9.

Notes

Authors' addresses: J.N. Rutger, Dale Bumpers National Rice Research Center, U.S. Department of Agriculture-Agricultural Research Service, P.O. Box 287, Stuttgart, Arkansas 72160, USA; D.J. Mackill, U.S. Department of Agriculture-Agricultural Research Service, and Department of Agronomy and Range Science, University of California, Davis, California 95616, USA.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

The Rockefeller Foundation's International Program on Rice Biotechnology

J.C. O'Toole, G.H. Toenniessen, T. Murashige, R.R. Harris, and R.W. Herdt

The Rockefeller Foundation's design of a long-term program on rice biotechnology was the product of a 2-year intensive survey and analysis of the genetic prospects for the world's major food crops conducted in the early 1980s. In late 1984, the Foundation's Board of Trustees approved a strategy for a 10–15-year program. That program was highly speculative and indicated substantial risk with regard to the status at that time of cereal plant molecular biology and rice in particular. During the first 5–7 years, projects supported by the Foundation laid the scientific basis for "rice biotechnology" as we know it today. Early successes were the first DNA molecular marker map of rice, the regeneration and transformation of rice, the use of rice pest genomic information to unravel age-old riddles of host-plant resistance, and numerous other discoveries that changed the way rice geneticists viewed breeding objectives such as insect resistance, abiotic stress tolerance, and hybrid rice. These discoveries culminated in the revelation of rice's pivotal genomic position in the evolution of cereal species. Over the ensuing 7–8 years, the program shifted its focus to the transfer of the resulting biotechnologies to institutions in rice-producing and -consuming countries. This task required the strengthening of both physical and human resources in cooperation with national and international rice research systems in Asia, Africa, and Latin America. The Foundation's program management sought to support further technology generation and application while promoting the program's greatest asset, *international collaborative research-cum-training*. This "win-win" component of the program linking fledgling national rice biotechnology efforts directly to advanced research institutes in the United States, Europe, Japan, and Australia became the hallmark of the Foundation's management strategy. During the program's 17-year lifetime, more than 400 (primarily Asian) rice scientists were trained in this manner. The successful linkage of research in cutting-edge biotechnology with the training of rice scientists often produced long-term collaborative relationships that outgrew dependence on Foundation support and continue today (such as the IRRI-managed Asian Rice Biotechnology Network). Some of these successes were undoubtedly a consequence of the basic research progress in rice plant molecular genomics, which brought greater financial support for rice-centered research as rice became the "model cereal" for genomic research, rivaling even *Arabidopsis*.

The Rockefeller Foundation has a long, complex, and rich history in promoting agricultural development throughout the developing world. The Foundation began its major field-based program in Mexico in the 1940s, which led to the series of technologies, insights, and processes collectively known as the "Green Revolution." During the 1950s, success in Mexico led the Foundation to establish similar country programs in Colombia, Chile, and India. The 1960s saw the Foundation establish, jointly with the Ford Foundation, four international agricultural research centers. In 1971, the Foundation helped establish a consortium of donors to support the international centers, the Consultative Group on International Agricultural Research (CGIAR). Today, 16 CGIAR-supported centers throughout the world have a total annual budget of about $350 million. The resulting growth in agricultural production and farm incomes, together with educational, health, policy, and environmental improvements, helped increase food availability and entitlements and enabled hundreds of millions to escape hunger over the past 50 years.

During the 1970s, many changes occurred in the direction and staffing of the Foundation's Agricultural Sciences Division. In 1982, an external review team intensively examined the Foundation's agricultural program (Kearl et al 1982). Prominent among the suggested future activities was the application of molecular biology to plant breeding. Over the next two years, Foundation officers consulted experts and assessed the relative status and merits of a program focusing on a few or a single crop species. Finally, the decision was made in late 1984 to implement a comprehensive program on rice, ranging from fundamental research through to the application of new molecular-based techniques in rice breeding (Toenniessen and Herdt 1988). The Foundation Board of Trustees approved the program in December 1984 and was aware from the outset of the high risk involved and probable 10–15-year time frame to accomplish the objectives. (At that time, the rice genome was relatively unknown and even its size was considerably overestimated, no DNA molecular markers/maps were available, no cereal had been regenerated from a protoplast, and hence there was no experimental evidence to support the proposal that transformation of rice with novel genes would ultimately become a tool for rice genetic improvement.)

This chapter sets out to chronicle the unique nature of the program's origin, guiding principles, and salient achievements in both scientific progress and capacity building in rice-consuming countries. In addition, we will illustrate the complex nature of an evolving program structure and management experience that fostered international collaborative research, training, and capacity building, focused on rice-consuming countries.

Evolution of program strategy and implementation

In 1984, the lowly position of rice (*Oryza sativa* L.) in the world of plant molecular biology represented a dramatic challenge to Foundation officers charged with generating a rice biotechnology knowledge base. Through a series of strategically placed grants, some of the world's premier laboratories were invited to participate in the program. The early output from these labs and others spurred on by the attention and promise of rice biotechnology was impressive: rice became the first cereal to be re-

generated from a protoplast in 1986-88, a comprehensive molecular (restriction fragment length polymorphism, RFLP) map was achieved by 1988, and the first experimental transformation was accomplished in 1988-90. These rapid developments were heartening and during that period the International Program on Rice Biotechnology (IPRB) goals were modified to include not only the discovery of scientific fundamentals and their transfer to the CGIAR-supported centers, but a more comprehensive and ambitious set of objectives was adopted related to the transfer of the new technology to rice researchers in rice-consuming countries.

Following a review of Foundation-wide development strategies, the IPRB goals were modified into four primary objectives:

1. To assure that the scientific tools of biotechnology were developed for tropical rice;
2. To create sufficient biotechnology capacity in rice-dependent countries to meet current and future challenges to rice production;
3. To better understand the consequences of agricultural technological change in Asia, in part to help in setting priorities for biotechnological applications; and
4. To apply this knowledge and capacity to the production of improved rice varieties and other materials that will enable farmers to produce more abundant supplies of nutritious food while causing less environmental damage.

Setting rice biotechnology research priorities

Setting research priorities for the program using a socioeconomic approach that balanced opportunities for rice productivity gains with costs of research was one of the unique aspects of the program. Herdt and Rieley (1987) pioneered the use of crop-loss estimates, weighted by equity considerations, and a global approach to provide the Foundation's rice biotechnology program with a set of the top-20 priority traits. The use of this mechanism to focus funding on the genetic solution of high-priority traits distinguishes the program and led to very cost-effective decision making by program management staff. The global study of 1986 was followed by several studies, of narrower geographic scale, intended to provide similar guidelines for resource allocation to national agricultural research systems at the country or regional level (eastern India—Widawsky and O'Toole 1990; China—Lin and Shen 1993; Nepal—Upadhyaya et al 1993; for others, see Evenson et al 1996). In addition, in light of changing circumstances and developing science, Foundation economists continued to refine both the methodology and the results used to direct investment in rice research (Herdt 1986, 1987, 1991). The nearly decade-long effort to prioritize rice research targets, with a primary purpose to guide the Foundation's IPRB, culminated in the publication of *Rice Research in Asia: Progress and Priorities* (Evenson et al 1996).

All research and capacity-building programs have finite budgets and the Foundation's rice biotechnology program was no exception. The continued use of updated output from "priority-setting" research was a salient component of the IPRB management decision making. In this way, national and international research targets were clearly defined and, when coupled with investments in human resource development (see below), represented a well-integrated and very cost-effective program management structure.

Evolution and implementation of the strategic plan

Evolution and implementation of the strategic plan were based on the Foundation's private mode of operation. This allowed the program to experiment with both scientific and management activities and to be responsive to the needs of grantees while also being opportunistic when appropriate. Figure 1 provides a greatly simplified conceptual model of the IPRB's operational mode.

The program was first of all ***research-driven***. All participants in the program entered only by this route. Few activities were supported that were not directly or indirectly associated with promoting the overall research goals. As illustrated in Figure 1, the benefits of participating in the IPRB were attractive to scientists in both the developed and developing countries. Early in the program's lifetime, it was determined to bring all participants together in periodic international meetings. This venue increased the probability of face-to-face meetings and the evolution of joint ***collaborative research*** proposal development. The Foundation fostered this outcome by providing an array of training opportunities in high-income countries' (HIC) labs and carefully selecting and matching candidates with host scientists based on common research interests and the needs of their home institutions with regard to future human capacity building. The IPRB supported several ***information technologies*** to serve participants' needs; distributed theses, reprints, books, and patents; and published the "Rice Biotechnology Quarterly," a newsletter serving program participants worldwide. At international meetings, participants could be updated on the latest (prepublication) rice biotechnology science as well as the latest in "priorities" established by the social scientists' network. Thus, the international collaborative research and training mechanism supported by the IPRB resulted in excellent synergies and benefits far outweighing the level of financial support available from a single donor organization.

Allocation of funds

Table 1 depicts the allocation of funds over the lifetime of the IPRB. Several trends are noteworthy. Since its inception, IPRB has dispensed almost $105 million, an average of about $6.2 million per year. The breakdown between research and training was

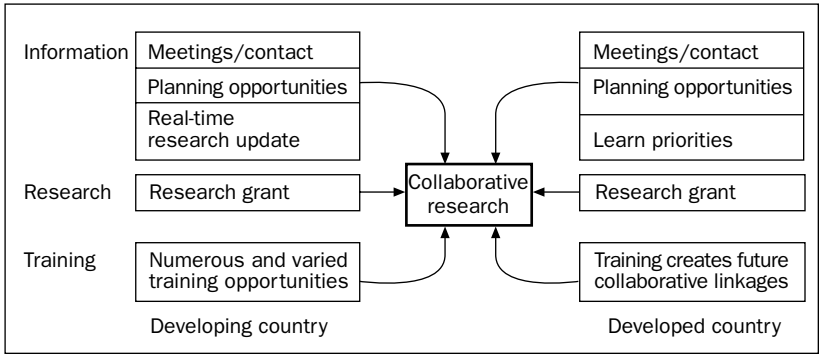


Fig. 1. Research-driven program components.

Table 1. The Rockefeller Foundation's International Program on Rice Biotechnology expenditures × $1,000.

Year	Basic research (HIC)[a]	Applied research (LIC)[b]	International centers	Social science	Meetings/ administration	Fellowships/ training	Total
2000	157	1,400	500	0	64	810	2,931
1999	468	1,689	1,006	55	466	2,266	5,950
1998	561	2,480	729	50	288	2,305	6,413
1997	566	2,068	936	0	523	2,418	6,511
1996	2,073	1,462	1,161	289	346	2,173	7,504
1995	1,974	1,845	1,289	280	240	2,071	7,699
1994	1,263	2,139	1,622	100	614	1,878	7,616
1993	2,400	1,537	1,857	307	176	2,525	8,802
1992	2,474	1,591	1,088	405	499	2,305	8,362
1991	2,081	1,309	800	69	385	2,160	6,804
1990	3,100	1,847	1,092	196	284	2,050	8,569
1989	3,049	2,811	773	181	372	1,038	8,224
1988	1,689	718	655	467	100	635	4,264
1987	4,753	170	621	1,217	100	368	7,229
1986	1,530	125	746	0	155	364	2,920
1985	859	131	15	0	34	427	1,466
1984	2,780	131	50	0	15	488	3,464
Total	31,777	23,453	14,940	3,616	4,661	26,281	104,728

[a]HIC = high-income countries of the industrialized world. [b]LIC = low-income countries of the developing world.

approximately 70% and 30%, respectively. However, as noted above, the training program was well integrated into achieving the research priorities and hence much of the support to training also contributed directly to the research achievements. In the same manner, the apparent allocation of funds to HIC and low-income country (LIC) institutions of 30% and 47%, respectively, is also flawed. Because of the highly integrated program implementation, the remaining 23% was used to create many "bridging" elements, such as meetings and workshops, which, along with the priority-setting research, and integrated *training-cum-research relationships often leading to future international collaboration*, contributed to the close linkages among HICs, LICs, and the international centers.

As noted above, the initial years were devoted primarily to basic research support in HICs. This peaked in 1989-90 with a concomitant increase in funding of LIC research and a sustained increase in training of rice scientists from Asia, made possible by the scientific knowledge base then existing in the HIC laboratories and CGIAR centers after the initial 5–6 years of IPRB support. In addition, although small in magnitude, the funding for social science research on priority setting was sustained from 1987 through 1996, reaching almost all countries participating in the IPRB. The impact on program direction and hence effectiveness was extremely valuable and the training accomplished in research priority setting and management continues to make research more effective across Asia's rice-consuming countries.

Scientific progress and outputs

In the limited space available, we are unable to recount the *many* success stories and scientific achievements associated with the IPRB's comprehensive scheme, long tenure, and many partnerships with associated supporters of rice research. In the following few paragraphs, we will merely touch on a few of the salient research outputs. We apologize to the innumerable researchers and fellows whose work will not be mentioned and trust that relevant reviews will more fully document their work. Here we will emphasize the role of international collaborative research and training that was pivotal to the program's multifaceted scientific and capacity-building output.

The tools of rice biotechnology

When the IPRB began in 1984, little was known about the rice genome at the molecular level and essentially few molecular tools were available for conducting rice biotechnology research. It was unknown then that rice had attributes that would make it especially amenable to genomic research. While initially reluctant to work on what was for them a new plant system, several leading laboratories accepted Foundation funding and quickly began generating results that were at the forefront of plant biotechnology and materials they readily shared with others. Training courses and workshops were sponsored that helped to rapidly transfer these methods and materials across the IPRB network. Rice became the first cereal regenerated from protoplasts and the first cereal transformed via protoplast, particle gun, and *Agrobacterium*-based methods. It was also discovered that rice had the smallest cereal genome, a relatively high percentage of single-copy DNA, and only one small chromosome duplication. By 1988, a molecular genetic map of rice was produced (McCouch et al 1988) and special funding was provided for dissemination of the map and its DNA markers worldwide. In the 1990s, rice became the model plant for cereal genomic research and full-scale rice genome sequencing projects began in Japan and the United States, which have now been combined and expanded to become the International Rice Genome Sequencing Project. And, at the last General Meeting of the IPRB, 20-24 September 1999 in Thailand, an international Rice Functional Genomics Working Group was formulated.

Molecular plant pathology

Molecular plant pathology showed perhaps the most significant and rapid research progress and applied product development as well as being a beacon for collaborative research and capacity building. One story that embodies nearly every aspect of the IPRB research agenda as well as the application of rice biotechnology tools was the discovery of the bacterial blight resistance gene *Xa21* and the interesting saga that followed (Box 1). One of the central publications in the series chronicling this effort stated, "Characterization of *Xa21* should facilitate understanding of plant disease resistance and lead to engineered resistance in rice" (Song et al 1995). Indeed, it did both and in this way contributed significantly to plant pathology's understanding of the basis of plant disease resistance (Ronald 1997) and the genetic mechanisms that are the basis of resistance gene family evolution (Song et al 1997, Wang et al 1998, Richter and Ronald 2000).

Box no. 1

The saga of *Xa21*

THE STORY OF BACTERIAL blight disease resistance gene *Xa21* epitomizes nearly all the IPRB program components in Figure 1 and above all the extension of the program goals from knowledge generation to production of improved rice varieties. The following is a brief "telegraphic" coverage of that voyage of discovery and ultimate application for the benefit of Asian rice farmers.

1977 *Oryza longistaminata* lines originating from Mali, Africa, noted to carry broad-spectrum resistance to bacterial blight

198? *Xa21* transferred into *O. sativa* background through interspecific hybridization (Khush et al 1991)

1990 *Xa21* locus RFLP mapped (Ronald et al 1992)

1992-95 Map-based cloning via bacterial artificial chromosome library construction (Wang et al 1995); *Xa21* sequencing and demonstration of engineered resistance of a susceptible genotype (Song et al 1995, Wang et al 1996)

1995 Patent filed and eventual U.S. Patent Number 5,859,339 granted to Ronald et al, 12 January 1999; innovative institution founded (Genetics Resources Recognition Fund—University of California, Davis) to use license fees/royalties to assist science capacity building in developing countries (Ronald PC, personal communication, 28 September 1997)

1997 *Xa21* pyramided with other *Xa* R genes via PCR-based marker-assisted selection (Huang et al 1997, Reddy et al 1997)

1998 *Xa21* experimentally transformed into elite rice varieties (Tu et al 1998, Zhang et al 1998)

1999 Field trials of pyramided *Xa* genes, including *Xa21*, reported in China, India, Indonesia, and Philippines (Rockefeller Foundation 1999)

2000 Commercial hybrid restorer line genetically improved by marker-assisted selection of *Xa21* and resulting hybrid rice demonstrates field-level efficacy (Chen et al 2000)

Over the 17-year period of the IPRB, we estimate that four Predoctoral, four Postdoctoral, and five Biotechnology Career Fellows took part in the above international research collaboration, transferring the basic skills, knowledge, and other materials to their home institutions. In one prominent publication, Song et al (1995), the 12 authors represented four research institutions in China, France, Korea, and the United States. The research background for that publication truly represented unprecedented international collaborative research, training, and capacity building spanning the globe.

Another salient example is that of the research output over approximately 15 years on the rice blast fungal pathogen. In both the bacterial blight and blast disease examples, DNA molecular tools were used, based on a solid 30+ years' foundation of conventional plant pathology and rice genetics, to provide dramatic new and informative insights into host-plant-pathogen interactions (Wang and Leung 1999). In both blast and bacterial blight diseases, DNA marker studies of the pathogens' genomes allowed new information on the geographic array of the pathogens' genetically divergent strains. This information was crucial in both basic understanding of the evolution and distribution of these diseases and in the applied art and science of field screening and "smart" deployment of specific resistance genes to specific geographic zones. Much like the bacterial blight saga (Box 1), the results of the past 15 years' molecular characterization of the blast fungus and the continuous discovery and characterization of blast R genes present in *Oryza* species led to significant basic knowledge of the fungal and rice genomes. This developed a new appreciation and understanding of the fungus's capacity for genotypic variation as well as the evolution of polymerase chain reaction (PCR) markers for specific blast resistance genes and their effective and rapid marker-assisted backcrossing into elite rice varieties. The capacity built by the IPRB in various national rice improvement programs has been used under the Asian Rice Biotechnology Network (ARBN) managed by the International Rice Research Institute (IRRI). Under the ARBN, the relevant PCR markers to genetically manipulate the bacterial and fungal R genes noted above are shared along with increasingly economical lab and field protocols. This international synergy has resulted in the most rapid and targeted deployment of new disease R genes possible. All concerned are to be congratulated!

As noted in the section on priority setting, tungro virus disease and other rice viral diseases ranked high as international and national constraints to rice yield. Again the advent of molecular tools allowed the revelation of an extremely intricate "natural history" story in the case of tungro virus and equally new scientific knowledge of other viruses. The discovery of two nucleic acid forms (spherical = single-stranded RNA and bacilliform = double-stranded DNA) of tungro virus and full sequencing of the virus genomes and the ability to then trace them through the insect vector (green leafhopper) contributed greatly to the understanding of the suite of biological characters responsible for field symptoms of the notoriously episodic damage feared by rice farmers. The high priority placed on viruses by the IPRB prioritization studies also meant that, of the 13 viruses known to attack Asian rice, almost all were partially or fully DNA/RNA-sequenced in the past ten years (Waterhouse and Upadhyaya 1999). With this knowledge in hand, a fruitful scientific dialogue ensued among rice scientists worldwide regarding the most effective way to use this knowledge base for the genetic improvement of rice.

Enhancing resistance to insects

Enhancing resistance to insects is the twin traditional goal of rice breeding with some 100+ species of insects attacking the rice crop. Unlike the study of the pest genome in diseases, insect genomics did not receive as much attention. Nevertheless, as with

viruses, molecular tools made possible a new understanding of both host-plant insect resistance and the natural history of some significant insect pests. In general, the use of molecular markers and the transgenic expression of insecticidal proteins received major attention. Bennett et al (1997) provide an excellent review of the contributions made by various biotechnology tools—interspecific hybridization, molecular markers, and rice plant transformation—toward the goal of enhanced insect resistance. Their review along with that of Mohan et al (1997), which ranges from genome mapping and cross-species synteny to the application of DNA markers in crop breeding, illustrate well the dramatic new genetic resources available to rice breeders regardless of the insect pest being targeted—stem borers, planthoppers, gall midge, or leaffolder.

Abiotic stresses—flood and drought

Flood/submergence tolerance research focused on studies of the basic response of rice to flooding/submergence and represents a classic example of international collaboration within the IPRB. Collaborative networking included scientists from Australia, Bangladesh, India, Japan, Philippines, Thailand, and the United States working interactively, as well as competitively, for more than a decade (Hossain et al 1996, Huq et al 1999). The studies embraced basic research related to critical gene isolation and regulation in transformed rice as well as molecular marker-assisted selection based on extensive field-level screening (Xu and Mackill 1996, Nandi et al 1997). The output from the various approaches (gene isolation/characterization versus quantitative trait loci identification) has begun to overlap as functional genomics became more pervasive than the "tools × traits" approach to genetic improvement. Those involved in submergence/flood tolerance research are in the forefront of candidate gene searches and international rice sequencing efforts (Normile 1999a). Other examples of the evolution and convergence of rice molecular genetics are widespread among IPRB grantee collaborative networks and illustrate yet another significant outcome of the network. This leads us to ask: Unless these researchers had been collaborating internationally and had a forum to communicate directly, would they be ready for the "next wave" in rice biotechnology—functional genomics and the bioinformatics revolution—now beginning?

Drought tolerance received high priority in the IPRB's earliest priority-setting exercise (1986) and subsequent national exercises further documented its importance. However, water deficit, unlike the abovementioned water excess, is one of crop genetic improvement's least understood genetic traits and is considered intractable by some. Nevertheless, over the past decade, a few dedicated researchers have made significant progress, both from the perspective of molecular marker tagging of traits thought to enhance drought tolerance—root system parameters (Champoux et al 1995, Ray et al 1996, Courtois et al 2000) and osmotic adjustment of tissues (Lilley et al 1996, Zhang et al 1999a,b)—but also in the creation of experimental rice transgenics with increasing levels of stress-inducible promoter and gene construct sophistication, which demonstrate striking responses to water deficit in growth chamber and greenhouse trials (Xu et al 1996, Su et al 1998, Bajaj et al 1999). The considerable accumulation of knowledge, information, and research experience was the subject of two

recent international workshops to plot future research strategy, not only for rice but also for the five major cereals that feed human populations globally (Ito et al 1999, Ribaut and Poland 2000).

The future use of molecular markers to combine or pyramid abiotic stress resistance genes with those of disease and insect resistance, in an efficient and timely manner, could only have been a rice breeder's dream just a decade ago. Today, however, with Internet access to molecular genotype and phenotype databases and the international sharing of many different types of genetic resources (germplasm and DNA-based technologies), even such complex traits and combinations of traits are approachable (Xu 1997).

Comparative genomics

Comparative genomics and the discoveries related to the study of synteny among cereal genomes demonstrated the unequivocal superiority of DNA-based molecular tools for investigating the long-standing questions of cereal evolution. In addition, it demonstrated what has become the ultimate discovery of the rice biotechnology adventure—the central role of the rice genome in understanding and technically accessing the far-larger genomes of such major crops as maize and wheat (Box 2).

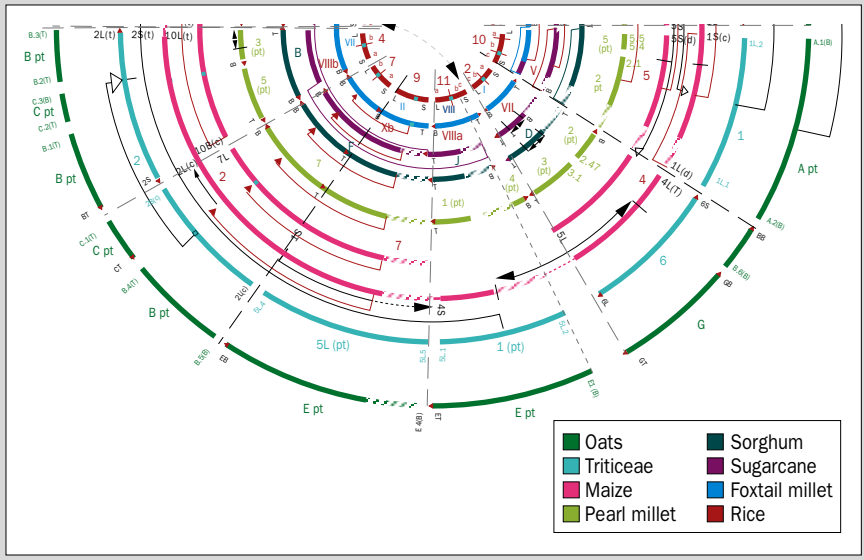
Early in the 1990s, it was apparent that higher plant species must share extensive "colinearity" of DNA markers and presumably genes across genera in both dicots as well as monocots. The prospect of conservation of linearity within linkage groups generated great excitement. For, if this were true, genetic information and related molecular markers might be exploited in related species with less well-developed molecular maps and markers. From 1988 to 1993, numerous papers related the news that indeed colinearity was a fact across species and even genera. However, even more striking and fortuitous for those working with the rice genome was the development of comparative or consensus maps within the cereals (Bennetzen and Freeling 1993). In their conceptual framework, Moore et al (1995) postulated that, based on the alignment of 19 linkage groups/segments from rice, one could form the building blocks of six of the world's major cereal food crops. A flood of research reports followed that not only confirmed the conceptual model but also quickly illustrated how gene discovery in the rice genome, partially because of its small size and saturated molecular maps, could rapidly and effectively be transferred to the other cereals. By 1998, Van Deynze et al (1998) had produced a set of RFLP anchor probes to facilitate comparative mapping across grass genera. Shortly thereafter, the applications forecast earlier began to be apparent. Leister et al (1999) demonstrated the RFLP and physical mapping of resistance gene homologues in rice and barley. Coincidentally, the R genes from rice were the genes with race-specific resistance to blast and bacterial disease noted at the beginning of this section.

These scientific advances, combined with the advent of electronic mail (an equalizer of time, space, and national origin without parallel), added a multiplier effect to the IPRB investment that could not have been imagined 15 years earlier. Researchers worldwide have experienced research synergy unheralded over the past 50 years of rice genetics, as rice, "the pivotal genome," was placed in a unique position along with those researchers who had helped to create the science of rice biotechnology.

Box no. 2

Rice—the pivotal genome

COMPARATIVE GENOMICS HAS revealed a level of conservation in gene content and order within the grasses that surprised even the most experienced geneticists. After the landmark publication of Moore et al (1995), which indicated that all major cereal crop genomes could be represented by 19 segments found in the rice genome, rice has taken a solid position at the *center* of both graphic presentations (adapted from Gale and Devos 1998, this box) as well as research efforts to use map-based cloning for gene discovery and isolation from the much larger genomes of oat, wheat, and maize. The rice genome, with only 400 million DNA base pairs (bp), is about four times larger than *Arabidopsis*, the model dicot. This, coupled with a dense genetic map of >2,500 markers, physical maps of the entire genome, large public collections of expressed sequence tags (ESTs), and the near-term prospect of the complete rice genome sequence, makes rice a model crop for the cereals.



Institutional and human resource capacity building[1]

Capacity building is an extremely difficult concept to define. When related to the IPRB, where it refers to national networks and institutions as well as individuals, it can result in a great lack of clarity, which might require a full-length paper to demystify.

[1]This section on capacity building under the IPRB relies heavily on the report "Capacity Building Evaluation of the International Program on Rice Biotechnology (IPRB)," June-October 1999, by Madan Mohan, Leocadio Sebastian, Kangle Zhang, and David Norman. Rockefeller Foundation. 249 p.

For the purposes of this discussion, we will use the following, not as a definition but as the major tools of capacity building under the IPRB.

The major elements of the IPRB that contribute to capacity building are (Toenniessen 1998)

1. Fellowships and courses offering specialized training, skills maintenance, and technology transfer.
2. The enabling environment provided by a network of scientists who are conducting related research and are eager to share ideas and materials.
3. Partnerships with national agencies that assume increasing responsibility for funding and management.
4. Access to relevant information and effective communications systems.
5. A rational process for establishing research priorities.
6. Renewable research grants having application and monitoring processes that place strong public emphasis on the use of rigorous scientific methods and peer review.
7. The emergence of centers of strength capable of playing a leadership role in the future.
8. The work of Foundation field staff scientists located in Asia.

In statistical terms, approximately 700 scientists from more than 30 countries have participated directly in the IPRB. Grantees and fellows came from 12 HIC, 16 LIC, 4 CGIAR centers, and the International Center for Genetic Engineering and Biotechnology. After the "invention of rice biotechnology" phase, the program's primary goal was directed toward building capacity in the rice research institutions of Asia's rice-consuming countries.

The types of institutions

The types of institutions that were contacted by Foundation staff and invited to submit research proposals were a departure from previous "traditional" agricultural research institutions. The fact that a new science was being promoted required the Foundation to include many traditional or basic science institutions in the constellation of institutional grantees. Hence, the IPRB became known early on for its efforts to include nontraditional agricultural institutions, even though the program theme was obviously genetic improvement of rice. However, from the outset it was clear that in the larger countries an intranational network apparatus would be required to achieve the national-level collaboration required for successful application of biotechnologies to rice genetic improvement. Table 2 illustrates the relative frequency of institutional types that eventually made up the intranational and international networks under the IPRB umbrella and the different types of LIC institutions supported by the IPRB. Universities made up 43% of the total, whereas research institutions accounted for 57%. An alternative view is that 47% of the institutions had an agricultural research focus, whereas the remaining 53% did not necessarily share such a focus. Since its inception, the IPRB has awarded grants to 77 different LIC institutions in 16 different countries, 75% of which are in Asia. Institutions in two countries with the largest human populations and rice research systems, China and India, have

Table 2. Types of institutions supported by the Rockefeller Foundation's International Program on Rice Biotechnology.[a]

Country	Type of institution				
	University		Research institution		
	Conventional	Agriculture	Basic	Agriculture	Rice
Bangladesh	1	–	–	–	1
China	3	6	6	5	2
India	8	6	6	2	2
Indonesia	–	–	1	1	–
Malaysia	–	–	–	1	–
Nepal	–	–	–	3	–
Pakistan	1	–	–	1	–
Philippines	1	–	–	1	–
South Korea	1	–	–	1	–
Sri Lanka	–	–	–	1	–
Thailand	3	–	2	–	1
Vietnam	1	–	3	1	1
Latin America	2	–	2	–	–
Total (77)	21	12	20	17	7
Percentage of institutions	27.3	15.6	26.0	22.1	9.0

[a]Only institutions that received research grants were considered.

been the major beneficiaries, together accounting for 62% of the grantees and about the same proportion of total research funds.

The CGIAR-supported international centers

The CGIAR-supported international centers played a primary role from the IPRB's inception. IRRI in the Philippines, the International Center for Tropical Agriculture (CIAT) in Colombia, and the West Africa Rice Development Association (WARDA) in Côte d'Ivoire have as part of their mandate building rice research capacity in developing countries. In 1985, however, these institutions had limited capabilities in biotechnology, as was the case with most international agricultural research institutions. In the late 1980s, Foundation funding enabled IRRI and CIAT to expand their capacities in tissue/anther culture and interspecific hybridization while beginning to develop DNA molecular biology capacity. WARDA followed suit in the 1990s.

By 1990, it was evident that biotechnology would significantly affect rice breeding and IRRI and CIAT began using their core funds to satisfy new staff requirements. Like the national networks' requirement for greater diversity in institutional scientific capability, these international centers now have significant capacity to use biotechnology tools, but they also maintain many international collaborative research partnerships to cover the more sophisticated and costly technologies. The centers have offered a series of training courses for colleagues from national programs. With support from the Asian Development Bank, IRRI began the ARBN in 1996. In this way, IRRI has significantly expanded its capacity to work collaboratively with national centers, many of which were assisted by the IPRB institutional and individual capacity-building process.

Because of the great diversity (scientifically as well as geographically) of institutions involved simultaneously in the IPRB, program management was challenged to provide the full spectrum of research and capacity-building opportunities. However, several of the innovative instruments noted in the following section have proved to be robust across the great institutional and individual diversity embodied in the IPRB.

Formal training

Formal training under the IPRB had six types of fellowships, all of which featured candidates from grantee institutions in LICs who were hosted at an advanced laboratory in an HIC or at IRRI or CIAT:

- Ph.D. Fellowship—Fellow receives training for 4–5 years at an advanced research institution.
- Dissertation Fellowship—Fellow receives Ph.D. degree from home institution but conducts dissertation research at an advanced research institution for 1–2 years.
- Postdoctoral Fellowship—Fellow conducts research at an advanced research institution for 2 years.
- Visiting Scientist Fellowship—Fellow serves as visiting researcher at an advanced research institution for 1–2 years.
- Biotechnology Career Fellowship—Fellow conducts part of collaborative research program at an advanced research institution for 3 months per year over 3 years.
- Technology Transfer Fellowship—Fellow from advanced research institution conducts collaborative research at an institution in a rice-dependent country for 3 months per year over 3 years.

Impact of the formal training program. The impact of the formal training program under the IPRB is extremely difficult to gauge. Table 3 indicates the number of individuals trained by the different number of institutions in the 12 Asian countries and Latin America. China and India clearly have the largest number of institutions in which a broad range of skills and in-depth expertise can be found. Overall, the numbers varied substantially across countries with about 36% of the institutions having five or more people trained, whereas 22% had zero trained under the IPRB. Thus, it appears that the impact of the IPRB formal training program should be greatest in six countries: China, India, Philippines, South Korea, Thailand, and Vietnam.

Common problems associated with international training. The common problems associated with international training efforts were addressed in the planning and management of the IPRB using the six primary fellowships above and research grants:

- "Brain drain" or immigration out of the home country was addressed by the incentives of eligibility for a research grant and, after one year of in-country research, eligibility for a Biotechnology Career Fellowship.
- "Irrelevant training" was minimized because fellows were, as much as possible, directed to host labs with similar research interests and, later in the program lifetime, dependent on jointly agreed upon research proposals as a part of the home-institution endorsement and host-institution acceptance.

Table 3. Breakdown of number of institutions by country where formal training was sponsored under the Rockefeller Foundation's International Program on Rice Biotechnology.[a]

Country	Number of institutions according to number of researchers trained[b]						
	>20	15–19	10–14	5–9	3–4	1–2	0
Bangladesh	–	–	–	1	1	–	–
China	–	1	4	4	2	6	5
India	1	1	–	5	7	7	3
Indonesia	–	–	–	1	–	1	–
Malaysia	–	–	–	–	–	1	–
Nepal	–	–	–	1	–	–	2
Pakistan	–	–	–	1	–	–	1
Philippines	–	–	1	1	–	–	–
South Korea	–	–	1	–	–	–	1
Sri Lanka	–	–	–	–	–	–	1
Thailand	–	–	1	1	–	3	1
Vietnam	–	–	1	2	–	3	–
Latin America	–	–	–	–	–	1	3
Total	1	2	8	17	10	22	17

[a]Only institutions that received research grants were considered. [b]The individual column headings reflect the number of scientists trained under IPRB sponsorship.

- "Re-posting" or transfer to an unrelated field of research. The IPRB management had little influence on this matter. But persistent invitations to national and international meetings and indication of eligibility for research funds were occasionally effective in redressing the re-posting issue.
- "Lack of indigenous support" was temporarily overcome with eligibility to apply for a research grant.
- "Nonresponsive home-institution administration" was also little influenced by the IPRB. However, receiving a research grant, invitations to international meetings, and other types of supportive services were often an incentive for institutional administrative personnel to be more responsive to the researcher's needs.

The role of periodic international meetings

The role of periodic international meetings of the entire IPRB network, numbering 350–400+ participants in the final five years, was of immeasurable benefit to the program collaborative research and capacity-building goals. As Figure 1 shows, the international meetings served as a physical venue for networking with colleagues, particularly from other countries, and conducting real-time, face-to-face negotiations of future collaborative research and training plans. The meetings featured well-organized full schedules (day and night sessions); covered extremely wide-ranging biotechnology topics while focusing on a single unifying theme, rice; "forced" accountability took place because a poster and/or verbal presentation was expected along with critical review by peers and Foundation expert advisors; and "think tanks" stimulated future plans to assure that intermediate research products went into the hands of rice

breeders where final products to enhance the productivity and welfare of farmers and consumers could be realized.

Sustainability

In 1999 the Foundation determined that, after 16 years, the IPRB had accomplished its goals and it was time for an orderly closing of the program during 2000. Soon afterward, the news was publicly available (Normile 1999b) and the question of the IPRB's "sustainability" came to the fore. Sustainability of the gains from such a complex and long-term (17 years) international program can be considered for many facets of the IPRB: physical infrastructure and its maintenance, funding for research, maintaining and enhancing the investment in human capital, etc. Indeed, capacity building may be viewed at the individual, institutional, or national level. Much has been written on this subject (Cohen et al 1998, Falconi 1999, Byerlee and Gregory 1999) in recent years. Although the IPRB attempted to incorporate a national perspective, at least for larger countries such as China and India, questions of sustainability are most realistically contemplated at the institutional level. This is indeed a complex issue, but our experience points to one major factor—***leadership***. Institutions in which enlightened and energetic leadership was coupled with a suitable foundation of trained and motivated scientists appear to be well positioned to sustain their research momentum from both national and international sources. Some IPRB participants have used their network linkages to acquire funding from both public- and private-sector sources, thus enhancing their prospects for future support in a public-private partnership world.

The future of rice biotechnology

The past 15 years have so dramatically changed our knowledge of rice genetics that we would like to paraphrase Lander and Weinberg's (2000) recent *Science* article because we can clearly see that the *future of rice biotechnology is now*: "biology enters this century in possession, for the first time, of the mysterious instruction book first postulated by Hippocrates and Aristotle. How far will this take us in explaining the vast complexity of the biological world?".... The solutions to many problems long resistant to attack are now within our reach. The prospects of 21st century biology are surely breathtaking."

We are tempted to point to but a handful of future rice biotechnology outcomes that appear certain to be realized in the relatively near future:

- Rice genomics and functional genomics will demonstrate the enormous benefits of combining traditional (genetics, physiology, biochemistry) and new approaches (bioinformatics) into a fuller understanding of rice biology.
- A major application of the above will be in the regulation of gene expression and resulting manipulation of reproductive development in tandem with the concomitant control of senescence in leaves and other organs.
- Crop response to adverse environmental conditions will see the merger of genomics with rice physiology and biochemistry to more fully understand rice's potential for adaptation to marginal environments and to genetically improve rice for the constraints of the 21st century—water and salinity.

- Transformation of chloroplast genomes will be recognized as a means of containment for such novel genes as herbicide resistance and as such will significantly modify rice farming, especially in rainfed ecosystems.
- The "pocket laboratory" will soon bring PCR, microarrays, and other molecular biology protocols to the greenhouse and field, making marker-assisted selection a powerful practical tool for rice breeders.
- Rice grain quality and nutrition will be addressed to make new rice varieties available that can improve the overall health of rice consumers and generate a multitude of rice-based products for specialty markets.

The foundation laid for these advances, in part through the collaborative research, training, and capacity building accomplished under the IPRB, is only the beginning of the fruits that the world's poor farmers and rice consumers will need if the world's neediest are not to go hungry in the 21st century. These advances in rice biotechnology must be equitably deployed by a strong public-sector agricultural research effort if the ultimate gains are to be shared among those most deserving of them (Conway and Toenniessen 1999).

Concluding remarks

Those of us at the Rockefeller Foundation who have been associated with the International Program on Rice Biotechnology have found it to be an exciting, rewarding, and learning experience. Our grantees, fellows, advisors, and consultants made the important contributions that came together and were readily shared in a highly successful and truly international program. In the process, many became our good friends and colleagues. It was difficult for us to make the decision to bring Foundation funding for the program to a close. However, we are confident that the rice research community, particularly in Asia, now has the capacity to keep rice at the forefront of biotechnology research and to produce the new rice varieties the world urgently needs, as long as the community continues to share and work collaboratively.

References

Bajaj S, Targolli J, Liu L-F, Ho T-HD, Wu R. 1999. Transgenic approaches to increase dehydration-stress tolerance in plants. Mol. Breed. 5:493-503.

Bennett J, Cohen MB, Katiyar SK, Ghareyazie B, Khush GS. 1997. Enhancing insect resistance in rice through biotechnology. In: Carozzi N, Koziel M, editors. Advances in insect control. London (UK): Taylor & Francis. p 75-93.

Bennetzen JL, Freeling M. 1993. Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet. 9:259-261.

Byerlee DE, Gregory P. 1999. Agricultural biotechnology and rural development: options for the World Bank. Biotechnology Task Force, World Bank, Washington, D.C.

Champoux MC, Wang G, Sarkarung S, Mackill DJ, O'Toole JC, Huang N, McCouch SR. 1995. Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. Theor. Appl. Genet. 90:969-981.

Chen S, Lin XH, Xu CG, Zhang Q. 2000. Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci. 40:239-244.

Cohen JL, Falconi C, Komen J. 1998. Proprietary biotechnology inputs and international agricultural research. ISNAR Briefing Paper No. 38. The Hague (Netherlands): International Service for National Agricultural Research.

Conway G, Toenniessen GH. 1999. Feeding the world in the twenty-first century. Nature 402(suppl):C55-C58.

Courtois B, McLaren G, Sinha PK, Prasad K, Yadav R, Shen L. 2000. Mapping QTLs associated with drought avoidance in upland rice. Mol. Breed. 6:55-66.

Devos KM, Gale MD. 2000. Genome relationships: the grass model in current research. Plant Cell 12:637-646.

Evenson RE, Herdt RW, Hossain M, editors. 1996. Rice research in Asia: progress and priorities. Manila (Philippines): International Rice Research Institute, and Wallingford (UK): CAB International. 418 p.

Falconi C. 1999. What is agricultural biotechnology research capacity in developing countries? ISNAR Discussion Paper No. 99-10. The Hague (Netherlands): International Service for National Agricultural Research.

Gale MD, Devos KM. 1998. Plant comparative genetics after 10 years. Science 282:656-658.

Herdt RW. 1986. Information needs for research resource allocation. Prepared for the workshop on Biotechnology for Crop Improvement, 13-17 Oct. 1986, International Rice Research Institute (IRRI). Los Baños (Philippines): IRRI. 32 p.

Herdt RW. 1987. Equity considerations in setting priorities for Third World rice biotechnology research. Dev. Seeds Change 4:19-24.

Herdt RW. 1991. Research priorities for rice biotechnology. In: Khush GS, Toenniessen GH, editors. Rice biotechnology. Manila (Philippines); International Rice Research Institute, and Wallingford (UK): CAB International. p 19-54.

Herdt RW, Rieley Jr. R. 1987. International rice research priorities: implications for biotechnology initiatives. Prepared for the Rockefeller Foundation Workshop on Allocating Resources for Developing Country Agricultural Research, 6-10 July 1987, Bellagio, Italy. 78 p.

Hossain MA, Huq E, Grover A, Dennis E, Peacock WJ, Hodges TK. 1996. Characterization of pyruvate decarboxylase genes from rice. Plant Mol. Biol. 31:761-770.

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95:313-320.

Huq E, Harrington S, Hossain MA, Wen F, McCouch SR, Hodges TK. 1999. Molecular characterization of *pdc2* and mapping of three *pdc* genes from rice. Theor. Appl. Genet. 98:815-824.

Ito O, O'Toole J, Hardy B, editors. 1999. Genetic improvement of rice for water-limited environments. Proceedings of the Workshop on Genetic Improvement of Rice for Water-Limited Environments, 1-3 December 1998, Los Baños, Philippines. Manila (Philippines): International Rice Research Institute. 353 p.

Kearl B, Adkisson P, McNamara R. 1982. A review of the Rockefeller Foundation Conquest of Hunger Program. New York (USA): Rockefeller Foundation.

Khush GS, Bacalangco E, Ogawa T. 1991. A new gene for resistance to bacterial blight from *O. longistaminata*. Rice Genet. Newsl. 7:121-122.

Lander ES, Weinberg RA. 2000. Genomics: journey to the center of biology. Science 287:1777-1782.

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Graner A, Schulze-Lefert P. 1999. RFLP and physical mapping of resistance gene homologues in rice (*O. sativa*) and barley (*H. vulgare*). Theor. Appl. Genet. 98:509-520.

Lilley JM, Ludlow MM, McCouch SR, O'Toole JC. 1996. Locating QTL for osmotic adjustment and dehydration tolerance in rice. J. Exp. Bot. 47:1427-1436.

Lin J, Shen M. 1993. Rice production constraints in China: implication for biotechnology initiative. Prepared for the 6th Annual Meeting of the Rockefeller Foundation's International Program on Rice Biotechnology, 1-5 Feb. 1993, Chiang Mai, Thailand. 33 p.

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed. 3:87-103.

Moore G, Devos KM, Wang Z, Gale MD. 1995. Grasses, line up and form a circle. Curr. Biol. 5:737-739.

Nandi S, Subudhi PK, Senadhira D, Manigbas NL, Sen-Mandi S, Huang N. 1997. Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. Mol. Gen. Genet. 255:1-8.

Normile D. 1999a. U.S. adds $12 million to global sequencing push. Science 286:24-25.

Normile D. 1999b. Rockefeller to end network after 15 years of success. Science 286:1468-1469.

Ray JD, Yu L, McCouch SR, Champoux MC, Wang G, Nguyen HT. 1996. Mapping quantitative trait loci associated with root penetration ability in rice (*Oryza sativa* L.). Theor. Appl. Genet. 92:627-636.

Reddy JN, Baraoidan MR, Bernardo MA, George MLC, Sridhar R. 1997. Application of marker-assisted selection in rice for bacterial blight resistance gene, *Xa21*. Curr. Sci. 73:873-875.

Ribaut J-M, Poland D, editors. 2000. Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. A strategic planning workshop held at CIMMYT, El Batán, Mexico, 21-25 June 1999. Mexico, D.F.: International Maize and Wheat Improvement Center. 172 p.

Richter TE, Ronald PC. 2000. The evolution of disease resistance genes. Plant Mol. Biol. 42: 195-204.

Rockefeller Foundation. 1999. Proceedings (abstracts) of the meeting of the International Program on Rice Biotechnology. Phuket, Thailand, 20-24 September 1999. 388 p.

Ronald PC. 1997. The molecular basis of disease resistance in rice. Plant Mol. Biol. 35:179-186.

Ronald PC, Albano B, Tabien R, Abenes L, Wu K, McCouch S, Tanksley S. 1992. Genetic and physical analysis of the rice bacterial blight resistance locus, Xa21. Mol. Gen. Genet. 236:113-120.

Song W-Y, Wang G-L, Chen L-L, Kim H-S, Pi L-Y, Holsten T, Gardner J, Wang B, Zhai W-X, Zhu L-H, Fauquet C, Ronald P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804-1806.

Song W-Y, Pi L-Y, Wang G-L, Gardner J, Holsten T, Ronald PC. 1997. Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9:1279-1287.

Su J, Shen Q, Ho T-HD, Wu R. 1998. Dehydration-stress-regulated transgene expression in stably transformed rice plants. Plant Physiol. 117:913-922.

Toenniessen GH. 1998. Rice biotechnology capacity building in Asia. In: Ives CL, Bedford BM, editors. Agricultural biotechnology in international development. Wallingford (UK): CAB International. p 201-212.

Toenniessen GH, Herdt RW. 1988. The Rockefeller Foundation's International Program on Rice Biotechnology. Presented at USAID-sponsored conference on Strengthening Collaboration in Biotechnology: International Agriculture and the Private Sector, 17-21 April 1988, Washington, D.C. p 291-317.

Tu J, Ona I, Zhang Q, Mew TW, Khush GS, Datta SK. 1998. Transgenic rice variety 'IR72' with *Xa21* is resistant to bacterial blight. Theor. Appl. Genet. 97:31-36.

United States Patent No. 5,859,339. 12 Jan. 1999. PC Ronald, G-L Wang, W-Y Song. Nucleic acids, from *Oryza sativa*, which encode leucine-rich repeat polypeptides and enhance *Xanthomonas* resistance in plants.

Upadhyaya HK, Pokhrel TP, Shivakoti GP, Vaidya SN, Bimb HP. 1993. Rice research in Nepal: current state and future priorities. Prepared for the 6th Annual Meeting of the Rockefeller Foundation's International Program on Rice Biotechnology, 1-5 Feb. 1993, Chiang Mai, Thailand. 40 p.

Van Deynze AE, Sorrells ME, Park WD, Ayres NM, Fu H, Cartinhour SW, Paul E, McCouch SR. 1998. Anchor probes for comparative mapping of grass genera. Theor. Appl. Genet. 97:356-369.

Wang G-L, Leung H. 1999. Molecular biology of host-pathogen interactions in rice diseases. In: Shimamoto, K, editor. Molecular biology of rice. Tokyo (Japan): Springer-Verlag. p 201-232.

Wang G-L, Holsten TE, Song W-Y, Wang H-P, Ronald PC. 1995. Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant J. 7:525-533.

Wang G-L, Ruan D-L, Song W-Y, Sideris S, Chen L, Pi L-Y, Zhang S, Zhang Z, Fauquet C, Gaut BS, Whalen MC, Ronald PC. 1998. Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10:765-779.

Wang G-L, Song W-Y, Ruan D-L, Sideris S, Ronald PC. 1996. The cloned gene, *Xa21*, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic plants. Mol. Plant-Microbe Interact. 9:850-855.

Waterhouse PM, Upadhyaya NM. 1999. Genetic engineering of virus resistance. In: Shimamoto K, editor. Molecular biology of rice. Tokyo (Japan): Springer. p 257-281.

Widawsky DA, O'Toole JC. 1990. Prioritizing the rice research agenda for Eastern India. New York (USA): Rockefeller Foundation. 86 p.

Xu D, Duan X, Wang B, Hong B, Ho T-HD, Wu R. 1996. Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. Plant Physiol. 110:249-257.

Xu K, Mackill DJ. 1996. A major locus for submergence tolerance mapped on rice chromosome 9. Mol. Breed. 2:219-224.

Xu Y. 1997. Quantitative trait loci: separating, pyramiding, and cloning. Plant Breed. Rev. 15:85-139.

Zhang J, Nguyen HT, Blum A. 1999a. Genetic analysis of osmotic adjustment in crop plants. J. Exp. Bot. 50:291-302.

Zhang J, Zheng HG, Ali ML, Tripathy JN, Aarti A, Pathan MS, Sarial AK, Robin S, Nguyen TT, Babu RC, Nguyen BD, Sarkarung S, Blum A, Nguyen HT. 1999b. Progress on the molecular mapping of osmotic adjustment and root traits in rice. In: Ito O, O'Toole J, Hardy B, editors. Genetic improvement of rice for water-limited environments. Manila (Philippines): International Rice Research Institute. p 307-317.

Zhang S, Song W-Y, Chen L, Ruan D, Taylor N, Ronald P, Beachy R, Fauguet C. 1998. Transgenic elite indica rice varieties, resistant to *Xanthomonas oryzae* pv. *orzae*. Mol. Breed. 4:551-558.

Notes

Authors' addresses: J.C. O'Toole, Senior Scientist and Field Representative, Food Security, The Rockefeller Foundation, Suite 1602, Central Chidlom Tower, 22 Soi Somkid, Ploenchit Road, Bangkok 10330, Thailand; G.H. Toenniessen, Director, Food Security, The Rockefeller Foundation, 420 Fifth Ave, New York, NY, USA; T. Murashige, Senior Scientist (Retired), Food Security, The Rockefeller Foundation, New York; R.R. Harris, Senior Program Associate, Food Security, The Rockefeller Foundation, New York; and R.W. Herdt, Vice President for Program Administration, The Rockefeller Foundation, New York.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Molecular markers, genetic diversity, and evolution

Evolution and domestication of rice

H. Morishima

This chapter intends to explore some implications of rice evolution from the viewpoints of genetics and ecology. Core issues are (1) what genetic changes are associated with differentiation among and within species of cultivated rice and their wild relatives and (2) what factors are responsible for the domestication process. First, genetic diversity among and within AA genome species is summarized. Second, four directions of differentiation within the Asian AA genome gene pool are clarified: differentiation from wild to cultivated type (domestication), ecotype differentiation from perennials to annuals in wild races, geographical variation in wild races, and indica-japonica differentiation in cultivars. Third, the genetic basis of the domestication syndrome is discussed. Our recent study demonstrated that mapped genomic locations of quantitative trait loci (QTLs) tended to cluster, reflecting the domestication syndrome as well as the indica-japonica syndrome. This phenomenon was explained by "multifactorial linkages."Domestication might be a process driven by conscious and unconscious selection of adaptive gene blocks distributed over the genome.

In the past decade, a wealth of data provided by molecular markers, together with phenotypic, ecological, and archaeological data, significantly increased our evolutionary understanding of the genus *Oryza*. The target species dealt with in this chapter are diploid AA genome species—cultivated rice and its wild relatives. Several important problems such as the genetic basis of reproductive isolation are not included, but some new information obtained from our recent studies is discussed.

Inter- and intraspecific genetic diversity in the *O. sativa* complex

Delimitation of the related wild taxa of the cultigens

The two cultivated rice species, *Oryza sativa* L. and *O. glaberrima* Steud., belong to a species group called *Oryza sativa* complex together with the five wild taxa, *O. rufipogon* (sensu lato), *O. longistaminata* Chev. et Roehr., *O. barthii* A. Chev., *O. glumaepatula* Steud., and *O. meridionalis* Ng. This species complex was first

defined as the diploid species having the genome A in common. Later, the genetic similarity among these species was confirmed using isozymes (Second 1991), restriction fragment length polymorphisms (RFLPs) of nuclear DNA (Wang ZY et al 1992), cpDNA (Dally and Second 1990) and mtDNA (Second and Wang 1992), and amplified fragment length polymorphism (AFLP) (Aggarwal et al 1999). This might indicate that nuclear and organellar genomes that are considered to undergo different evolutionary rates evolved concertedly at this level.

Among these taxa, only *O. rufipogon* produces fertile F_1 hybrids with *O. sativa* and therefore these two species are considered to belong to a single biological species. Together with all circumstantial evidence, this suggests that *O. rufipogon* is the ancestor of *O. sativa*. Similarly, it leaves no doubt that *O. barthii* is the ancestor of African rice *O. glaberrima*.

Species relationships among the AA genome wild taxa

Oryza rufipogon is widely distributed in tropical and subtropical areas in Asia and Oceania and tends to differentiate into perennial and annual types. *O. barthii* is the annual species found mainly in West Africa. Another African species, *O. longistaminata*, is strongly perennial. *O. glumaepatula* is distributed in Latin America and varies in perenniality. *O. meridionalis* is an annual species found in Australia. Taxonomically, *O. longistaminata* is distinguished by its strong rhyzomatous habit and *O. barthii* by its short ligule from other species. But Asian, American, and Oceanian taxa are barely distinguishable from each other by morphology only because of a lack of clear key characters. Therefore, those species are primarily classified on the basis of geographical origins. These five wild taxa are isolated from each other by various reproductive barriers.

Nomenclature of the Asian AA genome wild taxon has been a controversial subject. The perennial and annual types are often designated as *O. rufipogon* Griff. and *O. nivara* Sharma et Shastry, respectively. Our group has classified both types in a single species, *O. rufipogon*, since they are interfertile and variation between the two types is continuous (Oka 1988).

With the hope of elucidating their phylogenetic relationships, many workers have studied genetic diversity among and within species. Nuclear gene markers demonstrated the distinct difference among species (Second 1985, Wang ZY et al 1992, Doi et al 1995, Ishii et al 1996). A trend commonly obtained from those studies was that *O. meridionalis* was remotely related to other species that are relatively close. The variation pattern revealed in organellar markers was more complex (Dally and Second 1990, Second and Wang 1992, Ishii et al 1988). The species relationships revealed were not always consistent among the different studies, perhaps partly because of the quantity and quality of the accessions used in the respective studies. Since wild taxa are polymorphic and heterogeneous, a good sample representing the whole genetic variation involved in the respective species should be used.

Recently, Akimoto (1999) made a biosystematic study using 183 accessions of the AA genome species preserved at the National Institute of Genetics, Mishima. Figure 1 illustrates the rough sketch of his results. In isozyme and nuclear DNA RFLP,

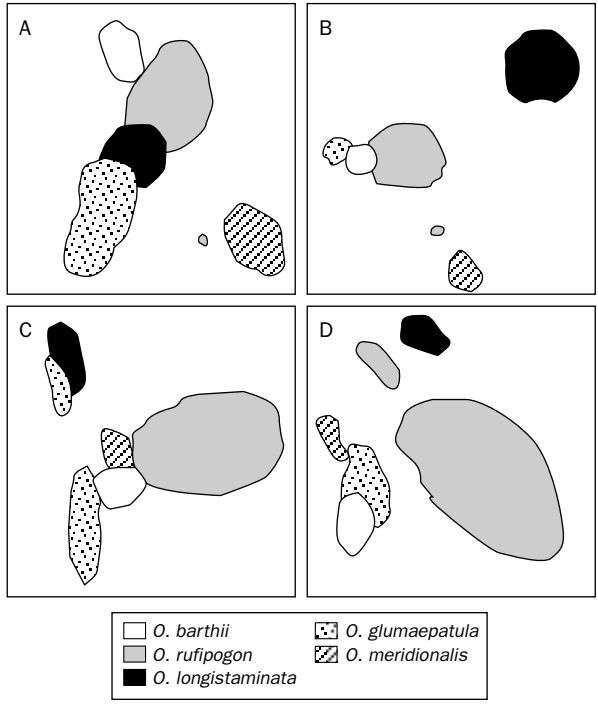


Fig. 1. Schematic diagrams showing genetic diversity among and within species of AA genome wild taxa based on isozyme (A), nuclear DNA restriction fragment length polymorphism (RFLP) (B), mtDNA RFLP (C), and cpDNA RFLP (D). (Redrawn from Akimoto 1999.)

as perceived from previous studies, *O. meridionalis* (or *O. meridionalis* and *O. longistaminata*) is remote from other species that are distinct but relatively close to each other. *O. glumaepatula* showed closer similarity with *O. longistaminata* in isozyme and with *O. barthii* in RFLP. In organellar markers, species difference was distinct except for *O. glumaepatula*. Strains collected from northern South America together with those from Central America showed similarity with *O. longistaminata* in mtDNA, whereas those from the Amazon basin showed similarity with *O. barthii* in mtDNA as well as in cpDNA. Genetic closeness between *O. glumaepatula* and both African taxa was previously pointed out by Wang ZY et al (1992), Doi et al (1995), and Ge et al (1999) based on a smaller number of accessions. On the other hand, similarity between some accessions of *O. glumaepatula* and *O. rufipogon* was pointed out by many authors, such as Dally and Second (1990), Doi et al (1995), and Juliano et al (1998). In Akimoto's study (1999), 57 *O. glumaepatula* accessions, excluding apparently weedy or introgressed types, did not show similarity with *O. rufipogon*. Yet, *O. glumaepatula* is undoubtedly a heterogeneous species and its phylogenetic status is still problematic.

Interspecific relatedness was also discussed based on the precise analysis of particular genes including mobile elements (rDNA, Sano and Sano 1990, Cordesse et al 1990; SINE, Hirano et al 1994; *CatA*, Iwamoto et al 1999; *Adh*, Ge et al 1999; MITE, Kanazawa et al 2000). The finding of such specific DNA will increase our understanding of the differentiation process at the molecular level. It should be noted, however, that gene lineage thus revealed does not necessarily reflect speciation.

Genetic diversity within taxa

The amount of genetic diversity within a species is important for understanding the evolutionary status of the species. *O. sativa* has lower genetic diversity at the molecular level than *O. rufipogon*. This contrasts with the variation pattern observed at the phenotypic level, in which *O. sativa* is more diverse than *O. rufipogon*. The genetic diversity in *O. glaberrima* is much lower than that in *O. sativa*. This can be accounted for by the difference in breeding system of their respective wild ancestors.

Among wild taxa, *O. rufipogon* generally showed the highest diversity, followed by *O. longistaminata* or *O. glumaepatula* (Table 1). This may be attributable to its variability in perenniality associated with allogamy, which conferred a potential to preserve genetic diversity in the populations (Barbier 1989). Evolution from perennials to annuals and from outbreeders to inbreeders is a general trend found in higher plants. Yet, precise phylogeny in this species complex remains unsolved.

Evolutionary trends in the Asian AA genome gene pool

Rice has two primary gene pools, corresponding to *O. sativa* and *O. glaberrima*, which contain cultivated races and their wild and weedy relatives, respectively. Within the primary gene pool of *O. sativa*, four directions of differentiation are recognized (Fig. 2): (1) differentiation from wild to cultivated types, (2) differentiation from perennial to annual types in wild races, (3) geographical differentiation in wild races, and (4) varietal differentiation toward indica and japonica types.

Table 1. Genetic diversity within species at isozyme and DNA levels. H and H′ stand for average gene diversity and diversity index, respectively. (Modified from Morishima et al 1992.)

Species	Isozyme		nDNA	rDNA	mtDNA	cpDNA		MITE[e]
	H[a]	H[b]	H[b]	H′[c]	H[b]	H[d]	H[b]	H′[f]
O. rufipogon	0.50	0.40	0.43	2.21	0.39	1.60	0.24	1.27
O. longistaminata	0.26	0.19	0.43	2.07	0.19	0	0.12	1.35
O. barthii	0.21	0.15	0.19	0	0.19	0.35	0.05	0
O. glumaepatula	0.28	0.18	0.22	0.90	0.24	0.69	0.05	0.38
O. meridionalis	0.11	0.18	0.18	0	0.16	0.32	0.06	0.64

[a]Recomputed from Second (1985). [b]Akimoto (1999). [c]Sano and Sano (1990). [d]Recomputed from Dally and Second (1990). [e]MITE = miniature inverted-repeat transposable elements. [f]Kanazawa et al (2000).

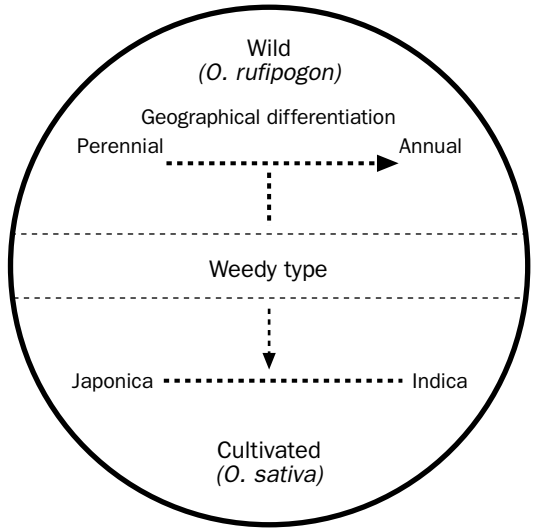


Fig. 2. Differentiation within the primary gene pool of *O. sativa*.

Differentiation from wild to cultivated types: domestication

In seed crops, the cultivated type is characterized by nonshedding of seeds, rapid and uniform germination, efficient seed production, and determinate growth in comparison with the wild type. At the incipient stage of domestication, planting harvested seeds by man automatically selected this "adaptive syndrome of domestication" (Harlan 1975). This holds true in rice.

Oryza sativa and *O. rufipogon* are genetically very close in spite of their clear phenotypic difference, and barely distinguishable by molecular markers. Wild and cultivated plants easily interbreed if grown nearby. Gene flow is mainly from predominantly inbreeding cultivated races to partially outbreeding wild races (outcrossing rate ranging from 10% to 60%, Oka 1988). Gene flow might have played an important role in diversification of the domesticates as in many other crops. Even now, natural hybridization between wild and cultivated rice occurs frequently and hybrid derivatives are found abundantly as weed types. We rarely find truly wild populations without introgression of genes from cultivated rice in tropical rice-growing areas.

Differentiation from the perennial to annual type in wild races

Perennial and annual types exhibit contrasting life-history traits that characterize fecundity/survival schedules of the individuals. The perennial (polycarpic) type shows vigorous vegetative growth, low seed productivity, late flowering, and a high outcrossing rate, whereas the annual (monocarpic) type shows the opposite characteristics (Oka 1988, Sano and Morishima 1982).

Within the geographical range of *O. rufipogon* spreading in Asia and Oceania, distribution of the truly wild annual populations is restricted to tropical continental Asia. In this area, perennial and annual populations are allopatric because of their different habitats. The perennial populations prefer habitats characterized by deeper water and less disturbance. The annual populations are in temporary swamps that are parched in the dry season. These two types reflect differentiation of adaptive strategy resulting from natural selection of habitat conditions (Sano and Morishima 1982, Morishima et al 1984).

Perennial and annual types are phenotypically different but genetically close and no reproductive barrier exists. Life-history traits characterizing the two types vary continuously in nature and segregate continuously in an F_2 population, suggesting that those traits are controlled by multiple factors. Few markers tend to distribute differentially between perennial and annual populations. Polymorphism at some isozyme loci such as *Pox1* and others (Morishima 1991), the presence or absence of the open reading frame (ORF) 100 region of cpDNA (Chen et al 1993), and the presence or absence of a locus of miniature inverted-repeat transposable elements—MITE (Kanazawa et al 2000)—were found to associate with perennial versus annual differentiation to some extent.

Geographical variation in wild races

In phenotypic characters, geographical variation is not so distinct in *O. rufipogon*. Polymorphism at molecular markers, however, revealed a trend of geographical differentiation. Our isozyme study demonstrated (Fig. 3) that the strains collected in

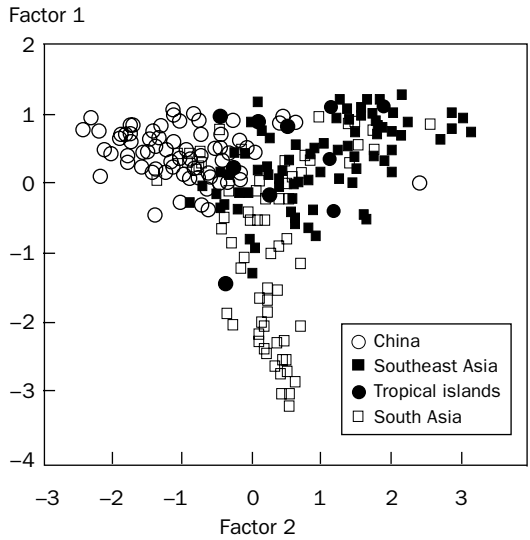


Fig. 3. Scatter diagram of *O. rufipogon* accessions plotted by the first and second scores of factor analysis based on 29 polymorphic isozymes. (Cai and Morishima, unpublished.)

South Asia (particularly on the west coast of India), Southeast Asia (including the east coast of India), and China tend to differentiate (Cai and Morishima 1996). Recently, similar geographical differentiation was observed in organellar markers (Akimoto 1999). The geographical differentiation in *O. rufipogon* is accounted for by "isolation by distance" and adaptation to local environments that occurred during the expansion of this species in Asia and Oceania. A reproductive isolation barrier does not seem to develop within *O. rufipogon*, though a more extensive survey is needed.

Indica versus japonica differentiation in cultivated races

The indica-japonica problem has been argued repeatedly as reviewed by Oka (1988). Various molecular studies consistently showed a distinct difference between these two major varietal groups in nuclear (Second 1982, Glaszmann 1987, Wang and Tanksley 1989, Nakano et al 1992) and organellar genomes (Kadowaki et al 1988, Dally and Second 1990, Second and Wang 1992, Ishii et al 1993, Chen et al 1993). When typical indica and japonica types were compared, many genes and characters showed nonrandom association with each other to clearly separate the two types.

When a large number of primitive cultivars are analyzed without a priori criteria, indica versus japonica differentiation is the principal variation, but some varieties that do not belong to either of the two types are also found. Such atypical cultivars are not necessarily intermediate on the axis distinguishing indica and japonica types or a recombined type, but seem to vary on a different variation axis. Himalayan hilly areas are known as the homeland of the atypical varieties (Sano and Morishima 1992). Further, some cultivars grown in low-lying deepwater areas in the Bengal delta are also atypical types (Glaszmann 1987, Hakoda et al 1990). A recent study by Cai and Morishima (2000a) demonstrated that primitive cultivars collected in deepwater areas in Bangladesh included a japonica-like group and an atypical group, which did not belong to either type (Table 2). It was further suggested that geographical differentiation could precede seasonal ecotype differentiation (aman, aus, and boro).

Table 2. Classification of indigenous cultivars collected in Bangladesh deepwater areas. Numbers in parentheses show the number of cultivars collected in Khulna District. (From Cai and Morishima 2000a.)

Isozyme cluster[a]	Seasonal ecotype					
	Boro	Aus	T. aman	B. aman	Rayada	Ashina
1 (japonica)	1		3			
2	1		1	1		2(2)
3		1(1)		10(4)	4(4)	
4	3(3)	1(1)				1(1)
5	1	5(2)				
6 (indica)	1	2	8	22		

[a]Based on cluster analysis of 8 polymorphic isozymes.

Search for the immediate ancestor of the domesticates

The four directions of differentiation mentioned above are conceptually independent from each other. Yet, they have proceeded, probably interacting with each other in the process of domestication. The evolutionary role of perenniality, geographical variation, and indica versus japonica differentiation will be discussed below in relation to the domestication process.

Which is the immediate ancestor of domesticates, the perennial or annual type? It has been a subject of discussion whether the perennial or annual type is the ancestor of *O. sativa*. Sano et al (1980) inferred that the perennial-annual intermediate type is most probably the immediate ancestor. "Intermediate type" implies the population that is habitually clonal and partially outbreeding but that can propagate sexually if seed propagation is advantageous. Such populations are now mostly the secondary products of natural hybridization between perennial wild and cultivated plants. When the primitive perennial population was exposed to disturbed or dry conditions, the population genotype probably shifted toward wild annuals in natural habitats or a primitive cultivated type in man-made habitats through the "intermediate type." A trade-off which constrains energy allocation into sexual and asexual reproduction might have moved the plants toward higher seed yielders.

Which geographical races of wild rice evolved into incipient domesticates? Older rice remains excavated to date are concentrated in the middle and lower basin of the Yangtze River in China. Though *O. rufipogon* is not distributed at present in this area, some evidence suggests the existence of wild rice in the past (Sato et al 1991). It seems difficult, however, to determine the exact place of origin of cultivated rice until more archaeological evidence, particularly on wild rice, will accumulate.

For genetic relationships between particular wild and cultivated rice, the following subjects are worth considering. First, Chinese wild rice strains, in particular those collected in the northern fringe of distribution of *O. rufipogon*, have some japonica-specific genes and characters (Second 1985, Sano et al 1989, Morishima and Gadrinab 1987, Nakano et al 1992, Cai et al 1995). Second, some indigenous cultivars grown in deepwater areas in Bangladesh carry particular isozyme alleles such as *Est10-4* and *Amp5-4*, which are rare in cultivars but not rare in wild populations (Cai and Morishima 2000a). Third, the annual wild rice distributed on the west coast of India and primitive cultivars growing nearby showed high genetic similarity in isozymes, and both showed similarity with the japonica type in cpDNA as well as in rDNA according to Lolo and Second (1988).

Are indica and japonica types monophyletic or diphyletic? Oka and his group considered that indica and japonica types have diverged as domestication proceeded. This view is mainly based on the fact that wild rice has a potential to evolve into indica as well as japonica types (Oka and Morishima 1982) and that indica versus japonica differentiation was not found in *O. rufipogon*. A recent collection of *O. rufipogon* obtained from a broader geographical range, however, yielded a slightly different variation pattern. A tendency toward indica-japonica differentiation in terms of particular association among genes or characters was observed among wild strains (Second 1985, Morishima and Gadrinab 1987, Dally and Second 1990, Nakano et

al 1992, Sun et al 1996a,b,c), though the degree of nonrandom association is much lower than in cultivars.

The diphyletic hypothesis postulates that indica and japonica types originated from different lineages of *O. rufipogon* (Second 1982, Sato 1996). Accumulated observations that indica and japonica types consistently showed a clear difference between each other and closer affinity with different wild accessions than with each other seem to support the diphyletic hypothesis. Recent archaeological excavations in China and analysis of rice remains (phytolith, Wang et al 1998; DNA, Sato et al 1995) suggest that Chinese wild rice played an important role in the origin of japonica rice.

I am inclined to the view that rice domestication has been a diffused process in both space and time. During a long period, prototypes of indica and japonica types have probably become two dominant groups and have accumulated marked differences in morphological and adaptive traits, keeping their respective genic constitution inherited from their founders.

Genetic basis of the domestication syndrome

Multilocus system in evolution

Related species or ecotypes are differentiated by a particular pattern of association between states of different characters or between alleles at different loci. Such multilocus covariation is called gametic disequilibrium. In rice, variations between wild and cultivated types, between perennial and annual types, and between indica and japonica types are good examples of gametic disequilibrium. Nonrandom association is caused by various factors such as selection (coadaptation), linkage, pleiotropy, and founder effect, as discussed by Hedrick et al (1978). It is difficult to dissect the underlying mechanism without elaborate experiments.

In rice, linkage blocks harboring genes for internal barriers and fitness traits were detected by isolating the relevant chromosomal segments using the backcross method (Sano 1992). Key factors for gametic disequilibrium in quantitative traits can be elucidated to some extent by observing the shift of correlations in the hybrid generations. Parental associations that disappear in the F_2 are mostly due to selection for coadapted traits. On the other hand, those that remain in the F_2 or early generations are probably due to linkage and/or pleiotropy. Many studies showed that parental associations partly persisted in the F_2 (character coherence) though they were weak but significant. In the crosses of wild $\times$ cultivated, annual $\times$ perennial, and indica $\times$ japonica rice strains, we have experienced that parental associations in quantitative traits mostly disappeared in the F_2. This indicates that those associations are not due to linkage or pleiotropy of a few major genes.

QTL clusters responsible for character associations

Advances in molecular genetics enabled us to dissect the genetic basis of quantitative differences between species or ecotypes, which has been studied only by statistical-genetics methods until recently. Targeting differentiation in the Asian AA genome

gene pool, we performed QTL analysis in the cross *O. sativa* (indica) × *O. rufipogon* (perennial type) (Cai and Morishima, n.d.). Based on 125 recombinant inbred lines (RILs), 148 markers were mapped. Because this wild parent carried some japonica-specific characteristics, indica-japonica diagnostic traits segregated in this mapping population in addition to domestication-related traits. Among 34 quantitative traits examined, 22 traits revealed 1–29 putative QTLs, respectively. Several QTLs for domestication-related traits were mapped over the whole genome, showing a tendency to form clusters, each reflecting the domestication syndrome. The same situation was observed in QTLs for indica versus japonica diagnostic traits. In addition, some of the clusters joined the domestication syndrome clusters. Figure 4 presents examples of gene clusters. Several isozyme loci that serve as diagnostic markers to distinguish

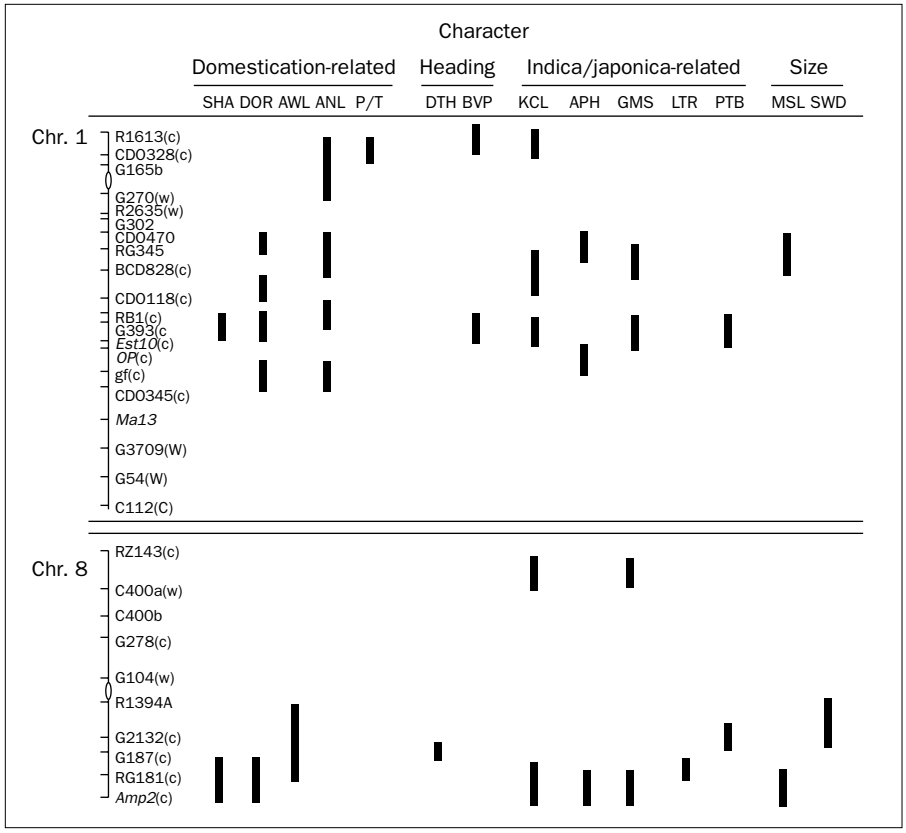


Fig. 4. Examples of QTL clusters (chromosomes 1 and 8). Genomic regions affecting the respective traits (LOD>2.8) are shown by the bars to the right of linkage groups. Character code: SHA = seed shattering, DOR = seed dormancy, AWL = awn length, ANL = anther length, P/T = panicle number/tiller number, DTH = days to heading, BVP = basic vegetative phase, KCL = $KClO_3$ resistance, APH = apiculus hair length, GMS = germination speed, LTR = low-temperature resistance, PTB = panicle neck to lowest branch, MSL = mesocotyl length, SWD = spikelet width. The small spindle on the chromosome shows the approximate position of the centromere. (Adapted from Cai and Morishima, n.d.)

wild and cultivated types (*Est10*, Wang XK et al 1992) and indica and japonica types (such as *Cat1* and *Amp2*) were found in or near the QTL clusters. Similar cluster phenomena of domestication-related QTLs were reported by Koinange et al (1996) in common bean and by Xiong et al (1999) in rice.

To determine whether QTL clusters are loosely linked loci or a single locus with a pleiotropic effect, more precise analysis is necessary. It is interesting to note that QTLs for mesocotyl length, which is considered to reflect endogenous hormone level, tended to join those clusters. This might suggest a possibility that the clustering phenomenon is partly due to the pleiotropic effect of an unknown key factor controlling various traits through hormonal regulation.

In this *O. sativa* × *O. rufipogon* cross, character correlations were generally weak among RILs that represent essentially F_2 variation. This implies that nonrandom character associations of the domestication syndrome are mainly due to natural selection for coadapted traits. The present analysis revealed several QTL linkages reflecting the domestication syndrome. Since they were located on different chromosomes, they segregated into RILs, resulting in decreased correlations. This phenomenon could be understood by "multifactorial linkages" advocated by Grant (1975). This phenomenon is inevitably caused by random distribution of multiple factors over the limited number of chromosomes that determine the differences in two or more quantitative characters.

QTL clusters thus identified were mapped on the regions in which cultivar-derived alleles segregated at higher frequencies than expected. This coincidence could be explained by unconscious selection, which worked under cultivation pressure when establishing RILs favoring the combination of cultivar-derived gene blocks.

Inference for domestication dynamics

The domestication process is undoubtedly a gradual process directed by humans. In rice, Oka and Morishima (1971) demonstrated early that "cultivation pressure" brought about a rapid change in population genotype toward the cultivated type as suggested by Harlan (1975). Seed shedding is the critical trait disruptively selected in natural and cultivated fields. The genetic basis of seed shedding is simpler than other domestication-related traits. Eiguchi and Sano (1990) identified two loci conferring high seed shedding on *O. rufipogon*, one of which was linked to the spreading panicle locus on chromosome 4. Our QTL analysis detected five loci for seed shedding and one of them located on chromosome 1 was linked with another spreading panicle locus (Cai and Morishima 2000b). A spreading panicle is advantageous for wild rice to disperse seeds but disadvantageous for cultivated rice. It may be reasonable to infer that selection for nonshedding genes with relatively large effects played an important role as a trigger for domestication.

Our experiments with another cross (*O. sativa* × *O. rufipogon*) showed that association among seed shedding, seed dormancy, and awn length tended to increase in later hybrid generations without deliberate selection (Table 3). In the cross propagated in bulk, the outcrossing rate was higher in the lines with wild characteristics than in those with cultivated-rice characters (Table 3). This suggests that flower char-

Table 3. Changes in population parameters in cultivated × wild hybrids.

Cross	Characters	F_3	F_6
1. Correlations between domestication-related traits. Breeding in bulk, seeds from ground and panicles, no deliberate selection (Morishima, unpublished).			
Taichung 65 × W120	Seed shedding-awn length	0.01	0.34**
	Seed shedding-seed dormancy	0.06	0.29**
	Seed dormancy-awn length	0.05	0.17*

Cross	Characters	F_3	F_7	
			Wild type	Cultivated type
2. % outcrossing rates estimated from the ratio of within- and between-line genetic variance components. Breeding in bulk without selection, F_6 lines (F_7) were classified into wild and cultivated types. (Recomputed from Oka and Morishima 1971.)				
Taichung 65 × W152	Spikelet length	40	39	13
	Spikelet width	42	32	17
Pei-khu × W152	Spikelet length	38	28	2
	Spikelet width	42	29	11

**, * = significant at the 1% and 5% levels, respectively.

acteristics to increase the selfing rate would be selected by linkage drag. Our QTL analysis indicated that loci governing those traits, including anther length, which is known to positively correlate with outcrossing rate, are linked with each other (Fig. 4).

Pernes (1983) predicted that linkages of domestication-related traits could be adaptive, particularly in the outbreeding species, in which recovery of the cultivated type following frequent outcrossing between wild and cultivated types might be facilitated by such linkages. In rice, wild rice is partially outbreeding though current cultivars are predominantly inbreeding, and natural hybridization has played a significant role in rice evolution during and after domestication. The domestication syndrome could be modified by unconscious and conscious selection of adaptive gene blocks distributed over the genome.

References

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS. 1999. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor. Appl. Genet. 98:1320-1328.

Akimoto M. 1999. Bio-systematics of the *Oryza* AA genome wild taxa (*Oryza sativa* complex) by morpho-physiological traits, isozymes and RFLPs. PhD thesis. Hokkaido University, Japan.

Barbier P. 1989. Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. Jpn. J. Genet. 64:273-285.

Cai HW, Morishima H. 2000a. Diversity of rice varieties and cropping system in Bangladesh deepwater areas. JARQ 34(4):225-231.
Cai HW, Morishima H. 2000b. Genomic regions affecting seed shattering and seed dormancy in rice. Theor. Appl. Genet. 100:840-846.
Cai HW, Morishima H. n.d. QTL clusters reflect character associations in wild and cultivated rice. Theor. Appl. Genet. (Submitted.)
Cai HW, Wang XK, Morishima H. 1995. Isozyme variation in Asian common wild rice, *Oryza rufipogon*. Rice Genet. Newsl. 12:178-180.
Cai HW, Morishima H. 1996. Geographical differentiation found in annual types of *O. rufipogon*. Annu. Rep. Natl. Inst. Genet. Jpn. 47:99-100.
Chen WB, Nakamura I, Sato YI, Nakai H. 1993. Distribution of deletion type in cpDNA of cultivated and wild rice. Jpn. J. Genet. 68:597-603.
Cordesse F, Second G, Delseny M. 1990. Ribosomal gene spacer length variability in cultivated and wild rice species. Theor. Appl. Genet. 79:81-88.
Dally AM, Second G. 1990. Chloroplast DNA diversity in wild and cultivated species of rice (genus *Oryza*, section *Oryza*): cladistic-mutation and genetic-distance analysis. Theor. Appl. Genet. 80:209-222.
Doi K, Yoshimura A, Nakano N, Iwata N. 1995. Phylogenetic study of A-genome species of genus *Oryza* using nuclear RFLP. Rice Genet. Newsl. 12:160-162.
Eiguchi M, Sano Y. 1990. A gene complex responsible for seed shattering and panicle spreading found in common wild rice. Rice Genet. Newsl. 7:105-107.
Glaszmann JC. 1987. Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. 74:21-30.
Ge S, Sang T, Lu BR, Hong DY. 1999. Phylogeny of rice genomes with emphasis on origin of allotetraploid species. Proc. Natl. Acad. Sci. USA 96:14400-14405.
Grant V. 1975. Genetics of flowering plants. New York (USA): Columbia University Press. 563 p.
Hakoda H, Inouye J, Morishima H. 1990. Isozyme diversity found among Asian deepwater rices. Rice Genet. Newsl. 7:91-93.
Harlan JR. 1975. Crops and man. Madison, Wis. (USA): American Society of Agronomy. 295 p.
Hedrick PW, Jain SK, Holden L. 1978. Multilocus systems in evolution. Evol. Biol. 11:101-184.
Hirano HY, Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E, Sano Y. 1994. Retrotransposition of a plant SINE into the *wx* locus during evolution of rice. J. Mol. Evol. 38:132-137.
Ishii T, Terauchi T, Tsunewaki K. 1988. Restriction endonuclease analysis of chloroplast DNA from A-genome diploid species of rice. Jpn. J. Genet. 63:523-536.
Ishii T, Terauchi T, Mori N, Tsunewaki K. 1993. Comparative study on chloroplast, mitochondrial and nuclear genome differentiation in two cultivated rice species, *O. sativa* and *O. glaberrima*, by RFLP analysis. Theor. Appl. Genet. 86:88-96.
Ishii T, Nakano T, Maeda H, Kamijima O. 1996. Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. Genes Genet. Syst. 71:195-201.
Iwamoto M, Nagashima H, Nagamine T, Higo H, Higo K. 1999. *p-SINE1*-like intron of the *CatA* catalase homologs and phylogenetic relationships among AA-genome *Oryza* and related species. Theor. Appl. Genet. 98:853-886.
Juliano AB, Naredo MEB, Jackson MT. 1998. Taxonomic status of *Oryza glumaepatula* Steud. I. Comparative morphological studies of New World diploids and Asian AA genome species. Genet. Res. Crop Evol. 45:197-203.

Kadowaki K, Yazaki K, Osumi T, Harada K, Katsuta M, Nakagahra M. 1988. Distribution of mitochondrial plasmid-like DNA in cultivated rice (*Oryza sativa* L.) and its relationship with varietal groups. Theor. Appl. Genet. 76:809-814.

Kanazawa A, Akimoto A, Morishima H, Shimamoto Y. 2000. Inter- and intra-specific distribution of *Stowaway* transposable elements in AA genome of wild rice. Theor. Appl. Genet. 101:327-335.

Koinange EMK, Singh SO, Gepts P. 1996. Genetic control of the domestication syndrome in the common bean. Crop Sci. 36:1037-1045.

Lolo OM, Second G. 1988. Peculiar genetic characteristics of *O. rufipogon* from Western India. Rice Genet. Newsl. 5:67-70.

Morishima H. 1991. Association between *Pox-1* variation and seed productivity potential in wild rice. In: Rice genetics II. Manila (Philippines): International Rice Research Institute. p 1-9.

Morishima H, Sano Y, Oka HI. 1984. Differentiation of perennial and annual populations of the Asian wild rice *Oryza perennis*. Plant Syst. Evol. 144:119-135.

Morishima H, Gadrinab LU. 1987. Are the Asian common wild rices differentiated into the Indica and Japonica types? In: Hsieh SC, editor. Crop exploration and utilization of genetic resources. Proceedings of the Conference. Changhua (Taiwan): Taichung District Agricultural Improvement Station. p 11-20.

Morishima H, Sano Y, Oka HI. 1992. Evolutionary studies in cultivated rice and its wild relatives. Oxford Surv. Evol. Biol. 8:135-184.

Nakano M, Yoshimura A, Iwata N. 1992. Phylogenetic study of cultivated rice and its wild relatives by RFLP. Rice Genet. Newsl. 9:132-134.

Oka HI. 1988. Origin of cultivated rice. Tokyo (Japan): Japanese Science Society Press. 254 p.

Oka HI, Morishima H. 1971. The dynamics of plant domestication: cultivation experiments with *Oryza perennis* and its hybrid with *O. sativa*. Evolution 25:356-364.

Oka HI, Morishima H. 1982. Phylogenetic differentiation of cultivated rice. 23. Potentiality of wild progenitors to evolve the Indica and Japonica types of rice cultivars. Euphytica 31:41-50.

Pernes J. 1983. The genetics of domestication in cereals. La Rechereche 14:910-919. (In French.)

Sano R, Morishima H. 1992. Indica-Japonica differentiation of rice cultivars viewed from the variation in key characters and isozymes with special reference to land races from the Himalayan hilly areas. Theor. Appl. Genet. 84:266-274.

Sano Y. 1992. Genetic comparisons of chromosome 6 between wild and cultivated rice. Jpn. J. Breed. 42:561-572.

Sano Y, Morishima H, Oka HI. 1980. Intermediate perennial-annual populations of *Oryza perennis* found in Thailand and their evolutionary significance. Bot. Mag. Tokyo 93:291-305.

Sano Y, Morishima H. 1982. Variation in resource allocation and adaptive strategy of a wild rice, *Oryza perennis* Moench. Bot. Gaz. 143:518-523.

Sano Y, Yi HX, Shao QQ, Iyama S. 1989. Ribosomal DNA spacer-length variations in a wild rice population from Dongxiang, China. In: Breeding research: the key to the survival of the earth. Proceedings of the 6th International SABRAO Congress, Tokyo, Japan. p 493-496.

Sano Y, Sano R. 1990. Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. Genome 33:209-218.

Sato YI. 1996. Origin of rice and rice cultivation based on DNA analysis. Tokyo (Japan): NHK Books. 227 p. (In Japanese.)

Sato YI, Tang SX, Yang LU, Tang LH. 1991. Wild rice seeds found in an oldest rice remain. Rice Genet. Newsl. 8:76-78.
Sato YI, Tang LH, Nakamura I. 1995. Amplification of DNA fragments from charred rice grains by polymerase chain reaction. Rice Genet. Newsl. 12:260-261.
Second G. 1982. Origin of the genetic diversity of cultivated rice: study of the polymorphism scored at 40 isozyme loci. Jpn. J. Genet. 57:25-57.
Second G. 1985. Evolutionary relationships in the *Sativa* group of *Oryza* based on isozyme data. Genet. Sel. Evol. 17:89-114.
Second G. 1991. Molecular markers in rice systematics and the evaluation of genetic resources. In: Bajaj YPS, editor. Biotechnology in agriculture and forestry. Vol 14. Rice. New York (USA): Springer. p 468-494.
Second G, Wang ZY. 1992. Mitochondrial DNA RFLP in genus *Oryza* and cultivated rice. Genet. Res. Crop Evol. 39:125-140.
Sun C, Wang XK, Yoshimura A, Iwata N. 1996a. Genetic differentiation of nuclear genome in common wild rice and cultivated rice. In: Wang XK, Sun C, editors. Origin and differentiation of Chinese cultivated rice. Beijing (China): China Agricultural University Press. p 120-133. (In Chinese with English summary.)
Sun C, Wang XK, Yoshimura A, Iwata N. 1996b. Genetic differentiation of mitochondrial DNA in common wild rice and cultivated rice. In: Wang XK, Sun C, editors. Origin and differentiation of Chinese cultivated rice. Beijing (China): China Agricultural University Press. p 134-139. (In Chinese with English summary.)
Sun C, Yoshimura A, Zichao L, Haiying Z, Wang XK, Iwata N. 1996c. Genetic differentiation of chloroplast genome in *O. rufipogon* Griff. and *O. sativa* L. In: Wang XK, Sun C, editors. Origin and differentiation of Chinese cultivated rice. Beijing (China): China Agricultural University Press. p 140-146. (In Chinese with English summary.)
Wang C, Udatsu T, Tang L, Zhou J, Zheng Y, Sasaki A, Yanagisawa K, Fujiwara H. 1998. Cultivar group of rice cultivated in Caoxieshan site (B.P. 6000-present) determined by the morphology of plant opals and its historical change. Breed. Sci. 48:387-394. (In Japanese with English summary.)
Wang XK, Cai HW, Cheng KS. 1992. The discovery of an *Est* locus related to the origin, evolution and classification of Asian rice. Chin. Rice Res. Newsl. 7:1-2.
Wang ZY, Tanksley SD. 1989. Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32:1113-1118.
Wang ZY, Second G, Tanksley SD. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. 83:565-581.
Xiong LZ, Liu KD, Dai KK, Xu CG, Zhang Q. 1999. Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theor. Appl. Genet. 98:243-251.

Notes

Author's address: Formerly National Institute of Genetics, Mishima, 411-8540 Japan; mailing address: H. Morishima: Saiwai-cho 15-2, Hiratsuka, 254-0804 Japan.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Rice: a central genome for the genetics of all cereals *

M. Gale, G. Moore, and K. Devos

The past 15 years have seen an intense research drive to apply the new molecular biology to rice. Initiatives such as the Rockefeller Foundation's International Program on Rice Biotechnology, begun in the mid-1980s, have been underpinned by the in-depth corporate knowledge of the crop built up by research organizations in Southeast Asia such as the International Rice Research Institute and accelerated by the application of a vast, and previously uncoordinated, research capacity in national programs in the area. Moreover, unlike the other two 500-million-ton crops, wheat and maize, rice has a small tractable genome, and the development of genetic and genomic tools not available in any other cereal has ensured the promotion of rice as a favored research target. On top of all this, the discovery that gene content and gene order—genome colinearity—have been maintained over the whole grass family, which includes all cereals and many forage crops, has elevated rice still further to the status of a "model" organism. The initiation of genomic DNA sequencing efforts in the public and private sector will further ensure rice's central position in plant science.

In this chapter, we will describe the ways in which rice genomic tools and knowledge of the rice genome are already being applied in research on the other major cereals, wheat and maize. Moreover, many aspects of rice genetics can be transferred to the "orphan" crops, the several minor economic grass species that have not themselves warranted extensive research and breeding.

* This is a modified version of Gale et al (2001): "Rice—the pivotal genome in cereal comparative genomics," in *Rice Biotechnology: Improving Yield, Stress Tolerance and Grain Quality*, published by John Wiley & Sons Ltd, Chichester, West Sussex, England, p 46-52; copyright© Novartis Foundation 2001; with permission from the publisher.

The solid base laid by classical rice geneticists and breeders around the world over the past century has allowed successful launches of several major research initiatives. One, the Rockefeller Foundation's International Program on Rice Biotechnology (Toenniessen 1998), was particularly farsighted and started in 1984. This was followed by national programs such as the Korean effort centered in Suwon and the Japanese Rice Genome Program based at Tsukuba. Rice is an important staple food crop and it has a much smaller genome than many other cereals—11 times smaller than that of maize and 40 times smaller than that of wheat. As a result, rice has rapidly become the most-researched cereal. Denser molecular maps and more DNA resources, including random fragment length polymorphism (RFLP) probes, microsatellite markers, expressed sequence tags (ESTs), bacterial and yeast artificial chromosome (BAC and YAC) libraries, and the genomic sequence, are available in rice than in any other cereal. Because considerable sequencing of rice has already taken place in Monsanto and Syngenta and the public international rice genomic sequencing effort is well under way will only increase the pace and central importance of rice genetics research.

Cereal comparative genetics

Increasingly, rice is being studied in laboratories that previously concentrated on only wheat or maize. The single major factor that has led to the promotion and use of rice as a model has been the discovery of an unexpectedly high degree of similarity between different cereal genomes in terms of gene content and gene order. It is now being recognized that conserved colinearity is the rule rather than the exception and that it exists to a greater or lesser extent in all groups of plants, animals, and microbes. However, the critical observations were first made in the cereals in the late 1980s. Early reports included the comparisons of the three genomes of hexaploid bread wheat (Chao et al 1989), comparisons of the genomes of wheat and rye (Devos et al 1993), and comparative maps of rice and maize (Ahn and Tanksley 1993) and rice and wheat (Kurata et al 1994, Van Deynze et al 1995b). We now understand that the relationships between various cereal genomes are so close that all of the genomes investigated to date can be shown as a single synthesis, irrespective of chromosome number or total DNA content (Fig. 1).

Other grasses of lesser global economic significance have later been shown to display similar close synteny with rice. These comparative studies include sorghum (Dufour et al 1997), sugarcane (Glaszmann et al 1997), foxtail millet (Devos et al 1998), oats (Van Deynze et al 1995a), pearl millet (Devos et al 2000), and *Zizania* (Kennard et al 1999). Others that will follow over the next year or so include the forage grasses, fescue (O.-A. Rognli, personal communication) and ryegrass (J. Foster, personal communication), finger millet (M.M. Dida and K.M. Devos, personal communication), and tef (M. Sorrells, personal communication). There will be differences of evolutionary origin in the detail, but overall it is clear that, in the end, all grass genomes will be described by only a few rice linkage blocks (Gale and Devos 1998).

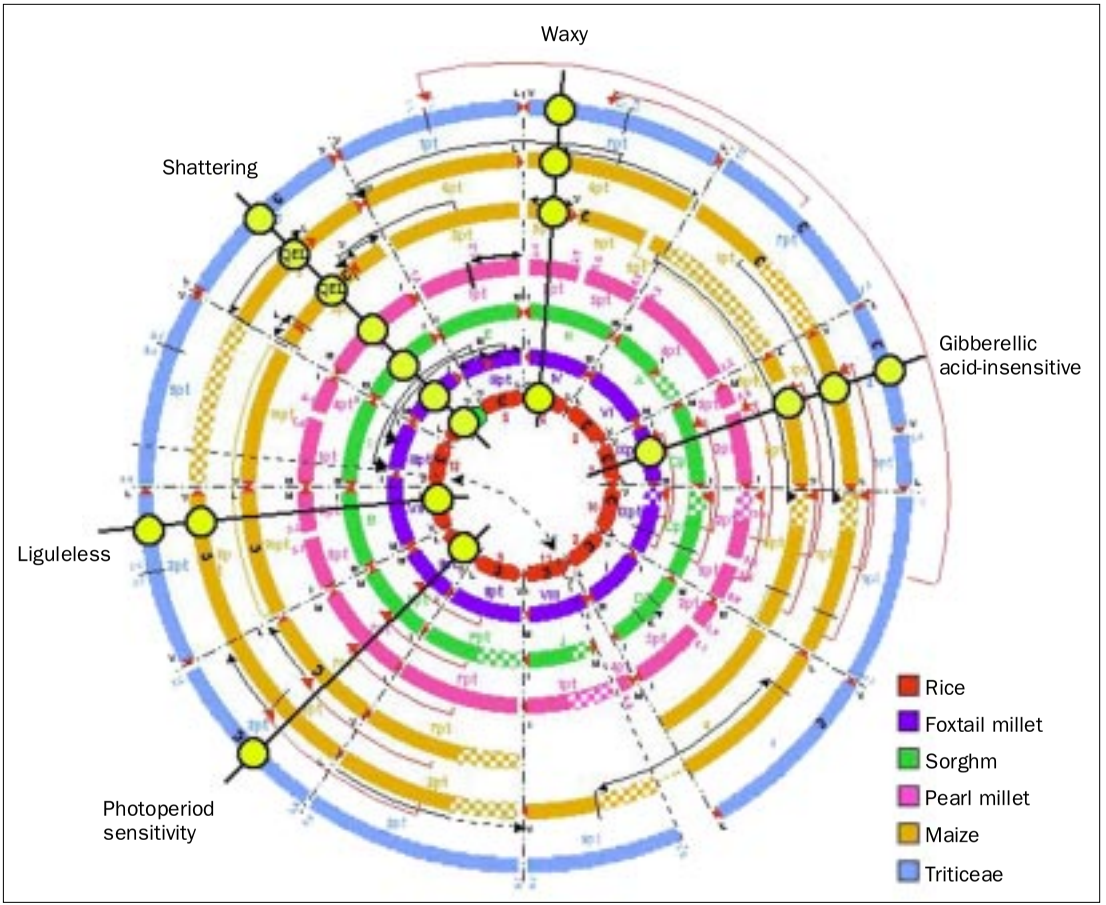


Fig. 1. Comparative maps of several of the world's cereal crop species genomes. Note: The 12 chromosomes of rice, ten chromosomes of maize, and the basic seven chromosomes of the Triticeae, together with the genomes of pearl millet, sorghum, and foxtail millet, are aligned so that homoeologous (orthologous) loci lie on the same radius. The arrangement is simply the most parsimonious, requiring the fewest number of adjustments from the present-day chromosome linear gene orders, and is not intended to represent any genome as the most primitive. Evolutionary translocations and inversions, relative to rice, that have to be invoked are shown by the arrows. Chromosomal regions where the syntenic relationships are still not clear are hatched (▩). This is an extension of the diagram first shown in Moore et al (1995) and modified from Gale and Devos (1998). For the additional detail of liguleless, waxy loci, etc., see text. B and T represent telomeres of the bottom and top of the chromosome arms; S and L represent short and long arms of the chromosome; pt = part. (Reprinted from Gale et al, 2001, with permission from the publisher.)

The synthesis shown in Figure 1 is derived from comparisons of genetic maps, particularly RFLP marker data. It is now possible to explore these relationships at the megabase and even DNA sequence levels. Initial results indicate that gene order is generally conserved but that small local duplications and other rearrangements are also common (Bennetzen 2000). Comparisons of the *sh2/a1* regions of rice and sorghum (Chen et al 1998) showed strong conservation of gene content and gene order. However, analysis of maize, sorghum, and rice *Adh* regions indicated considerable gene rearrangements (Tikhonov et al 1999, Tarchini et al 2000). In this latter study, sorghum and maize share nine genes in a colinear order but a further three genes appear to have been deleted from the maize genome. Comparisons of the region of wheat chromosome 5B containing the *Ph1* gene with the homoeologous region of rice chromosome 9 (Foote et al 1997) demonstrated good colinearity over a 30-Mb region of rice but also showed evidence for the duplication of one region containing three markers to a location some 10 cM distant on the same chromosome. Feuillet and Keller (1999) showed commonalities between the regions of wheat chromosome 3, barley chromosome 3H, rice chromosome 1, and maize chromosome 8 containing homoeologues of receptor-like kinases *Lrk10* and *Tak10.* Nevertheless, some duplications were observed and, in wheat and barley, the whole region was found duplicated on the short arms of the group 1 chromosomes. Interestingly, a lot of the evidence for rearrangement at the megabase level appears to be associated with disease resistance genes and their analogues. It is possible that these genes in particular are prone to more rapid evolution than most genes for adaptation, since they have to respond to pathogens that evolve rapidly themselves.

Poaceae evolution

The differences in segmental chromosome organization of the various economic grass crop species, expressed relative to the present-day rice genome, can be used to track evolutionary relationships. This approach exposed maize as an almost complete tetraploid (Fig. 1) very early on. Other major rearrangements, dating back 60 million years, can be identified that define two major subfamilies, the Pooideae, as exemplified by oats and wheat (and more recently the forage grasses, O.-A. Rognli, personal communication), and the Panicoideae, as exemplified by pearl and foxtail millet, sorghum, maize, and sugarcane. These translocations are shown in red in Figure 1. For example, in the Pooideae, insertion of present-day rice chromosome 10 (R10) into R5 represents the structural organization of the Triticeae group 1 chromosomes. The same situation exists for oat chromosome A (Van Deynze et al 1995a). Interestingly, in both wheat and oats, the region corresponding to rice chromosome 10, which displays normal recombination with a map length of 150 cM in rice, is highly compressed to just a few map units in the low recombinogenic centromeric regions of the cereals with larger genomes.

A question of interest to cereal taxonomists that will probably be hotly debated for many years to come is the nature of the primeval grass genome. One small step toward resolving the structure of the ancestral genome may be provided by the com-

parative organization of the Pooideae and Panicoideae chromosomes. The involvement of R10 in two different insertion events would suggest that an independent R10, as in rice itself, is the more primitive configuration.

Although it is tempting to suggest that the extent of evolutionary chromosomal rearrangements will be correlated with evolutionary time (Paterson et al 1996), it is becoming clear that the different chromosome structures can become fixed at varying rates in different species. For example, the rearrangements between relatively closely related Triticeae genomes that have been isolated by speciation for less than 10 million years, such as wheat, rye (Devos et al 1993), or *Aegilops umbellulata* (Zhang et al 1998), may be almost as extensive as those between wheat and rice. Almost no rearrangements are apparent between wheat and barley, however, which are more distantly related than wheat and rye (Fig. 2).

Segmental chromosome duplication is also becoming recognized as being more common than we originally thought. Extensive duplication will be a hindrance to many applications, such as in cross-genome map-based cloning in which "walks" could get deflected by inadvertent jumping from one segment to its duplicate. In addition, duplicated genes will probably be more prone to divergence and deletion, or silencing prior to eventual deletion. These processes will all give rise to apparent local disruptions in colinearity. Information on the extent of duplication emerges slowly where it depends on genetic mapping of duplicated pairs of genes; however, where an organism has been completely sequenced, the analysis is much more straightforward.

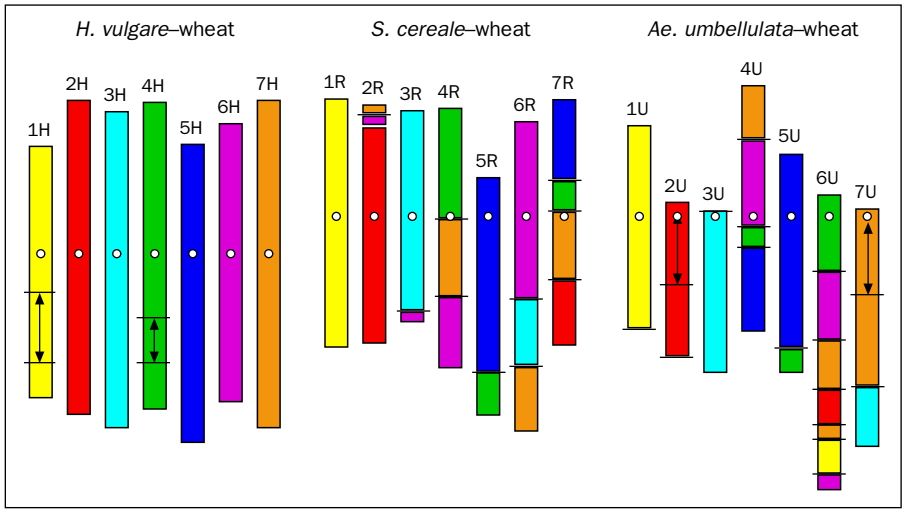


Fig. 2. Chromosome rearrangements within the Triticeae tribe—the rye and *Aegilops umbellulata* genomes relative to wheat. Although the basic chromosome number is constant at 2x=14 throughout the tribe, the juxtaposition of linkage blocks carried on individual chromosomes can be considerably rearranged in some species relative to the cultivated wheat and barley genomes. The extent of rearrangement appears not to be a reflection of evolutionary time or breeding system—*S. cereale* is an outbreeder, whereas *Ae. umbellulata* is an inbreeder. (Reprinted from Gale et al, 2001, with permission from the publisher.)

Arabidopsis, long heralded as a "simple" model diploid, turns out to be more than 60% duplicated (The Arabidopsis Genome Initiative 2000). Rice is almost certainly no exception.

Information concerning ancient large duplications in the grasses is beginning to emerge from analyses at the map level. The R11/R12 duplication, which involves substantial portions of the short arms of both rice chromosomes, was originally thought to be specific to rice (Nagamura et al 1995). However, the same duplication has now been shown to exist in pearl and foxtail millet (Fig. 3). Thus, this duplication clearly predates the divergence of the Bambusoideae and Panicoideae subfamilies. As more duplications are identified, these will be useful predictors for the taxonomists; however, they will also disrupt other applications that rely on close colinearity.

Underresearched cereals—the "orphan" crops

The discovery of general syntenic relationships among groups of plants will be of immense benefit to "orphan" crops. In fact, all initial mapping projects for underresearched minor crops should include a comparative approach at the earliest stage, because, once maps have been aligned, the presence, location, and modes of action of key genes can be predicted from knowledge gained in other cereals. An example is foxtail millet, *Setaria italica*, a crop of agricultural significance only in northern China. Although foxtail millet has received very little research attention in the past, it has recently achieved a considerable genetic base by the simple strategy of having its 2n=2x=18 genome aligned with rice (Devos et al 1998). QTL analyses of an interspecific foxtail millet and *S. viridis* segregating population have revealed the

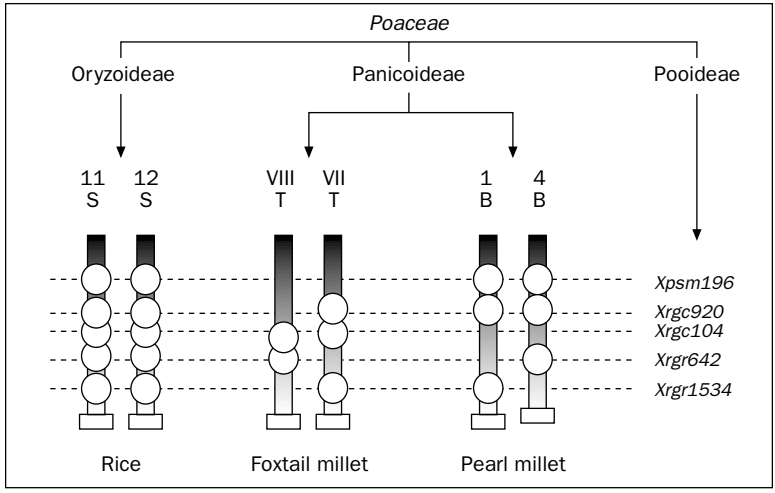


Fig. 3.The R11/R12 duplication in rice, foxtail millet, and pearl millet. Note: The figure shows that RFLP markers that define rice duplication are similarly duplicated and map in the same linear orders in the other two species. T and B represent the top and bottom of the chromosome arms. S is the short arm.

clear alignment of major factors controlling agronomic traits with genes in wheat, maize, or rice for which the physiological, biochemical, and genetic control has already been well studied (Wang et al 2001).

Gene prediction

The predictive power of the comparative alignments works extremely well for biochemical genes and genes controlling morphological traits. Figure 1 shows the comparative locations of a few such genes. These include major structural loci, such as the *waxy* genes, and classical mutants, such as genes for liguleless and dwarfism. Often, major genes in one species are aligned with QTLs for the same trait in another, as is the case with the domestication-related shattering genes in wheat and rice, which are aligned with QTLs for the same trait in both maize genomes. It has been suggested that some genes, such as those controlling disease resistance, find themselves under such intense selection pressure that colinearity, particularly involving copy numbers, may quickly be lost (Leister et al 1998).

Cross-genome gene isolation

Strong colinearity between the genomes of the different cereals opens up new possibilities for gene isolation. For example, until recently, wheat genes were considered to be well beyond the reach of conventional map-based cloning technology, because the genomic tools were not available and because of the large genome size. Clearly, however, the homoeologues of the same genes in rice are far more tractable to use in map-based cloning, even before the complete rice DNA sequence becomes available from the commercial and public programs. The tools—very dense genetic maps, large EST collections, large-insert BAC, P1-derived artificial chromosome (PAC), and YAC libraries—are already available and in use in most cereal genetics laboratories.

This approach is being used to isolate the wheat *Ph* gene, which controls chromosome pairing by limiting it to homologues, rather than homoeologues (Riley and Chapman 1964). Detailed maps of the critical regions on wheat chromosome 5B and rice chromosome 9 show a very high level of conserved colinearity (Foote et al 1997). However, because allelic variation is not known at the locus in wheat, let alone in rice, a new approach has been developed. Rather than use genetic recombination to narrow down the region that contains *Ph*, Roberts et al (1999) developed a set of overlapping deletions, induced by fast-neutron mutagenesis, spanning the region of 5BL carrying the gene. The 300-kb region of the rice YAC contig corresponding to the section of wheat 5B containing *Ph* has been sequenced and the painstaking functional genomics exercise to identify the roles of the 20 or so open reading frames in the region has begun. This strategy should be applicable to any gene in any crop in which the null phenotype can be clearly recognized. Allelic variation is not a prerequisite and large recombinant populations are not necessary. Availability of the complete rice genomic sequence could make map-based cloning by deletions an even more popular approach.

Monocots and eudicots: Is there still colinearity between rice and *Arabidopsis?*

The close alignment of genes between genomes spread over 60 million years of cereal evolution begs the question as to whether synteny remains over the 140–200-million-year monocot-eudicot divide. Significant residual colinearity would mean that many cereal genomics approaches could be serviced by the *Arabidopsis* DNA sequence.

This question has been addressed by several studies (Paterson et al 1996, Devos et al 1999, van Dodeweerd et al 1999). The consensus is that, although there is some evidence of residual synteny in some regions of the genome, gene sequences and orders have diverged so much that very little predictive power remains in the comparisons. The recent revelations of extensive local chromosome segment duplication that was preceded by a polyploidization event in *Arabidopsis* (The Arabidopsis Genome Initiative 2000) would seem to definitely preclude the wide-scale use of *Arabidopsis* gene orders to approach cereal genetics. Nevertheless, the question should remain open a little longer as enough contiguous genomic rice sequence becomes available to compare directly with that of the model eudicot.

References

Ahn SN, Tanksley SD. 1993. Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA 90:7980-7984.

Bennetzen JL. 2000. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12:1021-1029.

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD. 1989. RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor. Appl. Genet. 78:495-504.

Chen M, San Miguel P, Bennetzen JL. 1998. Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 148:435-443.

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX, Gale MD. 1993. Chromosomal rearrangements in the rye genome relative to that of wheat. Theor. Appl. Genet. 85:673-680.

Devos KM, Beales J, Nagamura Y, Sasaki T. 1999. *Arabidopsis*-rice: Will colinearity allow gene prediction across the eudicot-monocot divide? Genome Res. 9:825-829.

Devos KM, Pittaway TS, Reynolds A, Gale MD. 2000. Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor. Appl. Genet. 100:190-198.

Devos KM, Wang ZM, Beales J, Sasaki T, Gale MD. 1998. Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). Theor. Appl. Genet. 96:63-68.

Dufour M, Deu M, Grivet L, D'Hont A, Paulet F, Glaszmann JC, Hamon P. 1997. Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor. Appl. Genet. 94:409-418.

Feuillet C, Keller B. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. Proc. Natl. Acad. Sci. USA 96:8265-8270.

Foote T, Roberts M, Kurata N, Sasaki T, Moore G. 1997. Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics 147:801-807.

Gale MD, Devos KM. 1998. Plant comparative genetics after 10 years. Science 282:656-659.
Gale M, Moore G, Devos K. 2001. Rice—the pivotal genome in cereal comparative genetics. Chichester (England): John Wiley & Sons. p 46-52. Copyright© Novartis Foundation 2001.
Glaszmann JC, Dufour P, Grivet L, D'Hont A, Deu M, Paulet F, Hamon P. 1997. Comparative genome analysis between several tropical grasses. Euphytica 96:13-21.
Kennard W, Phillips RL, Porter R, Grombacher A. 1999. A comparative map of wild rice (*Zizania palustris* L. 2n=2x=30). Theor. Appl. Genet. 99:793-799.
Kurata N, Moore G, Nagamura Y, Foote T, Yano M, Minobe Y, Gale MD. 1994. Conservation of genome structure between rice and wheat. Bio/Technology 12:276-278.
Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos KM, Graner A, Schulze-Lefert P. 1998. Rapid reorganisation of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci. USA 95:370-375.
Moore G, Devos KM, Wang ZM, Gale MD. 1995. Cereal genome evolution: Grasses, line up and form a circle. Curr. Biol. 5:737-739.
Nagamura Y, Inoue T, Antonio BA, Shimano T, Kajiya H, Shomura A, Lin SY, Kuboki Y, Harushima Y, Kurata N, Minobe Y, Yano M, Sasaki T. 1995. Conservation of duplicated segments between rice chromosomes 11 and 12. Breed. Sci. 45:373-376.
Paterson AH, Lan T-H, Reischmann KP, Chang C, Lin Y-R, Liu S-C, Burow MD, Kowalski SP, Katsar CS, Del Monte TA, Feldmann KA, Schertz KF, Wendel JF. 1996. Toward a unified genetic map of higher plants transcending the monocot-dicot divergence. Nat. Genet. 14:380-382.
Riley R, Chapman V. 1964. Cytological determination of the homoeology of chromosomes of *Triticum aestivum*. Nature 203:156-158.
Roberts MA, Reader SM, Dalgliesh C, Miller TE, Foote TN, Fish LJ, Snape JW, Moore G. 1999. Induction and characterization of *Ph1* wheat mutants. Genetics 153:1909-1918.
Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A. 2000. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381-391.
The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.
Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z. 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. USA 96:7409-7414.
Toenniessen GH. 1998. Rice biotechnology capacity building in Asia. In: Ives CL, Bedford BM, editors. Agricultural biotechnology in international development. Wallingford (UK): CAB International. p 201-212.
Van Deynze AE, Nelson JC, O'Donoughue LS, Ahn S, Siripoonwiwat W, Harrington SE, Yglesias ES, Braga DP, McCouch SR, Sorrells ME. 1995a. Comparative mapping in grasses: oat relationships. Mol. Gen. Genet. 249:349-356.
Van Deynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells ME. 1995b. Comparative mapping in grasses: wheat relationships. Mol. Gen. Genet. 248:744-754.
van Dodeweerd AM, Hall AM, Bent EG, Johnson SJ, Bevan MW, Bancroft I. 1999. Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis*. Genome 42:887-892.
Wang ZM, Thierry A, Panaud O, Gale MD, Sarr A, Devos KM. 2001. Trait mapping in foxtail millet. Theor. Appl. Genet. (In press.)

Zhang H, Jia J, Gale MD, Devos KM. 1998. Relationship between the genomes of *Aegilops umbellulata* and wheat. Theor. Appl. Genet. 96:69-75.

Notes

Authors' address: John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Phylogeny of the genus *Oryza* as revealed by molecular approaches

Song Ge, Tao Sang, Bao-Rong Lu, and De-Yuan Hong

The genus *Oryza*, consisting of approximately 24 species, contains an enormous gene pool for the genetic improvement of rice cultivars. To realize the potential agricultural value of wild rice germplasm, continuous efforts have been made to understand the taxonomy, genomic composition, genetic diversity, and phylogeny of the *Oryza* species. Based on genome analyses using cytogenetic methods and genomic DNA hybridization, nine genome types (A, B, BC, C, CD, E, F, G, and HJ) were determined for various *Oryza* species. Recently, we amplified by polymerase chain reaction, cloned, and sequenced two nuclear genes, *Adh1* and *Adh2*, and a chloroplast gene, *matK*, from all of the *Oryza* species. Phylogenetic relationships of rice genomes and species were reconstructed by comparing three gene phylogenies. The results supported previous recognition of nine genome types. A new genome type, HK, was recognized for *Oryza schlechteri* and *Porteresia coarctata*, suggesting that *P. coarctata* might be included in *Oryza*. The study further revealed that the EE genome species is most closely related to the DD genome progenitor that gave rise to the CD genome. In contrast to a single origin of three CCDD genome species, the BBCC genome species had different origins because their maternal parents had either a B or C genome. The G genome is the most basal lineage on the phylogeny. The AA genome group, which contains the two cultivated rice species, is a recently diverged and rapidly radiated lineage within the rice genus. Despite our better understanding of phylogenetic relationships of *Oryza* species based on the three genes, relationships of rice genomes and species still remain partially resolved or weakly supported. Finally, we discuss some remaining questions and future perspectives concerning studies on the phylogeny and evolution of the rice genus.

Rice is the world's most important crop. It is the staple food for nearly one-half of the global population. Intensive selection in modern breeding practices has led to a severe loss of genetic diversity in the cultivated rice gene pool, rendering rice varieties more vulnerable to disease and insect epidemics and consequently affecting the sta-

bility of rice productivity (Lu 1999). As a consequence, the sustainability of world food security is threatened. Future increases in rice production will essentially rely on the use of genetic resources in the rice gene pool, including wild species of rice.

Wild species of rice have diversified in a wide range of environments over 40 million years (Stebbins 1981) and have become a vast gene pool valuable for the genetic improvement of rice varieties (Khush 1997). Recent advances in molecular breeding approaches hold tremendous potential for such improvement through the transfer of beneficial genes from wild rice species to cultivars (Brar and Khush 1997, Xiao et al 1996, Tanksley and McCouch 1997). However, the effective utility and conservation management of the valuable genetic diversity in the rice gene pool rely significantly on a clear understanding of the evolutionary relationships of rice species and, subsequently, the development of a natural classification of the genus *Oryza* (Ge et al 1999, Lu 1999).

In this chapter, we briefly review the historical changes of classification in the genus *Oryza* and the related phylogenetic studies, and introduce our recent phylogenetic study on the rice genus using sequences of two nuclear *Adh* genes and one chloroplast gene, *matK*. We emphasize relationships among genome types and the origin of the tetraploid species in the genus, and compare the multiple gene phylogenies with species relationships based on morphology, cytogenetics, and other molecular studies. Furthermore, we discuss some remaining questions and perspectives concerning studies on the phylogeny and evolution of the rice genus.

Historical perspective

Intrageneric grouping and classification of the genus

The genus *Oryza* L. is classified under the tribe Oryzeae, subfamily Oryzoideae, of the grass family Poaceae (Gramineae) (Lu 1999). This genus includes two cultivated species and more than 20 wild species, distributed in tropical Asia, Africa, Australia, and Central and South America (Khush 1997). Since the description of *Oryza* by Linnaeus (1753), many studies on taxonomy, biosystematics, and phylogeny of the *Oryza* species have been conducted (Prodoehl 1922, Roschevicz 1931, Sampath 1962, Tateoka 1963, Sharma and Shastry 1965, Second 1985, Vaughan 1989, 1994, Morishima et al 1992, Wang et al 1992, Aggarwal et al 1999, Lu 1999). Taxonomists such as Roschevicz (1931), Sampath (1962), Tateoka (1963), Sharma and Shastry (1965), Vaughan (1989, 1994), and Lu (1999) have made major taxonomic revisions of the genus, which gave rise to modern classification systems of the genus *Oryza* (Table 1).

The delimitation of *Oryza* has varied over time since Roschevicz's (1931) fundamental revision, which recognized 20 species in four sections (Table 1). Two previously recognized *Oryza* species, *O. subulata* and *O. coarctata*, were removed from *Oryza* to form two monotypic genera, *Rhynchoryza* and *Porteresia*, respectively (Tateoka 1965a). Three species, *O. angustifolia, O. perrierri*, and *O. tisseranti*, were transferred to the closely related genus *Leersia* (Launert 1965). These changes have been accepted in the recent taxonomic treatments (Vaughan 1989, 1994, Lu 1999).

Table 1. Species recognized in *Oryza* with their genome types and distributions (only includes species recognized by the authors).

Complex[a] Species	Genome	Distribution	Intrageneric classification systems		
			Roschevicz (1931)	Sharma and Shastry (1965)	Vaughan (1994)
O. sativa complex					
O. sativa	AA	Worldwide	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. nivara	AA	Tropical and subtropical Asia	–[b]	Sect. *Oryza*	Sect. *Oryza*
O. rufipogon	AA	Tropical and subtropical Asia, tropical Australia	–[b]	Sect. *Oryza*[c]	Sect. *Oryza*
O. glaberrima	AA	West Africa	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. barthii	AA	Africa	Sect. *Sativa*[c]	Sect. *Oryza*	Sect. *Oryza*
O. glumaepatula	AA	South and Central America	–[b]	Sect. *Oryza*	–[b]
O. longistaminata	AA	Africa	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. meridionalis	AA	Tropical Australia and Irian Jaya	–[b]	–[b]	Sect. *Oryza*
O. officinalis complex					
O. punctata	BB	Africa	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. chweinfurthiana	BBCC	Africa	Sect. *Sativa*	Sect. *Oryza*	–[b]
O. officinalis	CC	Tropical and subtropical Asia, tropical Australia	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. malampuzhaensis	BBCC	India	–[b]	Sect. *Oryza*	–[b]
O. rhizomatis	CC	Sri Lanka	–[b]	–[b]	Sect. *Oryza*
O. minuta	BBCC	Philippines, Thailand, and Papua New Guinea	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. eichingeri	CC BBCC	South Asia and East Africa	–[b]	Sect. *Oryza*	Sect. *Oryza*
O. alta	CCDD	South and Central America	–[b]	Sect. *Oryza*	Sect. *Oryza*
O. grandiglumis	CCDD	South and Central America	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. latifolia	CCDD	South and Central America	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. australiensis	EE	Tropical Australia	Sect. *Sativa*	Sect. *Oryza*	Sect. *Oryza*
O. ridleyi complex					
O. ridleyi	HHJJ	South Asia	Sect. *Coarctata*	Sect. *Padia*	Sect. *Ridleyanae*
O. longiglumis	HHJJ	Irian Jaya, Indonesia, and Papua New Guinea	–[b]	Sect. *Padia*	Sect. *Ridleyanae*

continued on next page

Table 1 continued

Complex[a] Species	Genome	Distribution	Intrageneric classification systems		
			Roschevicz (1931)	Sharma and Shastry (1965)	Vaughan (1994)
O. meyeriana complex					
O. meyeriana	GG	South and Southeast Asia	–[b]	Sect. *Padia*	Sect. *Granulata*
O. granulata	GG	Southeast Asia	Sect. *Granulata*	Sect. *Padia*	Sect. *Granulata*
O. neocaledonica	GG	New Caledonia	–[b]	–[b]	–[b]
Species not included in any of the complexes					
O. brachyantha	FF	Africa	Sect. *Coarctata*	Sect. *Angustifolia*	Sect. *Ridleyanae*
O. schlechteri	HHKK	Papua New Guinea	Sect. *Coarctata*	Sect. *Padia*	Sect. *Ridleyanae*
O. coarctata (=*Porteresia coarctata*)	HHKK	Bangladesh	Sect. *Coarctata*	–[b]	–[b]

[a]The concept of complex essentially follows Vaughan (1989). [b]Species was not included in the system. [c]Species was given a different name in the system.

However, the subdivisional classification remains somewhat controversial, and three to six sections in *Oryza* were generally recognized in most taxonomic treatments at different times (Lu 1999).

To explain the variation in *Oryza* species based on morphology, Tateoka (1962a,b) grouped *Oryza* species into several complexes. Following this concept, Vaughan (1989) developed a classification system in which 20 species in four complexes were recognized, but he left two species with no clear affinity association. Vaughan's systems (1989, 1994) appropriately reflected the current circumscription and enumeration of the genus, and have actually become a practical framework for rice taxonomists and breeders (Table 1), although the taxonomic ranking of the complex has no legal standing in the International Code of Botanic Nomenclature (ICBN).

Genome analysis

Genome analysis through assessment of chromosome pairing in an interspecific hybrid can be traced back to the beginning of the 20th century (Kihara 1924). It has profoundly influenced plant taxonomy and classification, particularly those of economically important groups such as *Triticum*, *Hordeum*, *Oryza*, *Avena*, *Brassica*, *Gossypium*, and *Solanum* because of the important value of chromosome-pairing data in plant breeding (Seberg and Petersen 1998). Traditional genome analysis has contributed greatly to the understanding of genome constitutions of *Oryza* species (Nayar 1973). By studying the meiotic pairing of hybrids between *Oryza* species, various genome types have been determined. These are the A, B, C, E, and F genomes in diploids and BC and CD genomes in tetraploids (Morinaga 1943, 1964, Li 1964, Nayar 1973). However, genome types of a few species could not be determined by this method because of the difficulties in obtaining F_1 hybrids.

Using the technology of total genomic DNA hybridization, Aggarwal et al (1997) assigned the G genome to *O. granulata* and *O. meyeriana* and the HJ genome to *O. longiglumis* and *O. ridleyi*. Radioactive probes made from the genomic DNA of each of these two groups of species hybridized strongly only to their own total genomic DNA, suggesting that each group of species has a unique genome type (Aggarwal et al 1997).

Traditional genome analysis is based on the assumption that only similar chromosomes pair during meiosis, and the extent of chromosome pairing in hybrids reflects the degree of relationship between the parental species (Seberg and Petersen 1998). As indicated by Seberg and Petersen (1998), data from chromosome pairing are captured as pair-wise comparisons and are amenable only to phenetic analysis and hence are not suited for phylogenetic inferences. Similarly, the recent genomic DNA hybridization identified genetic similarity among the rice genomes and species, but could not determine phylogenetic relationships within or between the genome types (Aggarwal et al 1997). As a result, little is known on the phylogenetic relationships of the *Oryza* species, although much work has been done in genome identification (Cordesse et al 1992, McIntyre and Winberg 1998).

Molecular markers and evolutionary relationships of rice species

Considerable effort has been devoted to studying the relationships of rice genomes and species, although the phylogeny of *Oryza* has been less explored than that of other major crop plants such as maize (Gaut et al 2000), soybean (Doyle et al 1996), and cotton (Wendel and Albert 1992).

Morishima and Oka (1960) studied 42 morphological characters of 16 *Oryza* species by numerical taxonomic methods. They found that Roschevicz's section *Sativa* could be divided into two groups. One group was represented by *O. sativa*, with five species (A genome), and the other by *O. officinalis*, with seven species (C, BC, CD genomes) (Morishima and Oka 1960). Based on 15 morphological, ploidy, and habit characters and the distribution of *Oryza* species, Sharma (1986) reconstructed a phylogenetic tree of *Oryza* species and found that *O. ridleyi*, *O. meyeriana*, and *O. schlechteri* formed a clade at the basal position of this tree.

Evolutionary relationships among the rice genomes and species were also investigated by various methods, including protein (Sarkar and Raina 1992), isozyme (Second 1985), and organelle and nuclear DNA restriction fragment length polymorphisms (RFLPs) (Dally and Second 1990, Second 1991, Wang et al 1992, Abe et al 1999). These studies provided only limited information on the phylogeny of the genus because of either incomplete sampling or the nature of the data. Wang et al (1992) conducted the most comprehensive study on phylogenetic relationships of 21 *Oryza* species. The UPGMA (unweighted pair-group method, arithmetic average) dendrogram, generated by genetic distances estimated from 25 genomic RFLP clones, indicated four species complexes that corresponded to those of Vaughan (1989). In addition, Wang et al (1992) suggested that BBCC tetraploids (*O. malampuzhaensis*, *O. punctata*, and *O. minuta*) were likely to have independent origins and that CCDD tetraploids possibly had an ancient origin with the closest living diploids containing the C and E genomes. However, using phenetic methods and data other than DNA sequences, these studies often produced contradictory results and failed to reconstruct origins of allotetraploid species with their parents indicated clearly.

Recently, the PCR-based methods such as random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphisms (AFLPs), and inter simple sequence repeats (ISSRs) have been increasingly used for a variety of studies on population genetics and phylogenetic reconstruction at lower taxonomic levels (Wolfe and Liston 1998). In addition to the application of these markers to the analysis of population genetics and germplasm management, they have also been used to detect genetic similarity and phylogenetic relationships among *Oryza* species (Ishii et al 1996, Aggarwal et al 1999, Joshi et al 2000). However, PCR-based analyses are sensitive to minor modifications in experimental protocols and to problems of homology assessment (Wolfe and Liston 1998), particularly when involving distinct species. Moreover, the dendrogram generated based on genetic distance may not accurately reflect the phylogenetic relationships of the genomes and species. Therefore, there is a great need for a robust phylogenetic reconstruction of *Oryza* species based on molecular phylogenetic approaches and DNA sequence data.

Multiple gene trees: new insights into the phylogeny of *Oryza*

Recent advances in phylogenetic methodologies together with the availability of new types of data have significantly increased the accuracy of phylogenetic reconstruction (Donoghue 1994). In particular, the phylogenetic utility of sequences of low-copy nuclear genes has allowed us to address phylogenetic questions that were previously challenging. Low-copy nuclear genes, which contain fast-evolving introns and are less susceptible to concerted evolution, are especially powerful markers for reconstructing allopolyploidization at the interspecific level (Sang and Zhang 1999). Caution has to be exercised, however, when we infer a species tree from gene trees because gene trees may not represent species trees when factors such as paralogy (referring to the common ancestry of the sequences originated from a gene duplication event), hybridization, or lineage sorting are involved. Therefore, it is an increasingly widespread practice to apply multiple data sets to a common group of taxa (Wendel and Doyle 1998). Congruence among the multiple gene phylogenies provides a sound foundation for a robust reconstruction of phylogenetic relationships of the organisms themselves (Hillis and Huelsenbeck 1995).

Recently, we have sequenced two nuclear genes, *Adh1* and *Adh2*, and a chloroplast gene, *matK*, by PCR-amplifying and cloning of DNA from 23 *Oryza* species. By comparing *Adh1* and *Adh2* trees and the *matK* tree of the *Oryza* species, a new genome type (HK) was identified for the tetraploid *O. schlechteri*, the only *Oryza* species with an unknown genome type. The origin of the allotetraploid species was reconstructed based on the *Adh* gene phylogenies with their maternal parents identified by the *matK* phylogeny. In addition, phylogenetic relationships of the rice genomes and species were reconstructed based on the three gene phylogenies (Figs. 1, 2, 3; Ge et al 1999).

The origin of allotetraploid species

Cytogenetic studies have shown that all the polyploid *Oryza* species are allotetraploids (2n=4x=28) that make up more than one-third of the species in the genus (Vaughan 1994). The formation of so many allotetraploids obscured the morphological distinction between some diploid species, and consequently made *Oryza* a taxonomically and phylogenetically difficult group (Vaughan 1989).

Phylogenetic analysis of single or low-copy nuclear gene sequences offers a new and effective way to study the evolution of allopolyploids (Gaut and Doebley 1997, Small et al 1998, Sang and Zhang 1999). An allotetraploid, originated through hybridization of two diverged diploid species, includes two distinct diploid genomes. Homeologous loci contributed by diploid parents can be cloned and sequenced from an allotetraploid species. Analysis of these sequences together with the gene sequences of the putative diploid parents enables us to unravel the pattern of hybrid speciation. Furthermore, since the chloroplast genome is maternally inherited in rice, analysis of the chloroplast *matK* gene offers an opportunity to identify the maternal parents of allotetraploid species.

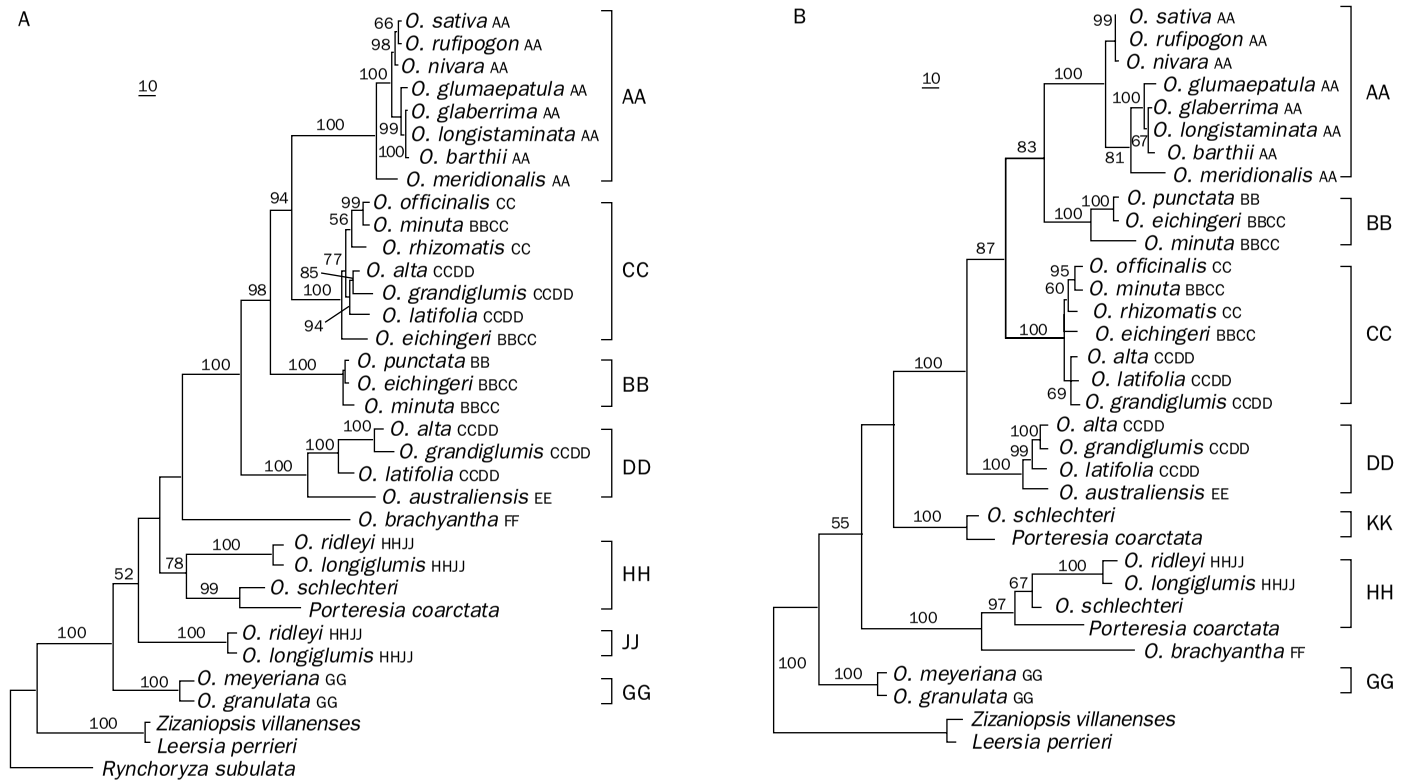


Fig. 1. Phylogenies of *Adh1* and *Adh2* genes of *Oryza* species. (A) The strict consensus tree of two equally most parsimonious trees of *Adh1* gene sequence (CI = 0.75, RI = 0.87). (B) The strict consensus tree of four equally most parsimonious trees of *Adh2* gene sequence (CI = 0.76, RI = 0.88). Numbers are bootstrap percentages above 50%. Branch lengths are proportional to the number of nucleotide substitutions and scale bars indicate 10 substitutions. Small letters following a species name indicate the recognized genome type of the species. On each *Adh* gene tree, the appearance twice of an allotetraploid species represents two distinct types of sequences cloned from the same individual of the species. The genome type of a monophyletic group is indicated (Ge et al 1999). CI = consistency index, RI = retention index.

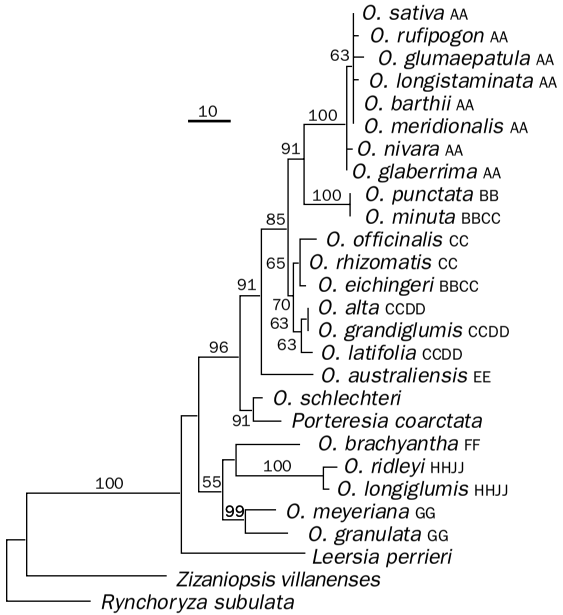


Fig. 2. The single most parsimonious tree generated from *matK* gene sequences of rice species (CI = 0.87, RI = 0.90). Numbers are bootstrap percentages above 50%. Branch lengths are proportional to the number of nucleotide substitutions and scale bars indicate 10 substitutions. Small letters following a species name indicate the recognized genome type of the species (Ge et al 1999). CI = consistency index, RI = retention index.

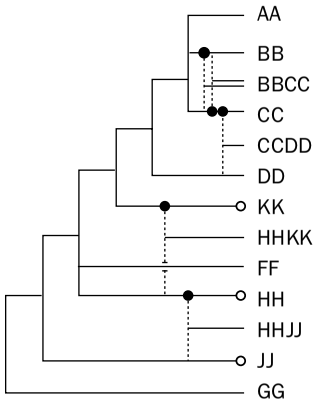


Fig. 3. Evolutionary relationships of the rice genomes inferred from *Adh1*, *Adh2*, and *matK* gene phylogenies. Broken lines indicate origins of allotetraploids, ● = maternal parents, ○ = unidentified diploid genomes (Ge et al 1999).

Two distinct types of sequences were identified for each of the *Adh1* and *Adh2* genes for the majority of tetraploid species. Sequences of the same gene cloned from both diploid and tetraploid species were aligned for phylogenetic analyses. For allotetraploid species with the BC and CD genomes, on both *Adh* gene trees, the types of sequences formed monophyletic groups with the diploid BB and CC genome species, respectively (Fig. 1). Two types of sequences cloned from the CCDD genome species formed monophyletic groups with the diploid CC and EE genome species, respectively. The congruent relationships between the *Adh1* and *Adh2* gene trees strongly support previous hypotheses of the allotetraploid nature of the BC and CD genomes. Similarly, the allotetraploid nature was also determined for *O. ridleyi*, *O. longiglumis* (HHJJ), and *O. schlechteri* and *Porteresia coarctata* (HHKK) (Fig. 1). However, the allotetraploid origin of the HHJJ and HHKK species was inferred only on the basis of sequence polymorphism at one of the *Adh* loci; thus, the confidence was not as high as that of the BC and CD genomes, which are supported by both *Adh* genes.

Based on the *matK* gene tree that represents a maternal genealogy of rice species, the maternal parents of each allotetraploid species can be easily determined. Because the CCDD genome species formed a monophyletic group with the CC genome species on the *matK* phylogeny (Fig. 2), the CC genome species apparently served as the maternal donor of the CCDD genome species. On the *Adh* and *matK* gene trees, three CCDD genome species formed monophyletic groups in which *O. alta* and *O. grandiglumis* were sister species, suggesting that these three allotetraploid species originated from a single hybridization event. In contrast, the two BBCC genome species, *O. minuta* and *O. eichingeri*, may have had different origins. The former had the B genome as the maternal donor and the latter the C genome as the maternal donor (Fig. 2). The multiple origins of the BBCC genome species have also been suggested previously by cpDNA RFLP analysis (Dally and Second 1990). This explains the taxonomic complexity in the BBCC groups (Vaughan 1989).

The position of *O. schlechteri* and *P. coarctata* on the *matK* gene tree was apparently the same as that of the K genome on the *Adh2* tree, suggesting that the K genome was the maternal donor of the HHKK genome species. However, the maternal parent of the HHJJ genome species seems less clear.

Genome types identified on the *Adh* gene phylogenies

Figure 1 shows that, on both the *Adh1* and *Adh2* gene trees, each clade of a previously recognized genome type (including sequences from both diploid and tetraploid species) is supported by a high bootstrap value, mostly 100%. This supports almost all genome types of *Oryza* species identified by traditional genome analysis or genomic DNA hybridization (Nayar 1973, Aggarwal et al 1997). Based on the *Adh* gene phylogenies, we therefore designated the clade containing the diploid BB genome species and one type of sequence cloned from the BBCC genome species as the B genome. The clade that includes the diploid CC genome species and sequences cloned from the BBCC and CCDD genome species was recognized as the C genome clade (Fig. 3). The D type of sequence cloned from the CD genome forms a monophyletic

group with the diploid E genome species, *O. australiensis*, on both *Adh* gene trees (Fig. 1). Although there were different opinions on the D genome donors of the CCDD species (Wang et al 1992, Fukui et al 1997), strong support of monophyly of the D genome sequence and the E genome sequence on both *Adh* trees suggested the closest relationship of the E genome to the D genome progenitors, which gave rise to the CCDD species. Thus, the clade containing the diploid E genome and sequences from the CCDD genome species could be treated as the D genome clade (Fig. 3).

The clade containing *O. ridleyi*, *O. longiglumis*, *O. schlechteri*, and *Porteresia coarctata* on each *Adh* gene tree (Fig. 1) was designated as the H genome clade. The other clade containing *O. ridleyi* and *O. longiglumis* on the *Adh1* phylogeny represented the J genome. The clade containing *O. schlechteri* and *P. coarctata* on the *Adh2* tree was then given a new genome type, K (Fig. 1). Consequently, all species in the genus *Oryza* have the genome type recognized, that is, the A, B, C, BC, CD, E, F, G, HJ, and HK genomes (Table 1).

Phylogeny of rice genomes and intrageneric classification

Based on the two *Adh* data sets, we generated a consensus tree (Fig. 3) in which the congruent relationships between the two gene trees were maintained and the incongruent relationships were left unresolved (Swofford 1991). Monophyletic groups revealed by the phylogenetic reconstruction are either concordant or discordant with taxonomic sections recognized in the most recent classification of the genus (Table 1). The A, B, and C genomes are most closely related and together form a sister group with the D (E) genome. This monophyletic group, containing the A to E genomes, corresponds to section *Oryza* (section *Sativa sensu* Roschevicz 1931) (Fig. 3, Table 1). The G genome, which occupies the most basal position of the genus, constitutes section *Granulata sensu* Roschevicz (1931). The remaining genome types that are included in section *Ridleyanae*, however, form a paraphyletic group in the phylogenetic hypothesis (Fig. 3). It is evident that the circumscription of section *Oryza* is consistent with those in most taxonomic treatments (Table 1). For other sections, however, the *Adh* gene phylogenies agree better with Roschevicz's (1931) classification that recognized section *Ridleyanae* and section *Granulata* although section *Ridleyanae* is paraphyletic (Figs. 1, 2). The section *Padia* recognized by Sharma and Shastry (1965) forms a polyphyletic group on the *Adh* gene phylogenies (Fig. 3, Table 1).

In addition to the relationships among genome types, three gene phylogenies provide some implications for interspecific relationships within genomes. The A genome, which is present in cultivated rice, is one of the most recently diverged lineages within the rice genus (Fig. 1). It contains only diploid species and has the widest geographic distribution compared with other genome groups in *Oryza*. Apparently, the A genome is a well-adapted group that not only diversified recently but also radiated rapidly. The *Adh* phylogenies further indicated that the Asian cultivated rice, *O. sativa*, is most closely related to two Asian wild species, *O. nivara* and *O. rufipogon*, supporting the hypothesis of an Asian origin of *O. sativa* (Second 1982, Khush 1997). The African cultivated *O. glaberrima* is closely related to two African wild species, *O.*

barthii and *O. longistaminata*, and to the New World species, *O. glumaepatula*. These results are in general agreement with studies using various markers (Wang et al 1992, Aggarwal et al 1999, Joshi et al 2000).

An important argument in our three gene phylogenies involves the position of *Porteresia coarctata* that was once recognized as a species of *Oryza* (*O. coarctata* Ref.), but treated as a monotypic genus by Tateoka (1965a). It is shown that *P. coarctata* and *O. schlechteri* share the same genome type, HHKK, and are closely related to each other according to the phylogenies of the two *Adh* genes and the *matK* gene (Figs. 1, 2). Therefore, *Porteresia coarctata* should be treated as a member of *Oryza* rather than as a separate genus. Recent AFLP and ISSR analyses also showed some affinities of *P. coarctata* and *Oryza* species (Aggarwal et al 1999, Joshi et al 2000).

Remaining questions and future directions

A robust phylogeny and the taxonomic position of particular species

Our recent phylogenetic study using *Adh* and *matK* genes has led to a better understanding of the phylogeny of *Oryza*, but relationships among the genomes still remain partially unresolved or weakly supported (Fig. 1). In the *Adh* phylogenies, for example, the A and C genomes are sister groups on the *Adh1* tree, whereas the A and B genomes are sister groups on the *Adh2* tree (Fig. 1). The topological incongruence was statistically significant and the reasons for the incongruence need to be further investigated (Ge et al 1999).

Furthermore, the relationship of the F genome remains unsolved on the overall phylogenetic hypothesis of the genus (Figs. 1, 3). It was grouped strongly with the H genome on the *Adh2* tree but did not form a strongly supported group with any genome type on the *Adh1* tree (Fig. 1). Previous morphological and molecular studies also showed conflicting results regarding the taxonomic position of *O. brachyantha* of the F genome. It has been associated with the *O. ridleyi* complex on the basis of its embryo structure and other morphological traits (Vaughan 1994), but showed some affinity to the *O. sativa* complex according to a nuclear RFLP study (Wang et al 1992). In contrast, data from recent AFLP and ISSR studies demonstrated that the F genome did not align with any species complex in *Oryza* (Aggarwal et al 1999, Joshi et al 2000). Because *O. brachyantha* is morphologically, cytologically, and genetically distinct from all other *Oryza* species (Vaughan 1989, Aggarwal et al 1999), Lu (1999) treated it as a separate section. The position of the F genome remains questionable and should be further clarified with additional gene markers (Ge et al 1999).

Additionally, we were not able to isolate one homeologous *Adh1* locus from *O. ridleyi* and *O. longiglumis* and one homeologous *Adh2* locus from *O. schlechteri* and *Porteresia coarctata* although various PCR strategies have been tried (Ge et al 1999). These results support the hypothesis of gene deletion, that is, deletion of the *Adh1* gene from the K genome of *O. schlechteri* and *P. coarctata* and deletion of the *Adh2* gene from the J genome of *O. ridleyi* and *O. longiglumis*. The deletion of one of the homeologous loci may be a mechanism of reduction of genetic redundance in these allotetraploid rice species. On both *Adh* phylogenies, the sequences of the HJ and HK

genomes occupy the basal positions relative to those of the BC and CD genomes (Fig. 1), suggesting that the former had more ancient origins, which may have allowed more extensive genomic rearrangement including possible deletions of some homeologous loci. Southern blotting and more gene phylogenies can be employed to further test the hypothesis of gene deletion.

Moreover, the study focused mainly on relationships among genome types and did not involve interspecific relationships in depth. A well-resolved and strongly supported phylogeny of *Oryza* at the species level is needed to provide a basis for an informative and predictive classification, a better understanding of evolution and biogeography, and more effective use of wild rice germplasm.

Evolution of genes and genomes in allotetraploid species

Allopolyploidy is a widely documented mechanism of speciation in flowering plants (Masterson 1994). Previous cytogenetic studies have inferred genomic compositions of the allotetraploid species and provided necessary background as well as intriguing hypotheses for further molecular phylogenetic analyses. The *Adh* phylogenies determined the origins of allotetraploid genomes of *Oryza* and suggested that these genomes originated at different times. The BC genome originated most recently, while the HJ and HK genomes are more ancient (Figs. 1, 3). Allotetraploid species with different origins and ages in the *Oryza* genus serve as a good system for investigating a variety of questions concerning the formation and evolution of polyploid species. As reviewed in recent literature (Soltis and Soltis 2000, Wendel 2000), the study of polyploidy represents an intriguing topic in plant evolutionary biology. We are particularly interested in the following questions in the rice genus.

Gene silencing or deletion following polyploidization. It has been suggested that cycles of polyploidization and diploidization have contributed to the species diversity of angiosperms (Wendel 2000). If this is true, gene silencing or deletion should occur at a large number of homeologous loci in older allopolyploids. For the *Adh* genes, both homeologous loci are retained in the recently originated BC and CD genomes. As mentioned above, however, multiple combinations of PCR primers failed to amplify one of the homeologous loci of the *Adh2* gene from the HJ genome and one of the homeologous loci of the *Adh1* gene from the HK genome (Ge et al 1999). This suggests that one of the *Adh* homeologous loci has been deleted from the allotetraploid genomes. It is thus of great interest to investigate whether gene silencing (e.g., pseudogene formation) or gene deletion also occurs at other nuclear loci of the HJ and HK genomes.

Adaptive evolution of duplicated genes in allotetraploid genomes. Although an allotetraploid genome tends to reduce its genetic redundance through gene silencing or deletion, it may also take advantage of genetic redundancy. Genome duplication could allow the relaxation of purifying selection on one of the homeologous loci and subsequently the evolution of a new function of the locus (Soltis and Soltis 2000, Wendel 2000). If the new function is acquired through fixation of mutations in the protein-coding region driven by positive selection, it can be detected by comparing the ratio of nonsynonymous to synonymous substitutions (Ka/Ks) (Li 1997). Further,

testing Ka/Ks between the homeologous loci of the allotetraploid and the genes of the diploid species with the corresponding genomes will determine which homeologous locus has undergone adaptive evolution.

Recurrent formation of an allotetraploid genome. Multiple independent formations of an allopolyploid genome or species from different populations of the diploid progenitors have been observed in both plants and animals (Soltis and Soltis 2000). The *O. officinalis* complex with the B, C, and BC genomes is well known for its taxonomic complexity in *Oryza*, and the polyploids, in particular, are difficult to group (Vaughan 1989). A recent study using *Adh* and *matK* genes suggested that the BC genome originated independently in Africa and Asia (Ge et al 1999). The BC genome may have also formed multiple times within each continent. The recurrent formation of the BC genome is probably responsible for the taxonomic difficulty in the species complexes containing the B, C, and BC genomes (Tateoka and Pancho 1963, Tateoka 1965b). Because allotetraploids in *Oryza* involved many genomes and originated from different lineages and at different times, *Oryza* should serve as a good model system for studying various aspects of polyploid genome evolution.

Beyond its agronomic importance, rice has become a model organism for biological research. Because of its small genome size, *Oryza sativa* will be the first monocot and the second plant species (following *Arabidopsis thaliana*) with the entire genome sequenced (Sasaki 1998). The genus *Oryza* thus offers an excellent system for studies of plant evolution at the genomic level. This will allow us to examine genomic changes that are adaptively or agriculturally important. The availability of genomic information of the selected nuclear gene markers, such as chromosomal locations, will allow us to investigate genome evolution. This will, on the one hand, help address a fundamental question of which genomic consequences have provided adaptive advantages for allopolyploids and have made this mode of speciation so prevalent in angiosperms (Wendel 2000). On the other hand, with a robust phylogeny and multiple gene markers, many intriguing questions involving the biogeography of particular species, the genome, and the whole genus can be addressed.

References

Abe T, Edanami T, Adachi E, Sasahara T. 1999. Phylogenetic relationships in the genus *Oryza* based on mitochondrial RFLPs. Genes Genet. Syst. 74:23-27.

Aggarwal RK, Brar DS, Khush GS. 1997. Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. Mol. Gen. Genet. 254:1-12.

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS. 1999. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor. Appl. Genet. 98:1320-1328.

Brar DS, Khush GS. 1997. Alien introgression in rice. Plant Mol. Biol. 35:35-44.

Cordesse F, Grellet F, Reddy AS, Delseny M. 1992. Genome specificity of rDNA spacer fragments from *Oryza sativa* L. Theor. Appl. Genet. 83:864-870.

Dally AM, Second G. 1990. Chloroplast DNA diversity in wild and cultivated species of rice (genus *Oryza*, section *Oryza*): cladistic-mutation and genetic-distance analysis. Theor. Appl. Genet. 80:209-222.

Donoghue MJ. 1994. Progress and prospects in reconstructing plant phylogeny. Ann. Mo. Bot. Gard. 81:405-418.
Doyle JJ, Kanazin V, Shoemaker RC. 1996. Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (*Glycine*: Leguminosae). Mol. Phylogen. Evol. 6:438-447.
Fukui K, Shishido R, Kinoshita T. 1997. Identification of the rice D-genome chromosomes by genomic *in situ* hybridisation. Theor. Appl. Genet. 95:1239-1245.
Gaut BS, d'Ennequin MLT, Peek AS, Sawkins MC. 2000. Maize as a model for the evolution of plant nuclear genomes. Proc. Natl. Acad. Sci. USA 97:7008-7015.
Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA 94:6809-6814.
Ge S, Sang T, Lu BR, Hong DY. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA 96:14400-14405.
Hillis DM, Huelsenbeck JP. 1995. Assessing molecular phylogenies. Science 267:255-256.
Ishii T, Nakano T, Maeda H, Kamijima O. 1996. Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. Genes Genet. Syst. 71:195-201.
Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK, Brar DS. 2000. Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. Theor. Appl. Genet. 100:1311-1320.
Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. 35:25-34.
Kihara H. 1924. Cytologische und genetische Studien bei wichtigen Getreidearten mit besonderer Rucksicht auf das Verhalten der Chromosomen und die Sterilitat in den Bastarden. Mem. Coll. Sci. Kyoto Imp. Univ., Ser. B 1:1-200.
Launert E. 1965. A survey of the genus *Leersia* in Africa. Senckenbergiana Biol. 46:129-153.
Li HW. 1964. Studies on genetic and cytogenetic evidence for species relationships in the Republic of China. In: Rice genetics and cytogenetics. Proceedings of the Symposium on Rice Genetics and Cytogenetics. Amsterdam (Netherlands): Elsevier Publishing Company. p 119-131.
Li WH. 1997. Molecular evolution. Sunderland, Mass. (USA): Sinauer Associates, Inc.
Linnaeus C. 1753. Species plantarum. Vol. 1. Stockholm, Sweden.
Lu BR. 1999. Taxonomy of the genus *Oryza* (Poaceae): historical perspective and current status. Int. Rice Res. Notes 24:4-8.
Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264:421-424.
McIntyre CL, Winberg BC. 1998. A rapid means of identifying wild rice species DNA using dot blots and genome-specific rDNA probes. Genome 41:391-395.
Morinaga T. 1943. Cytogenetical studies on *Oryza sativa* L. VI. The cytogenetics of F_1 hybrids of *O. minuta* Presl. and *O. latifolia*. Jpn. J. Bot. 12:347-357.
Morinaga T. 1964. Cytogenetical investigations on *Oryza* species. In: Rice genetics and cytogenetics. Proceedings of the Symposium on Rice Genetics and Cytogenetics. Amsterdam (Netherlands): Elsevier Publishing Company. p 91-102.
Morishima H, Oka HI. 1960. The pattern of interspecific variation in the genus *Oryza*: its quantitative representation by statistical methods. Evolution 14:153-165.
Morishima H, Sano Y, Oka HI. 1992. Evolutionary studies in cultivated rice and its wild relatives. Oxford Surv. Evol. Biol. 8:135-184.
Nayar NM. 1973. Origin and cytogenetics of rice. Adv. Genet. 17:153-292.
Prodoehl A. 1922. Oryzeae monographice describuntur. Bot. Arch. 1:211-224, 231-256.

Roschevicz RI. 1931. A contribution to the study of rice. Tr. Prikl. Bot. Genet. Sel. 27:3-133.
Sampath S. 1962. The genus *Oryza*: its taxonomy and species interrelationships. Oryza 1:1-29.
Sang T, Zhang DM. 1999. Reconstructing hybrid speciation using sequences of low copy nuclear genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. Syst. Bot. 24:148-163.
Sarkar R, Raina SN. 1992. Assessment of genome relationships in the genus *Oryza* L. based on seed-protein profile analysis. Theor. Appl. Genet. 85:127-132.
Sasaki T. 1998. The rice genome project in Japan. Proc. Natl. Acad. Sci. USA 95:2027-2028.
Seberg O, Petersen G. 1998. A critical review of concepts and methods used in classical genome analysis. Bot. Rev. 64:372-417.
Second G. 1982. Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Jpn. J. Genet. 57:25-57.
Second G. 1985. A new insight into the genome differentiation in *Oryza* L. through isozymic studies. In: Sharma AK, Sharma A, editors. Advances in chromosome and cell genetics. New Delhi (India): Oxford & IBH Publishing Co. p 45-78.
Second G. 1991. Molecular markers in rice systematics and the evaluation of genetic resources. In: Bajaj YPS, editor. Biotechnology in agriculture and forestry. Berlin and Heidelberg (Germany): Springer-Verlag. p 468-494.
Sharma SD. 1986. Evolutionary trends in genus *Oryza*. In: Rice genetics. Proceedings of the International Rice Genetics Symposium, 27-31 May 1985. Los Baños (Philippines): International Rice Research Institute. p 59-67.
Sharma SD, Shastry SVS. 1965. Taxonomic studies in the genus *Oryza*. VI. A modified classification of genus. Indian J. Genet. 25:173-178.
Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF. 1998. The tortoise and the hare: choosing between non-coding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. 85:1301-1315.
Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. USA 97:7051-7057.
Stebbins GL. 1981. Coevolution of grasses and herbivores. Ann. Mo. Bot. Gard. 68:75-86.
Swofford DL. 1991. When are phylogeny estimates from molecular and morphological data incongruent? In: Miyamoto MM, Cracraft J, editors. Phylogenetic analysis of DNA sequences. New York (USA): Oxford University Press. p 295-333.
Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063-1066.
Tateoka T. 1962a. Taxonomic studies of *Oryza*. I. *O. latifolia* complex. Bot. Mag. Tokyo 75:418-427.
Tateoka T. 1962b. Taxonomic studies of *Oryza*. II. Several species complexes. Bot. Mag. Tokyo 75:455-461.
Tateoka T. 1963. Taxonomic studies of *Oryza*. III. Key to the species and their enumeration. Bot. Mag. Tokyo 76:165-173.
Tateoka T. 1965a. *Porteresia*, a new genus of Gramineae. Bull. Natl. Sci. Mus. Tokyo 8:405-406.
Tateoka T. 1965b. A taxonomy study of *Oryza eichingeri* and *O. punctata*. Bot. Mag. Tokyo 78:156-163.
Tateoka T, Pancho JV. 1963. A cytotaxonomic study of *Oryza minuta* and *O. officinalis*. Bot. Mag. Tokyo 76:366-373.
Vaughan DA. 1989. The genus *Oryza* L.: current status of taxonomy. IRRI Research Paper Series 138. Los Baños (Philippines): International Rice Research Institute.

Vaughan DA. 1994. The wild relatives of rice: a genetic resources handbook. Los Baños (Philippines): International Rice Research Institute. 137 p.
Wang ZY, Second G, Tanksley SD. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. 83:565-581.
Wendel JF. 2000. Genome evolution in polyploids. Plant Mol. Biol. 42:225-249.
Wendel JF, Albert VA. 1992. Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. 17:115-143.
Wendel JF, Doyle JJ. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis DE, Soltis PS, Doyle JJ, editors. Molecular systematics of plants II: DNA sequencing. Norwell, Mass. (USA): Kluwer Academic Publishers. p 265-296.
Wolfe AD, Liston A. 1998. Contributions of PCR-based methods to plant systematics and evolutionary biology. In: Soltis DE, Soltis PS, Doyle JJ, editors. Molecular systematics of plants II: DNA sequencing. Norwell, Mass. (USA): Kluwer Academic Publishers. p 43-86.
Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L. 1996. Genes from wild rice improve yield. Nature 384:223-224.

Notes

Authors' addresses: Song Ge, De-Yuan Hong, Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; Tao Sang, Department of Botany and Plant Pathology, Michigan State University, East Lansing, MI 48824, USA; Bao-Rong Lu, Genetic Resources Center, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines.
Acknowledgments: This work was supported by the Chinese Academy of Sciences (KZ951-B1-102), Michigan State University, and the Michigan Agricultural Experiment Station.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Miniature inverted repeat transposable elements help create genomic diversity in maize and rice

S.R. Wessler, A. Nagel, and A. Casa

Miniature inverted repeat transposable elements (MITEs) are the most prevalent elements associated with the genes of maize and rice, in which they reside in introns and in 5′ and 3′ flanking regions. Several MITEs have recently been amplified in the genomes of maize and rice, thus giving rise to structurally distinct alleles. This MITE-mediated variation can be assayed and quantified using a modification of the amplified fragment length polymorphism method called transposon display. In this way, the genome-wide distribution of MITEs can be mapped and MITE markers isolated and added to geneticists' toolboxes.

Transposable elements (TEs) were discovered more than a half century ago as the genetic agents responsible for the sectors of anthocyanin on mutant maize kernels (McClintock 1950). This seemingly benign discovery led to a revolution in our concept of the content and stability of the genetic material in virtually all living things. This revolution continues to this day.

The existence of TEs implied that genomes are not just collections of the genes necessary to assemble and maintain living organisms. In fact, we now know that the TE component can account for the largest fraction of genomic DNA in a variety of organisms including many plant species. The differential amplification of TEs has recently been shown to be responsible for the C value paradox in the grasses. The C value paradox is the observed lack of correlation between DNA content and complexity of organisms (Thomas 1971). It has been documented for both animal and plant species but, to date, only appears to be "solved" for members of the grass clade. That is, the increase in genome size from rice to sorghum to maize to barley (from 0.4 to 5 Mb) reflects, in large part, TE amplification in the larger genomes (Kumar and Bennetzen 1999).

The existence of TEs also demonstrated that the genome is dynamic and not stable as had been thought. TEs are capable of moving from one locus to another and increasing their copy number relative to other genomic components. Again, from studies with members of the grass clade, we now know that TE mobility and amplification create an amazing amount of structural diversity between members of the same species. In the past year, the analysis of adjacent populations of *Hordeum spontaneum* (wild barley) revealed dramatic differences in TE copy number (Kalendar et al 2000). We also know that the presence or absence of a TE at a particular locus creates structural allelic diversity on a massive scale in both maize and rice (Zhang et al 2000, A. Nagel and S. Wessler, unpublished data).

These and other data suggest that the grasses are in an epoch of remarkable TE activity that is creating previously unimagined levels of genomic diversity. To date, most of these studies have been carried out in maize. In this review, we summarize our studies on the miniature inverted repeat transposable elements (MITEs) of maize and rice, which are a major source of structural allelic diversity. We also discuss how the *Oryza* genus will play a central role in determining whether TE-mediated diversity is largely neutral or whether it is contributing to the variation that is fueling micro- and macroevolutionary processes.

Classes of TEs

Transposable elements are divided into two classes: class 1, or retro-elements, includes the long terminal repeat (LTR) retrotransposons, the long interspersed nuclear elements (LINEs, also known as non-LTR retrotransposons), and short interspersed nuclear elements (SINEs). For all class 1 elements, it is the element-encoded mRNA, and not the element itself, that forms the transposition intermediate. For this reason, retro-elements are capable of attaining very high copy numbers in a relatively short time frame.

Class 2 or DNA elements are characterized by short terminal inverted repeats (TIRs) and transposition via a DNA intermediate (reviewed in Kunze et al 1997). Plant DNA elements (such as *Ac*/*Ds*, *Spm*/*dSpm*, and *Mutator*) generally excise from one site and re-insert elsewhere. Class 2 elements are themselves divided into two groups. Autonomous elements such as *Ac* and *Spm* encode all of the products necessary for their transposition in maize and in several other plant species (Baker et al 1986, Yoder 1990). Nonautonomous elements such as *Ds* and *dSpm* require the presence of the autonomous elements *Ac* and *Spm*, respectively, for transposition because they are usually deleted (defective) versions of autonomous elements.

The copy number of class 2 element families is usually less than 100 per haploid genome due to their conservative mechanism of transposition. One exception to this generalization is the newly described MITEs. MITEs are a special category of nonautonomous elements that display a very high copy number (up to tens of thousands), are uniformly small (less than 500 bp), and have, in most cases, a strong target site preference (Bureau and Wessler 1992, 1994a,b, Zhang et al 2000). Although first identified in association with the genes of several plant species including maize (Bu-

reau and Wessler 1992, 1994), rice (Bureau and Wessler 1994a,b, Bureau et al 1996), green pepper (Pozueta-Romero et al 1996), and *Arabidopsis* (Casacuberta et al 1998, Surzycki and Belknap 1999, Kapitonov and Jurka 1999, Le et al 2000), MITEs are also in several animal genomes including *Caenorhabditis elegans* (Oosumi et al 1995b, Surzycki and Belknap 2000), insects (Tu 1997), fish (Izsvak et al 1999), and humans (Morgan 1995, Oosumi et al 1995a).

A MITE family is loosely defined as a group of related elements (usually with >70% sequence identity, frequently >85%) whose amplification/transposition has been catalyzed by a transposase that is encoded by a class 2 autonomous element. Figure 1 shows the hypothetical position of MITEs in a traditional transposable element family. This hierarchy is called hypothetical because strains harboring active MITEs have not yet been described. Although computer analysis of genomic sequence has identified putative autonomous elements in *Arabidopsis* (Feschotte and Mouches 2000, Le et al 2000) and *C. elegans* (Oosumi et al 1995b), no genetic relationship between a MITE family and an active autonomous element has been established.

Although all TE classes are found in higher plants, both LINEs and SINEs are less prevalent in plant than in mammalian genomes (discussed in Wessler et al 1995). Instead, LTR retrotransposons and MITEs appear to predominate in plants. LTR retrotransposons are the most abundant element in many plant genomes. The discovery that the vast intergenic regions of maize are composed largely of LTR retrotransposons led to the hypothesis that these elements preferentially target other retrotransposons for insertion (San Miguel et al 1996). A similar insertion site preference was reported for *BARE-1* of barley (Suoniemi et al 1997) and *RIRE1* of wild rice (Noma et al 1997, Kumekawa et al 1999).

MITEs appear to be the most prevalent element associated with the genic regions of higher plants, especially those in the grass clade (Bureau et al 1996, Hu et al 2000, Tarchini et al 2000). An actual preference for genic regions has been established for two maize MITEs: *Hbr* (Zhang et al 2000) and *mPIF* (N. Jiang, X. Zhang, and

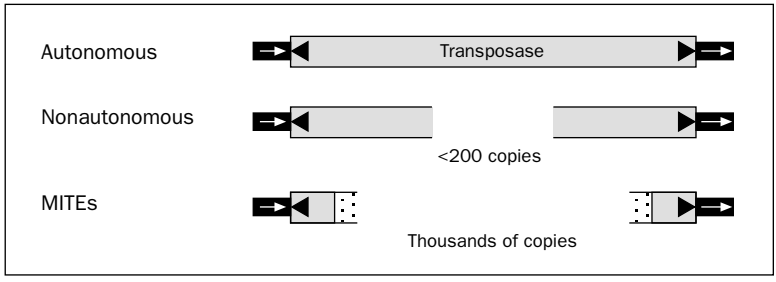


Fig. 1. Hypothetical position of miniature inverted repeat transposable elements (MITEs) in a transposable element family hierarchy. Arrows in black boxes represent the host sequence direct repeat. Arrowheads represent the terminal inverted repeat. The sequence of both repeats is hypothesized to be conserved in members of the same element family. The dotted region indicates that MITEs may not be derived from autonomous elements by internal deletion.

S. Wessler, unpublished data). Other DNA element families, including *Ac*/*Ds* and *Mutator*, also target genic, low-copy regions by a mechanism that has yet to be described (Chen et al 1992, Cresse et al 1995). It is intriguing that these previously characterized class 2 elements have not apparently given rise to MITEs. The maize *Ds1* element resembles MITEs in size and structure. However, like the rest of the *Ac/Ds* family, it has a very low copy number (<50) (Peacock et al 1984) and does not have a target sequence preference.

The TEs of *Oryza*

All classes of elements described above have been detected in *Oryza* species. A 340-kb contig around the *Adh1* gene of *O. sativa* was found to contain 28.5% repetitive DNA (Tarchini et al 2000). Based on the composition of this segment, the authors estimate that retrotransposons, DNA elements, and MITEs account for 14.4%, 8.7%, and 5.3%, respectively, of total genomic DNA. MITEs were clearly the most numerous with 78 elements (from 14 different families) or one per 4.4 kb. If this contig is representative of the rest of the genome, there could be more than 100,000 MITEs in the 438-Mb *O. sativa* genome.

The small size of the rice genome is correlated with a small percentage of retrotransposons compared with other domesticated cereals. If the 14.4% value for the amount of LTR-retrotransposons reflects the composition of the whole genome of *O. sativa* (Tarchini et al 2000), then this is much lower than the value of 50% for maize (San Miguel et al 1996). Amplification of TEs, especially retrotransposons, was shown to correlate with the almost 3-fold difference in genome sizes among *Oryza* species as determined by fluorescence *in situ* hybridization (FISH) (Uozo et al 1997). For example, *RIRE1* copy number was estimated at 180 in *O. sativa* (Nipponbare) and 7,500 in the 946-Mb genome of *O. australiensis* (Noma et al 1997, Uozo et al 1997).

Computer-assisted discovery of MITEs in rice

Bureau et al (1996) reported the results of a computer analysis of the repetitive DNA present in the 105 partial and complete rice genes then present in the GenBank and EMBL databases. Thirty-two common sequences belonging to nine putative mobile element families were found in the noncoding regions of rice genes. Seven of the nine families (*Tourist*, *Stowaway*, *Gaijin*, *Castaway*, *Ditto*, *Wanderer*, *Explorer*) had the features of MITEs. This study demonstrated that MITEs are the most common elements associated with the genes of rice. A computer-assisted search of 30 Mb of a bacterial artificial chromosome (BAC)-end sequence (Mao et al 2000) led to an estimate of the copy number of members of six MITEs [*Tourist* (~335 bp), 4,000; *Stowaway* (~245 bp), 3,000; *Gaijin* (~145 bp), 2,200; *Ditto* (~245 bp), 2,000; *Olo24* (~340 bp), 2,000; and *Castaway* (~345 bp), 1,500]. Many of these elements have greater than 85% within-family sequence identity (A. Nagel and S. Wessler, unpublished data).

MITEs as a source of allelic diversity in maize and rice

Through numerous studies, a great deal has been learned about the comparative structure of the genomes of members of the grass clade including rice, sorghum, maize, and wheat (reviewed in Bennetzen 2000). The take-home message from these studies is that grass genomes are highly conserved for coding exon sequences and gene order but completely different for the presence or absence of TEs at particular loci (defined herein as TE-insertion site polymorphism). This situation has led to speculation that TEs play a major role in creating the variation that fuels speciation in the grass clade (Zhang et al 2000).

This hypothesis cannot be tested through intergeneric comparisons or by deciphering the complete genome sequence of one strain. Although such studies are critical to understanding gene function, they do not reveal how genes or genomes evolve. To address evolutionary questions, it is necessary to first quantify the extent of TE-insertion site polymorphism between members of the same species and then to determine whether this polymorphism is functionally significant and/or generated in response to external cues. To understand the contribution of MITEs to the creation of allelic diversity, we are focusing on MITEs that have recently spread through the genome in maize and rice. Two features distinguish such MITEs: they have high within-family sequence identity (usually >90%) and their insertion sites are highly polymorphic within the species.

Several recently amplified MITEs in maize

The first member of the MITE family, *Hbr*, was isolated from a mutant allele of the maize *HM1* disease resistance gene (Johal and Briggs 1992). Subsequent studies revealed that there are about 3,000–4,000 *Hbr* elements in maize with more than 90% sequence identity (Zhang et al 2000). Although the majority of the maize genome contains moderate to highly repetitive DNA, randomly chosen *Hbr* elements are predominantly in single or low-copy genic regions. Preliminary studies have identified three additional maize MITE families each with more than 5,000 copies per haploid genome, with more than 90% sequence identity, and with target site preference: *B2-Tourist* (Bureau and Wessler 1992, N. Jiang and S. Wessler, unpublished data), *mPIF* (Walker et al 1997, X. Zhang, Q. Zhang, and S.Wessler, unpublished data), and *Hb2* (Spell et al 1988, Z. Magbanua and S. Wessler, unpublished data).

Transposon display: assaying TE polymorphism

MITE-transposon display (MITE-TD) is a modification of the amplified fragment length polymorphism (AFLP) technique and the transposon display technique (Vos et al 1995, Van den Broeck et al 1998) that permits the simultaneous detection of many MITEs from high copy number lines. Unlike conventional AFLP products that are delimited by two restriction sites, one restriction site (usually *Mse*1 or *Bfa*1) and a MITE delimit MITE-TD polymerase chain reaction (PCR) products. MITE-TD has been applied to several MITEs from maize including *Hbr* (Casa et al 2000). As with AFLP, the segregation of a MITE product that is present in the parent of one mapping

parent but not in the other can be scored in the progeny and used to determine the chromosomal location of the MITE. In this way, more than 250 *Hbr* markers were scored in a B73 × Mo17 recombinant inbred mapping population and found to be distributed relatively uniformly over the ten chromosomes of maize (Casa et al 2000). Given the genic preference of *Hbr* and other MITEs, MITE markers may be an extremely valuable addition to the mapmaker's toolbox.

A key feature in the success of MITE-TD is the identification of a suitable consensus sequence adjacent to the TIR. This sequence is employed to synthesize a primer(s) that is used in conjunction with the restriction site adapter primer in the PCR amplification steps. For the maize MITE *Hbr*, and for other maize MITEs, this sequence was derived by aligning the sequences of several related elements recovered from small insert genomic libraries (Casa et al 2000, Zhang et al 2000). In contrast, MITE primer selection is far simpler in rice, which has a large genomic sequence database. Related elements are simply identified through database searches and Boxshade alignments, like that shown for the MITE *Olo* (Fig. 2A). In this way, suitable primers can be easily designed and employed in TD (Fig. 2B). Database-assisted primer design led to successful TDs for several other MITEs in the rice genomes (Fig. 2B).

Oryza as a model genus to study TEs and evolution

The availability of the entire sequence of *Oryza sativa* cv. Nipponbare will permit identification of virtually all TEs and their relative positions with respect to known genes and open reading frames. This information can be used to understand how different TE families become established, amplified, and spread throughout the genome. The genome sequence can also be used to derive primers for transposon display of the high copy number MITEs and retrotransposons. Transposon display, in turn, when used in conjunction with the 20 or more species of *Oryza*, will provide data on the contribution of each TE family to intrageneric diversity. It is hoped that, eventually, these data will furnish the raw material for determining whether TEs are a significant factor in generating the diversity that fuels both natural and artificial selection.

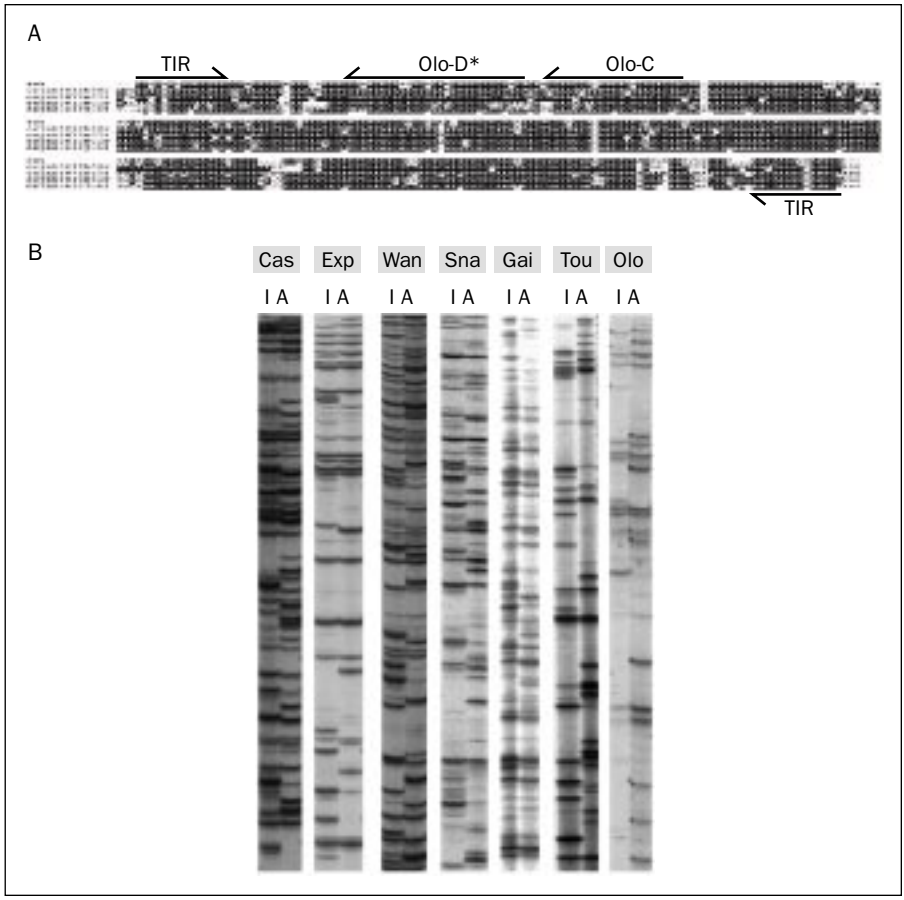


Fig. 2. Transposon display using rice miniature inverted repeat transposable elements (MITEs) identified through database searches. (A) Boxshade alignment of *Olo* elements. TIR = terminal inverted repeat; Olo-C = position of element-specific primer used in preselective amplification of transposon display; Olo-D = labeled element-specific primer used in selective amplification of transposon display (see Casa et al 2000 for methodology). (B) Autoradiograph of transposon display polymerase chain reaction products resolved on DNA sequencing gels (Casa et al 2000) for several MITEs from *Oryza sativa*. MITEs: Cas = *Castaway*, Exp = *Explorer*, Wan = *Wanderer*, Sna = *Snabo-4*, Gai = *Gaijin*, Tou = *Tourist-os6*, Olo = *Olo*. Genomic DNA from IR64 (*O. sativa* subsp. *indica*) or A = Azucena (*O. sativa* subsp. *japonica*) was digested with either *Mse*1 or *Bfa*1.

References

Baker B, Schell J, Lorz H, Fedoroff N. 1986. Transposition of the maize controlling element *Activator* in tobacco. Proc. Natl. Acad. Sci. USA 83:4844-4848.

Bennetzen JL. 2000. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12:1021-1030.

Bureau TE, Ronald PC, Wessler SR. 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc. Natl. Acad. Sci. USA 93:8524-8529.

Bureau TE, Wessler SR. 1992. *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283-1294.

Bureau TE, Wessler SR. 1994. Mobile inverted-repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. Proc. Natl. Acad. Sci. USA 91:1411-1415.

Bureau TE, Wessler SR. 1994. *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6:907-916.

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR. 2000. The MITE family *Heartbreaker (Hbr)*: molecular markers in maize. Proc. Natl. Acad. Sci. USA 97:10083-10089.

Casacuberta E, Casacuberta JM, Puigdomenech P, Monfort A. 1998. Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterization of the *Emigrant* family of elements. Plant J. 16:79-85.

Chen J, Greenblatt IM, Dellaporta SL. 1992. Molecular analysis of *Ac* transposition and DNA replication. Genetics 130:665-676.

Cresse AD, Hulbert SH, Brown WE, Lucas JR, Bennetzen JL. 1995. *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. Genetics 140:315-324.

Feschotte C, Mouches C. 2000. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. Mol. Biol. Evol. 17:730-737.

Hu J, Reddy VS, Wessler SR. 2000. The rice *R* gene family: two distinct subfamilies containing several miniature inverted-repeat transposable elements. Plant Mol. Biol. 42:667-678.

Izsvak Z, Ivics Z, Shimoda N, Mohn D, Okamoto H, Hackett PB. 1999. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J. Mol. Evol. 48:13-21.

Johal GS, Briggs SP. 1992. Reductase activity encoded by the *HM1* disease resistance gene in maize. Science 258:985-987.

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA 97:6603-6607.

Kapitonov VV, Jurka J. 1999. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107:27-37.

Kumar A, Bennetzen JL. 1999. Plant retrotransposons. Annu. Rev. Genet. 33:479-532.

Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E. 1999. Identification and characterization of novel retrotransposons of the *gypsy* type in rice. Mol. Gen. Genet. 260:593-602.

Kunze R, Saedler H, Lonnig WE. 1997. Plant transposable elements. Adv. Bot. Res. 27:331-470.

Le QH, Wright S, Yu Z, Bureau T. 2000. Transposon diversity in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 97:7376-7381.

Mao L, Wood TC, Yeisoo Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frisch D, Goff S, Dean RA, Wing RA. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res. 10:982-990.

McClintock B. 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36:344-355.

Morgan GT. 1995. Identification in the human genome of mobile elements spread by DNA-mediated transposition. J. Mol. Biol. 254:1-5.

Noma K, Nakajima R, Ohtsubo H, Ohtsubo E. 1997. *RIRE1*, a retrotransposon from wild rice *Oryza australiensis*. Genes Genet. Syst. 72:131-140.

Oosumi T, Belknap WR, Garlick B. 1995a. *Mariner* transposons in humans. Nature 378:672.

Oosumi T, Garlick B, Belknap WR. 1995b. Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 92:8886-8890.

Peacock WJ, Dennis ES, Gerlach WL, Sachs MM, Schwartz D. 1984. Insertion and excision of *Ds* controlling elements in maize. Cold Spring Harbor Symp. Quant. Biol. 49:347-354.

Pozueta-Romero J, Houlne G, Schantz R. 1996. Nonautonomous inverted repeat *Alien* transposable elements are associated with genes of both monocotyledonous and dicotyledonous plants. Gene 171:147-153.

San Miguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765-768.

Spell M, Baran G, Wessler SR. 1988. A RFLP adjacent to the maize *waxy* gene has the structure of a transposable element. Mol. Gen. Genet. 211:364-366.

Suoniemi A, Schmidt D, Schulman A. 1997. *BARE*-1 insertion site preference and evolutionary conservation of RNA and cDNA processing sites. Genetica 100:219-230.

Surzycki SA, Belknap WR. 1999. Characterization of repetitive DNA elements in *Arabidopsis*. J. Mol. Evol. 48:684-691.

Surzycki SA, Belknap WR. 2000. Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. Proc. Natl. Acad. Sci. USA 97:245-249.

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A. 2000. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381-391.

Thomas CA. 1971. The genetic organization of chromosomes. Annu. Rev. Genet. 5:237-256.

Tu Z. 1997. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. Proc. Natl. Acad. Sci. USA 94:7475-7480.

Uozo S, Ikehashi H, Ohmido N, Ohtsubo H, Ohtsubo E, Fukui K. 1997. Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. Plant Mol. Biol. 35:791-799.

Van den Broeck D, Maes T, Sauer M, Zethof J, De Keukeleire P, D'Hauw M, Van Montagu M, Gerats T. 1998. *Transposon Display* identifies individual transposable elements in high copy number lines. Plant J. 13:121-129.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters BA, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP, a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Walker EL, Eggleston WB, Demopulos D, Kermicle J, Dellaporta SL. 1997. Insertions of a novel class of transposable elements with a strong target site preference at the *r* locus of maize. Genetics 146:681-693.

Wessler SR, Bureau TE, White SE. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5:814-821.

Yoder JI. 1990. Rapid proliferation of the maize transposable element *Activator* in transgenic tomato. Plant Cell 2:723-730.

Zhang Q, Arbuckle J, Wessler SR. 2000. Recent, extensive and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker (Hbr)* into genic regions of maize. Proc. Natl. Acad. Sci. USA 97:1160-1165.

Notes

Authors' addresses: S.R. Wessler, A. Nagel, Department of Botany, University of Georgia, Athens, GA 30602, USA, phone: 706-542-1870, fax: 706-542-1805, E-mail: sue@dogwood.botany.uga.edu; A. Casa, present address: Institute for Genomic Diversity and Department of Plant Breeding, Cornell University, 153 Biotechnology Building, Ithaca, NY 14853, USA.

Acknowledgments: We thank Svieta Temnykh and Susan McCouch for rice genomic DNA used for transposon display and Cedric Feschotte for critical reading of the manuscript. MITE research is supported by grants from the National Institutes of Health, Pioneer Hi-Bred Int. (to S.R.W.), and a National Science Foundation Research Training Grant (to A.N).

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Microsatellite markers in rice: abundance, diversity, and applications

S.R. McCouch, S. Temnykh, A. Lukashova, J. Coburn, G. DeClerck, S. Cartinhour, S. Harrington, M. Thomson, E. Septiningsih, M. Semon, P. Moncada, and Jiming Li

A well-distributed set of 500 microsatellite markers has been genetically mapped onto the rice genome. These markers link the genetic and physical maps with the genomic sequence of rice, facilitating studies that seek to determine the relationship between the structure and function of genes and genomes. To facilitate the development of new microsatellite markers using publicly available DNA sequence information, a simple sequence repeat identification tool (SSRIT) has been developed for semiautomated identification of nonredundant simple sequence repeat (SSR) loci and for primer design (available at http://www.gramene.org/microsat/). Using this script, a total of 57.8 Mb of DNA sequence from rice was searched to determine the frequency and distribution of different SSRs in the genome. Because the length of the SSR unit in any single genome has proven to be a reasonably good predictor of overall polymorphism in related genotypes, SSR loci were categorized into two groups (class I and class II) based on the length of the repeat motif. Microsatellites with poly(AT)n repeats represented the most abundant and polymorphic class of SSRs but were frequently associated with the Micropon family of miniature inverted repeat transposable elements (MITEs) and were difficult to amplify. Estimates of total microsatellite frequencies in rice suggested that there were approximately 28,340 Class I and 70,530 Class II SSRs, or about 100,000 di-, tri-, and tetra-nucleotide SSR motifs in the rice genome. The distribution of SSR sequences showed that regions of the rice genome that were richer in expressed genes also tended to be richer in SSR sequences, underscoring their usefulness as genetic markers. Applications of SSRs in variety protection, diversity analysis, gene and QTL identification, marker-assisted selection, physical mapping, and gene isolation will be discussed.

Simple sequence repeats (SSRs), commonly referred to as microsatellites, are tandemly repeated mono- to hexa-nucleotide motifs that are ubiquitous in eukaryotic genomes and exhibit highly variable numbers of repeats at a locus. Their abundance and

hypervariability make them valuable as genetic markers (Weber and May 1989). In rice, more than 500 microsatellite markers have been developed and used to construct genetic maps (Wu and Tanksley 1993, Akagi et al 1996, Panaud et al 1995, 1996, Chen et al 1997, Temnykh et al 2000, 2001). These markers provide important codominant landmarks that are well distributed throughout the rice genome. SSRs have been widely used in rice genetics to fingerprint accessions, analyze diversity, identify introgressions in interspecific crosses, trace pedigrees, locate genes and QTLs on rice chromosomes, and in marker-assisted selection. In sequencing and genomics applications, SSRs are increasingly useful as sequence tag connectors (STCs) linking the physical and genetic maps of rice.

SSR marker development

In the past, the advantages of microsatellite markers were partially offset by the difficulty inherent in marker development, as laborious iterations of genomic DNA library screening with radio-labeled probes were required to isolate SSR-containing sequences. In rice, the growing pool of DNA sequence information being generated by the International Rice Genome Sequencing Project (IRGSP) and by independent organizations (see, for example, www.rice-research.org and www.syngenta.com) provides the basis for efficient, high-throughput *in silico* identification of SSR loci. The public availability of genomic sequences has dramatically altered the strategies used for marker development and has largely removed the technical and economic limitations that previously limited the number of SSR loci that were available for rice.

To facilitate the development of new microsatellite markers using publicly available DNA sequence information, a set of computer programs has been developed for semiautomated identification of nonredundant SSR loci and for primer design. The script consists of two parts and is publicly available at http://ars-genome.cornell.edu/rice/tools.html. The searching script identifies all microsatellites of a user-defined minimum repeat length in a set of downloaded sequences; records the database ID (i.e., GenBank accession number), motif type, number of perfect repeats, and sequence coordinates for each SSR; does a redundancy search based on flanking sequence identity; and reports the results in an output file. Primers are then designed using the program Primer 0.5 (Daly et al 1991). The search can be used to identify SSRs in different types and lengths of DNA sequence, including relatively short sequences (300–700 bp) derived from the ends of cDNA clones (known as expressed sequence tags, or ESTs), or from the ends of rice genomic DNA cloned into bacterial artificial chromosomes (BACs), or long contigs assembled from fully sequenced BAC and P1-derived artificial chromosome (PAC) clones that may be up to 1 Mb in length (Temnykh et al 2001).

Based on information about the frequency and average length of specific SSR motifs in rice, classes of SSRs most likely to provide good genome coverage and high levels of polymorphism can be selected for primer design and used as DNA markers. Akagi et al (1996) first noted that AT-rich microsatellites tended to show more length variation than tracts consisting of GC-rich motifs and suggested that these would

make the best SSR markers for rice. Cho et al (2000) demonstrated that, on average, long SSRs tend to be more polymorphic than shorter ones, no matter what the motif is, and this characteristic was in agreement with studies in human and other species (Weber 1990, Goldstein and Clark 1995).

Length classification of SSRs and relationship to polymorphism

Because the length of the SSR repeat unit in any single genome has proven to be a reasonably good predictor of overall polymorphism in related genotypes, microsatellite loci in this study were categorized into two groups based on the length of the repeat motif. As summarized in Figure 1, class I consists of hypervariable markers, classified as perfect SSRs ≥20 nt in length, and class II consists of potentially variable markers, defined as SSRs ≥12 nt and <20 nt in length. While these classes were derived empirically in this study, Tautz and Ranz (1984) provided a mechanistic explanation for this phenomenon. They suggested that long perfect repeats are a difficult template for DNA polymerase II to replicate faithfully and thus the enzyme is prone to replication slippage on long tracts, resulting in an accelerated rate of mutation for long SSRs. Shorter or imperfect SSRs represent sites where SSR expansion may begin, or may have already occurred in other genotypes, but these sites are less frequently the targets of replication slippage. Motifs shorter than 12 bp demonstrate a mutation potential that is no different from that of most other types of nonrepetitive sequence, and therefore they are not considered to be SSRs (Pupko and Graur 1999).

Estimates of SSR frequency in the rice genome

The first estimates of microsatellite abundance in rice were based on hybridization assays using clone libraries (Wu and Tanksley 1993, Panaud et al 1996). These ex-

	Length	Repeat number		
		di	tri	tetra
Class I (variable)	≥20 nt	≥10	≥7	≥5
Class II (potential)	12–20 nt	6–9	4–6	3–4
Class III (stochastic)	6–12 nt	3–5	2–3	2

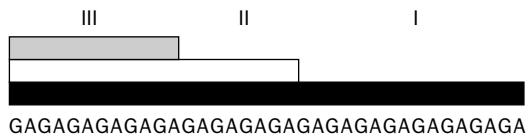


Fig. 1. Classification of SSRs by repeat length into class I (≥20 nt), class II (12–20 nt), and class III (6–12 nt), with the corresponding number of repeat units indicated for di-, tri-, and tetra-nucleotide motifs.

periments tended to detect only the longer SSR motifs (corresponding to class I microsatellites) because of the hybridization conditions used. Estimates from this early work predicted that there were from 5,500 to 10,000 microsatellite loci in rice, or one every 40–80 kb. Using a computational approach, Temnykh et al (2001) examined 47,430 kb of BAC end sequence, or about 0.11 genome equivalent, the largest single source of rice sequence information available at the time of the study. Extrapolating from the frequency of di-, tri-, and tetra-nucleotide SSRs observed in this single-pass BAC end sequence, it was predicted that the rice genome contained approximately 11,000 class I and an additional 22,000 class II microsatellites, or a total of approximately 33,000 di-tetranucleotide SSRs. These numbers are extrapolations of the observed frequencies summarized in Table 1. Two other independent sources of rice sequence were also examined by Temnykh et al (2001), namely, 27 fully sequenced BAC and PAC clones and 12,000 nonredundant sequenced cDNAs (expressed sequence tags, ESTs). This provided about 4,000 kb of continuous genomic sequence and 6,000 kb of ESTs. Estimates of total microsatellite frequencies in these sequences were approximately three times those based on BAC end sequences (Fig. 2), with one class I SSR observed every 15 kb and one class II SSR every 6 kb in the totally sequenced PACs and BACs, suggesting a total of approximately 28,340 class I and 70,530 class II SSRs, or about 100,000 SSR motifs in the rice genome (Table 1). As additional sequence information becomes available for rice, the accuracy of these estimates will improve.

Table 1. Observed number of microsatellites with di-, tri-, and tetra-nucleotide repeats in three sources of DNA sequence data: (A) BAC ends, (B) fully sequenced PAC and BAC clones, and (C) ESTs.

SSR category	Category I (n≥20 nt)	Category II (12< n <20)	All SSRs with n >12 nt
A) 74,127 BAC ends (47,430 kb)			
di	690 (1/69 kb)	969 (1/49 kb)	1,659 (1/29 kb)
tri	210 (1/226 kb)	1,035 (1/46 kb)	1,245 (1/38 kb)
tetra	289 (1/164 kb)	410 (1/115 kb)	699 (1/68 kb)
Total (observed)	1,189 (1/40 kb)	2,414 (1/20 kb)	3,603 (1/16 kb)
Predicted total in rice genome	10,780	21,885	32,665
B) 27 PACs and BACs (4,036 kb)			
di	132 (1/30.5 kb)	173 (1/23 kb)	305 (1/13 kb)
tri	97 (1/41.6 kb)	382 (1/11 kb)	479 (1/8 kb)
tetra	37 (1/109 kb)	107 (1/38 kb)	144 (1/28 kb)
Total (observed)	266 (1/15 kb)	662 (1/6 kb)	928 (1/4 kb)
Predicted total in rice genome	28,340	70,530	98,870
C) 12,532 ESTs (approx. 6,000 kb)			
di	74 (1/81 kb)	159 (1/38 kb)	233 (1/26 kb)
tri	235 (1/26 kb)	1,159 (1/5 kb)	1,394 (1/4 kb)
tetra	32 (1/188 kb)	No data	32
Total (observed)	341 (1/18 kb)	1,318 (1/5 kb)	1,659 (1/3 kb)
Predicted total in rice genome	24,440	94,460	118,900

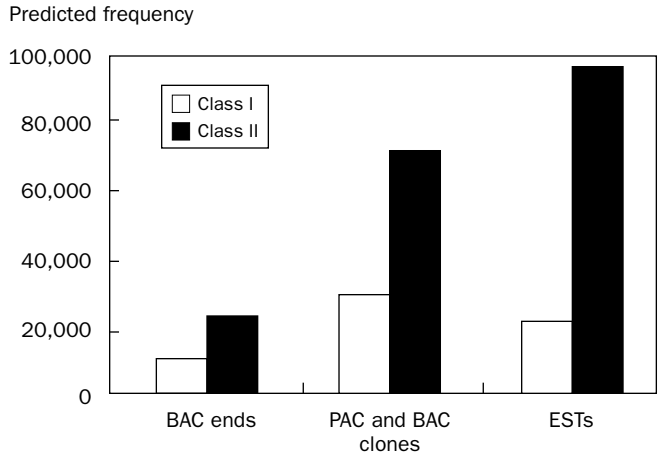


Fig. 2. Predicted frequency of class I and class II microsatellites in three sets of DNA sequence data: (1) BAC ends, (2) fully sequenced PAC and BAC clones, and (3) ESTs.

Although several thousand new SSR-containing sequences were recently released by Monsanto/Pharmacia (www.riceresearch.org), this information is not directly comparable to results from this study because the information included only SSRs ≥ 24 bp in length. In summary, current estimates based on available genomic sequence information of between 10,000 and 30,000 class I SSRs are slightly higher than previous predictions based on hybridization experiments, with a total prediction of up to 100,000 class I and class II SSRs in rice.

Frequency of SSRs in different fractions of the rice genome

We were interested in exploring why continuous genomic sequence appeared to be richer in SSRs than BAC end sequence. To address this question, we first compared the relative frequency of SSRs in coding and noncoding sequences, based on an analysis of the frequency of class I and class II SSRs in rice ESTs. The estimated frequency of microsatellites in ESTs was even higher than that in continuous genomic sequence (Fig. 2). This suggested (1) that SSRs are abundant in genic regions, (2) that the fully sequenced BAC and PAC clones that were available were gene-rich, and (3) that the BAC end sequences were relatively gene-poor. These conclusions are consistent with the fact that BAC end sequences, bounded by *Eco*RI (GAATTC) and *Hin*dIII (AAGCTT) restriction sites, are relatively AT-rich compared with the fully sequenced BAC and PAC clones or the ESTs. Although more precise estimates of class I and class II SSR frequency in rice await the completion of the rice genomic sequence, the fact that regions of the rice genome that are richer in expressed genes also tend to be richer in SSR sequences has important implications for the usefulness of SSRs as genetic markers.

Distribution of SSR motifs

Significant differences were observed when the frequencies of different class I SSR motifs extracted from the three independent sources of sequence data were compared, that is, BAC end sequences, completely sequenced BAC and PAC clones, and rice ESTs. As illustrated in Figure 3, the greatest differences were observed for GC-rich poly tri-nucleotides and poly(AT)n microsatellites. These observations suggested that SSR motif categories were not randomly distributed. In agreement with a previous study in rice (Cho et al 2000) and in maize (Chin et al 1996), GC-rich tri-nucleotide SSRs appear to be concentrated in coding regions, whereas (AT) di-nucleotides are scarce in EST sequence but abundant in and around the AT-rich intergenic regions represented by *Eco*RI- and *Hin*dIII-digested BAC end sequences. The relatively equal frequency of GC-rich tri- and poly(AT)n di-nucleotides in the fully sequenced BAC and PAC clones suggested that these clones have interspersed regions of coding and noncoding sequence.

When different motifs were compared, (AT)-rich tri-nucleotide and poly(GA)n microsatellites were unusual in that their frequencies were roughly equivalent in all three types of sequence. This observation is consistent with the idea that these motifs are relatively uniformly distributed in the rice genome and make good targets for SSR marker development. In contrast, poly(GC)n motifs were extremely rare in all sequence data and poly(CA)n microsatellites occurred at low frequencies (Fig. 3).

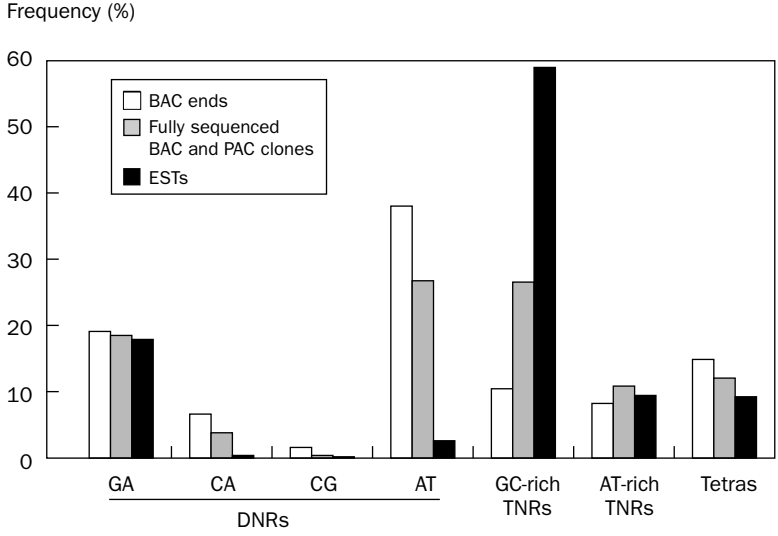


Fig. 3. Comparative frequencies of class I microsatellites from different sources of DNA sequence data classified by SSR motif category. Di-nucleotide repeats (DNR): poly(GA)n, poly(CA)n, poly(CG)n, poly(AT)n; tri-nucleotide repeats (TNR): GC-rich and AT-rich; and tetra-nucleotide repeats (tetramers).

Relationship between SSRs and genes

To further characterize the spatial relationship between class I microsatellites and genes, the distribution of SSRs was investigated in relation to open reading frames (ORFs), untranslated 3′and 5′ regions (UTRs), introns, and intergenic regions. From this analysis, we observed that about 80% of GC-rich tri-nucleotide repeats occurred in predicted exons, whereas AT-rich tri-nucleotide SSRs were distributed in all four genomic domains in this dataset. As reported by Temnykh et al (2001), poly di-nucleotide and poly tetra-nucleotide SSRs were predominantly situated in noncoding regions—preferentially in intergenic regions, and only more rarely in introns.

At a higher level of resolution, 27 SSRs found in expressed sequences were examined to determine their precise position in cloned and completely sequenced genes. Cho et al (2000) reported that 13 of the SSRs were positioned in UTRs, 6 were found in introns, and 8 were found in exons or ORFs. The eight markers in ORFs were all GC-rich tri-nucleotide repeats and showed very low levels of polymorphism. On the other hand, microsatellite sequences located in introns or in 5′ and 3′ UTRs were variable in motif (mostly poly(GA) or AT-rich di- and tri-nucleotide motifs) and tended to be much more polymorphic. These observations supported the hypothesis that SSRs found in ORFs experience functional constraints to their variability and therefore the most polymorphic microsatellite sequences are unlikely to be found in the coding portions of genes.

Strategies for successful marker design

Given the interest of our laboratory in developing a high-resolution set of microsatellite markers for use in genetics and breeding, we aimed to develop an efficient strategy for selecting useful SSR markers for this purpose. As outlined in Temnykh et al (2001), such a strategy was based on observations about the frequency, size variation potential, and PCR-amplification properties of different types of rice SSRs. Based on our experience, the highest rate of successful amplification was achieved for poly(GA)n, poly(GAA)n, and poly(CAT)n microsatellites in rice, and the first two classes were also highly polymorphic. Though poly(AT)n blocks were the most abundant and variable microsatellite sequences in the rice genome, primers designed from regions flanking this SSR motif frequently failed to amplify. Yet, primers for the commonly encountered compound motif, poly(TA)n(CA)n, performed better. In fact, the complex poly(TA)n(CA)n microsatellites had the longest runs of uninterrupted repeats and demonstrated the highest level of allelic diversity in our panel. Other types of motifs, such as poly(CA)n, poly-tetra-nucleotides, and GC-rich poly-tri-nucleotides, amplified moderately well but tended to have fewer alleles and lower polymorphism information content (PIC) values than the other classes.

Association between SSRs and transposable elements

Poly(AT)n SSRs were frequently found in association with dispersed miniature inverted repeat transposable elements (MITEs) in rice, and this was postulated to help explain the low amplification frequency of this particular motif (Temnykh et al 2001). MITE sequences that reside in regions flanking the poly(AT)n motifs may form hairpin structures that make it difficult for SSR primer pairs to bind to the DNA. When binding does occur, primers in these regions may recognize many targets dispersed throughout the genome and this results in weak or smeared amplification patterns. As a consequence of this finding in rice, BAC end sequences homologous to known repetitive elements or to other BAC end sequences are considered poor templates for SSR marker development.

Genetic mapping with SSR markers

Currently, more than 500 microsatellite markers have been integrated into the existing genetic map of rice (Fig. 4). This provides an average density of one SSR marker every 4 cM on the IR64/Azucena doubled-haploid (DH) map (Guiderdoni et al 1992, Huang et al 1994). As more genomic sequence information becomes available for rice, additional polymorphic microsatellite markers can be easily designed to saturate the genetic map at predicted densities of one SSR per 20–50 kb, and, in regions of particular interest or where overall polymorphism is not a limiting factor, higher density coverage (up to one SSR per 3–10 kb) is likely. Most SSR markers can be mapped as co-dominant single-locus markers and have easily scoreable banding patterns. Among the markers mapped in our laboratory, a total of ten (RM4, RM20, RM81, RM233, RM238, RM456, RM464, RM465, RM473, and RM476) amplified complex banding patterns, which segregated independently and mapped to multiple loci.

All markers genetically mapped in our laboratory were assigned RM locus names according to the following nomenclature guidelines: RM1–100 indicate markers from a randomly sheared library made from cv. IR36 (Panaud et al 1996) (with RM1–59 for di-nucleotides, RM60s for tetra-nucleotides, and RM70s and RM80s for ATT and TCT, respectively); RM101–199 indicate markers derived from EST sequences in GenBank; RM201–350 indicate markers from the *Tsp*509-digested library, with RM201–320 for poly-di-nucleotides and RM321–345 for poly-tri-nucleotides; RM346–351 indicate markers from other genomic libraries; and RM 400–600 indicate markers derived from BAC end sequences. Markers that mapped to more than one locus were given a suffix (A, B, C) following the RM designation.

Information about all microsatellite markers developed in the author's laboratory is available at the Gramene Web site (http://www.gramene.org/). Available data include the locus name, GenBank ID, accession number, clone name, SSR motif description, primer sequences, and polymorphism survey results, including allele number, PIC value, and allele molecular weight range based on a panel of 13 *O. sativa* genotypes described in Cho et al (2000). Markers identified in our automated searches that showed sequence similarity to those previously reported by Akagi et al (1996)

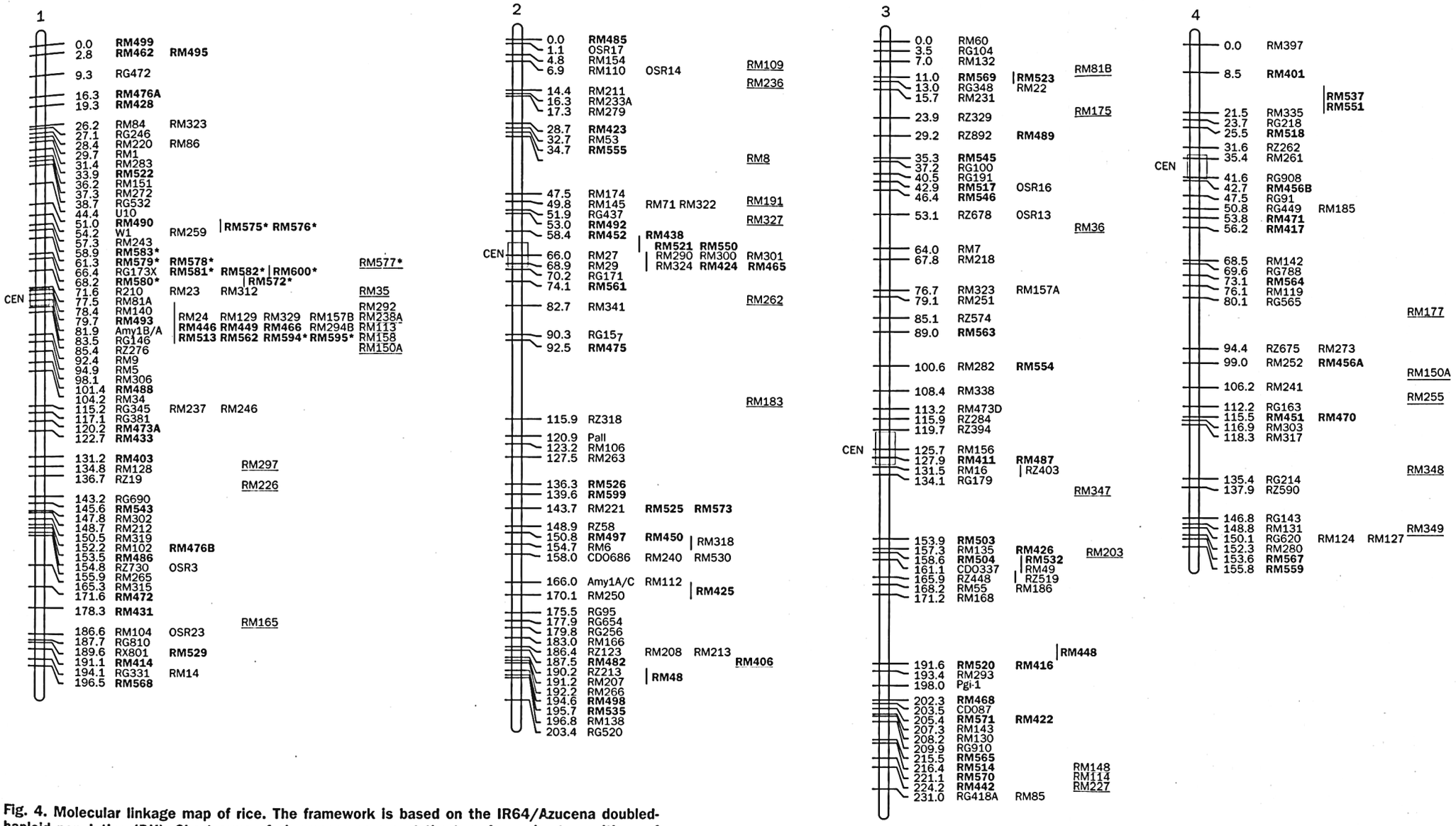
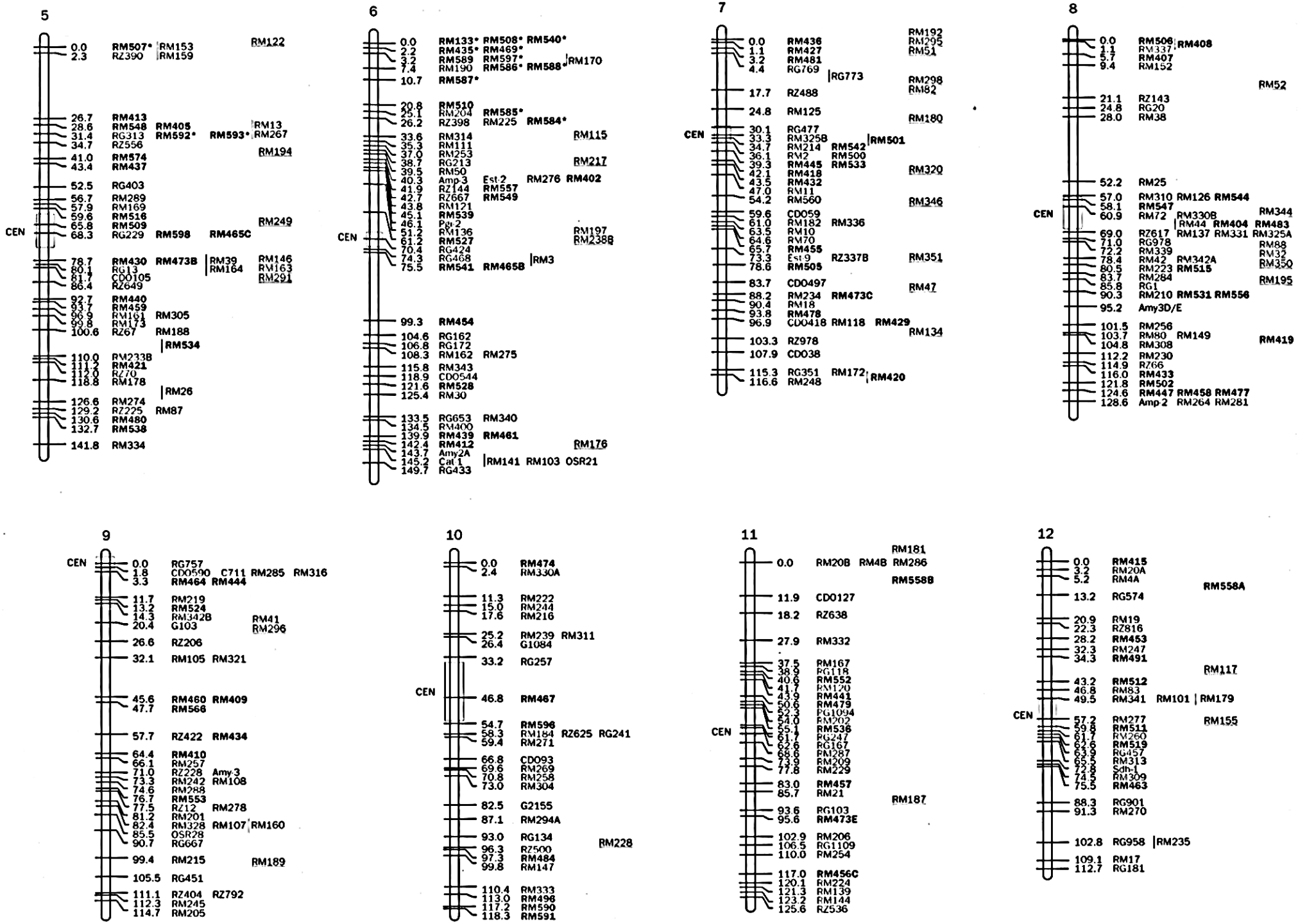


Fig. 4. Molecular linkage map of rice. The framework is based on the IR64/Azucena doubled-haploid population (DH). Short arms of chromosomes are at the top. Approximate positions of centromeres are indicated by CEN with an open box. Framework markers (those ordered at LOD score >2.0) have tick marks on chromosome bars. Co-segregating markers with absolute linkage are in the same row. Vertical lines delimit probable intervals for markers mapped with low LOD score. Markers mapped onto another population are underlined and placed to the side of the DH map based on their position in relation to common markers. The abbreviation RM is used for *R*ice *M*icrosatellite markers developed in the Cornell University laboratory. OSR loci correspond to microsatellite markers previously reported by Akagi et al (1996). New RM markers derived from BAC end sequences are shown in bold; those from fully sequenced large-insert clones are in boldface and marked by asterisks. Reproduced from Temnykh et al (2001) with permission from *Genome Research*.


5
0.0 RM507* RM153 RM122
2.3 RZ390 RM159
26.7 RM413
28.6 RM548 RM405 RM13
31.4 RG313 RM592* RM593* RM267
34.7 RZ556 RM194
41.0 RM574
43.4 RM437
52.5 RG403
56.7 RM289
57.9 RM169
59.6 RM516 RM249
65.8 RM509
CEN 68.3 RG229 RM598 RM465C
78.7 RM430 RM473B RM39 RM146
80.1 RG13 RM164 RM163
81.7 CD0105 RM291
86.4 RZ649
92.7 RM440
93.7 RM459
96.9 RM161 RM305
99.8 RM173
100.6 RZ67 RM188
RM534
110.0 RM233B
111.2 RM421
112.0 RZ70
118.8 RM178
RM26
126.6 RM274
129.2 RZ225 RM87
130.6 RM480
132.7 RM538
141.8 RM334
6
0.0 RM133* RM508* RM540*
2.2 RM435* RM469*
3.2 RM589 RM597* RM170
7.4 RM190 RM586* RM588*
10.7 RM587*
20.8 RM510
25.1 RM204 RM585*
26.2 RZ398 RM225 RM584*
33.6 RM314 RM115
35.3 RM111
37.0 RM253
38.7 RG213 RM217
39.5 RM50
40.3 Amp-3 Est-2 RM276 RM402
41.9 RZ144 RM557
42.7 RZ667 RM549
43.8 RM121
45.1 RM539
46.1 Pgi-2
CEN 51.2 RM136 RM197
61.2 RM527 RM238B
70.4 RG424
74.3 RG468 RM3
75.5 RM541 RM465B
99.3 RM454
104.6 RG162
106.8 RG172
108.3 RM162 RM275
115.8 RM343
118.9 CD0544
121.6 RM528
125.4 RM30
133.5 RG653 RM340
134.5 RM400
139.9 RM439 RM461
142.4 RM412 RM176
143.7 Amy2A
145.2 Cat 1 RM141 RM103 OSR21
149.7 RG433
7
RM192
RM295
0.0 RM436 RM51
1.1 RM427
3.2 RM481
4.4 RG769 RG773
RM298
RM82
17.7 RZ488
24.8 RM125
RM180
30.1 RG477
CEN 33.3 RM3258 RM501
34.7 RM214 RM542
36.1 RM2 RM500
39.3 RM445 RM533
42.1 RM418 RM320
43.5 RM432
47.0 RM11
54.2 RM560 RM346
59.6 CD059
61.0 RM182 RM336
63.5 RM10
64.6 RM70
65.7 RM455
73.3 Est-9 RZ337B RM351
78.6 RM505
83.7 CD0497 RM47
88.2 RM234 RM473C
90.4 RM18
93.8 RM478
96.9 CD0418 RM118 RM429 RM134
103.3 RZ978
107.9 CD038
115.3 RG351 RM172 RM420
116.6 RM248
8
0.0 RM506 RM408
1.1 RM337
5.7 RM407
9.4 RM152
RM52
21.1 RZ143
24.8 RG20
28.0 RM38
52.2 RM25
57.0 RM310 RM126 RM544
58.1 RM547
CEN 60.9 RM72 RM330B RM344
RM44 RM404 RM483
69.0 RZ617 RM137 RM331 RM325A
71.0 RG978 RM88
72.2 RM339 RM32
78.4 RM42 RM342A RM350
80.5 RM223 RM515
83.7 RM284 RM195
85.8 RG1
90.3 RM210 RM531 RM556
95.2 Amy3D/E
101.5 RM256
103.7 RM80 RM149 RM419
104.8 RM308
112.2 RM230
114.9 RZ66
116.0 RM433
121.8 RM502
124.6 RM447 RM458 RM477
128.6 Amp-2 RM264 RM281
9
CEN 0.0 RG757
1.8 CD0590 C711 RM285 RM316
3.3 RM464 RM444
11.7 RM219
13.2 RM524
14.3 RM342B RM41
20.4 G103 RM296
26.6 RZ206
32.1 RM105 RM321
45.6 RM460 RM409
47.7 RM566
57.7 RZ422 RM434
64.4 RM410
66.1 RM257
71.0 RZ228 Amy-3
73.3 RM242 RM108
74.6 RM288
76.7 RM553
77.5 RZ12 RM278
81.2 RM201
82.4 RM328 RM107 RM160
85.5 OSR28
90.7 RG667
99.4 RM215 RM189
105.5 RG451
111.1 RZ404 RZ792
112.3 RM245
114.7 RM205
10
0.0 RM474
2.4 RM330A
11.3 RM222
15.0 RM244
17.6 RM216
25.2 RM239 RM311
26.4 G1084
33.2 RG257
CEN 46.8 RM467
54.7 RM596
58.3 RM184 RZ625 RG241
59.4 RM271
66.8 CD093
69.6 RM269
70.8 RM258
73.0 RM304
82.5 G2155
87.1 RM294A
93.0 RG134 RM228
96.3 RZ500
97.3 RM484
99.8 RM147
110.4 RM333
113.0 RM498
117.2 RM590
118.3 RM591
11
RM181
0.0 RM20B RM4B RM286
RM558B
11.9 CD0127
18.2 RZ638
27.9 RM332
37.5 RM167
38.9 RG118
40.6 RM552
41.7 RM120
43.9 RM441
50.6 RM479
52.3 PG1094
54.0 RM202
CEN 55.1 RM536
61.7 PG247
62.6 RG167
68.6 RM287
73.9 RM209
77.8 RM229
83.0 RM457
85.7 RM21 RM187
93.6 RG103
95.6 RM473E
102.9 RM206
106.5 RG1109
110.0 RM254
117.0 RM456C
120.1 RM224
121.3 RM139
123.2 RM144
125.6 RZ536
12
0.0 RM415
3.2 RM20A
5.2 RM4A RM558A
13.2 RG574
20.9 RM19
22.3 RZ816
28.2 RM453
32.3 RM247
34.3 RM491 RM117
43.2 RM512
46.8 RM83
49.5 RM341 RM101 RM179
CEN 57.2 RM277 RM155
59.8 RM511
61.7 RM260
62.6 RM519
63.9 RG457
65.5 RM313
72.8 Sdh-1
74.5 RM309
75.5 RM463
88.3 RG901
91.3 RM270
102.8 RG958 RM235
109.1 RM17
112.7 RG181

are indicated with RM locus designations and primer sequences designed and tested in our laboratory, along with previously reported OSR names and links to the original GenBank and DDBJ entries. Approximately 500 bp of sequence from all SSR-containing genomic clones (summarized by Chen et al 1997 and Temnykh et al 2000) are also available in GenBank. These sequences were derived from the screening of small-insert clone libraries from IR36 and provide an immediate link to genomic sequence from cv. Nipponbare and an *in silico* view of polymorphism between indica and japonica cultivars.

Applications of SSRs to evolutionary studies, genetics, and breeding

Several useful reviews summarize the advantages of using SSRs in plant research and in applications such as marker-based breeding, improving the efficiency of germplasm management, evaluating seed purity, and as the basis of intellectual property protection (Hahn and Grifo 1996, Powell et al 1996, Mitchell et al 1997). Yang et al (1994) used SSRs to demonstrate that levels of allelic diversity in a collection of landraces were higher than in a corresponding collection of rice cultivars. Olufowote et al (1997) demonstrated that a selected set of highly informative SSR markers could be used to differentiate varieties, and that these markers were especially useful in identifying allele frequencies in complex mixtures of pure lines that were characteristic of many traditional (landrace) varieties. SSR-based fingerprinting was used to classify breeding material in Korea (Ji et al 1998) and red rice accessions in the United States (Vaughan et al 2001). These markers also provide an efficient methodology for checking varietal purity in commercial settings, where price differentials for certain premium rice varieties motivate unacceptable seed-mixing practices in the marketplace (Jain et al, personal communication).

Of particular interest to geneticists and breeders are studies in which SSR markers have been used to make inferences about the genetics, pedigrees, evolution, and/or identity of various traits and/or germplasm accessions. Temnykh et al (2000) suggested that certain regions of the rice genome are associated with higher than average levels of SSR diversity and speculated about the reasons for their observations. Studies by Bligh et al (1995, 1998) demonstrated that expansion/contraction of an SSR motif occurring at the splice site in the 3′ UTR of the *waxy* gene was correlated with amylose content of the grain, and Ayres et al (1997) used SSRs to make predictions about the inheritance of *waxy* alleles in complex pedigrees. Further investigation by Larkin and Park (1999) demonstrated that a single nucleotide polymorphism at this site conferred temperature sensitivity and altered transcript accumulation in developing rice endosperm.

Microsatellites have proved especially useful in evaluating diversity in narrowly defined gene pools in which other kinds of molecular markers such as amplified fragment length polymorphism, restriction fragment length polymorphism, or randomly amplified polymorphic DNA are unable to detect polymorphism (Powell et al 1996). Examples in rice include *O. glaberrima* accessions from West Africa (Lorieux et al 2000, Semon et al 2001, Talag et al 2000), basmati rice from northern India (S. Jain,

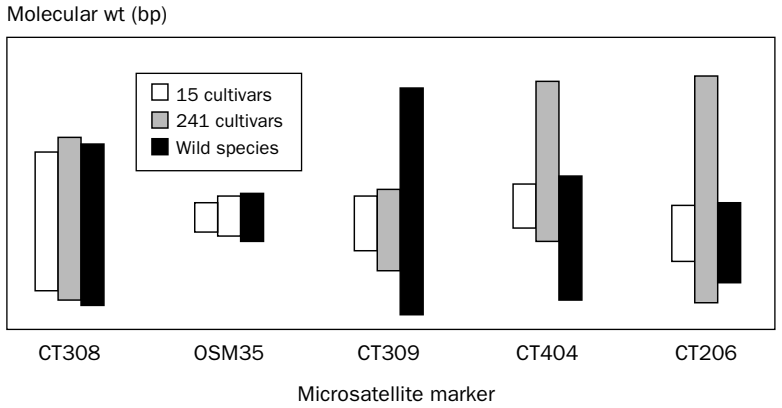


Fig. 5. Range of allele sizes at five SSR markers as detected in three groups of germplasm: (1) a group of 15 diverse *O. sativa* cultivars, (2) a group of 241 diverse *O. sativa* cultivars, and (3) a set of 76 wild AA genome *Oryza* species (described by Harrington 2000). Allele size ranges are illustrated in terms of molecular weight of alleles as detected on silver-stained gels for each set of germplasm.

personal communication), and japonica rice from California (Ni et al 2001) and Korea (Ji et al 1998, Kwon et al 2000).

When evolutionarily divergent accessions of *Oryza* and other grasses are being compared, chloroplast SSRs have proven to be a reliable tool for comparative phylogenetic analysis (Provan et al 1996, 1997, Ishii and McCouch 2000, Ishii et al 2001). SSR loci vary considerably in the amount of allelic diversity detected per locus. Harrington (2000) demonstrated that the relative number of alleles per locus detected in wild and cultivated AA genome species of *Oryza* varies greatly with the specific SSR marker (Fig. 5). Loci that tend to be hypervariable in a specific set of germplasm are most useful for differentiating among closely related genotypes, whereas SSR loci that are more conserved and have fewer alleles per locus can be used more reliably for evaluating genetic relationships among more distantly related accessions. Ultimately, sequence analysis is required to determine whether SSR amplicons that are determined to be identical in size based on electrophoretic migration patterns are homoplasic or are truly common by descent (X. Chen, personal communication). Thus, only when sequence analysis has confirmed the identity of alleles can phylogenetic reconstruction be reliably undertaken using SSR data (Doyle et al 1998).

In both public and private settings, SSRs are economically employed in hybrid rice breeding programs. These markers have also been used to help define heterotic groups in rice (Xiao et al 1996); to evaluate parental and hybrid seed purity (Q. Zhang and J. Mann, personal communication); to study the genetics of heterosis (Hua et al 2000), transgressive variation (Xiao et al 1998, J. Li, personal communication); and hybrid fertility (Zhang et al 1997); and to transfer traits via marker-assisted selection (He et al 2000).

SSRs have been used to define introgressions in wide hybridization programs (Brar et al 2000, Talag et al 2000), to construct ordered sets of substitution lines (Lorieux et al 2000, Z. Li, personal communication), and to identify genes and QTLs in both intra- and interspecific mapping populations (Blair et al 1997, McCouch et al 1997, Xiao et al 1998, Moncada et al 2000, Bao et al 2000, Yu et al 2000, Zou et al 2000, M. Thomson, personal communication).

Blair and McCouch (1997), M. Thomson (personal communication), and M. Lorieux (personal communication) have reported the use of SSRs to construct fine maps and select near-isogenic lines (NILs) for positional cloning of genes underlying QTLs (Fig. 6). Many groups are using SSRs to implement marker-assisted selection in breeding programs. It is predicted that the availability of an increasing number of SSR markers, well distributed in the rice genome, will provide an increasingly useful resource for many applications in genetics and breeding.

Automated SSR analysis

The use of fluorescently labeled microsatellite markers for genotyping on automated sequencers offers many advantages over analysis using traditional autoradiographic or silver-stained detection techniques. As a result, semiautomated methods of SSR genotyping are gradually replacing manual systems in plant breeding and genetics research (Mitchell et al 1997). The first report of automated SSR analysis in rice was Blair et al (2001), in which multiplex panels consisting of 27 markers were used for diversity analysis. More recently, a set of 160 well-distributed microsatellite markers was assembled into 21 multiplex panels, providing tools for facilitating semiautomated genotyping of rice (Coburn et al 2001). The panels comprised an average of eight markers each and provided genome-wide coverage of the 12 chromosomes of rice, with one marker approximately every 11 cM throughout the genome. As discussed by several authors and indicated in Figure 5, these markers amplify reliably in all AA-genome rice and many also amplify in more distantly related *Oryza* species (Wu and Tanksley 1993, Panaud et al 1996, Harrington 2000, Ishii et al 2001).

Advantages of the automated system include (1) a large increase in throughput made possible by the multiplexing of many polymerase chain reaction (PCR) products into a single lane, (2) a significant increase in accuracy of allele sizing achieved by the use of an internal size standard in each lane coupled with the availability of computerized allele-calling algorithms, and (3) a reduction in the required volume (and therefore the cost) of PCR reactions because of the high sensitivity of fluorescent detection, which also facilitates detection of loci that are difficult to amplify.

Future SSR development

In an effort to facilitate studies that aim to understand the relationship between phenotype and genotype, a new consortium has recently been formed that aims to enhance the development and evaluation of new SSR markers using publicly available sequence information. The importance of having a large set of SSR markers is to

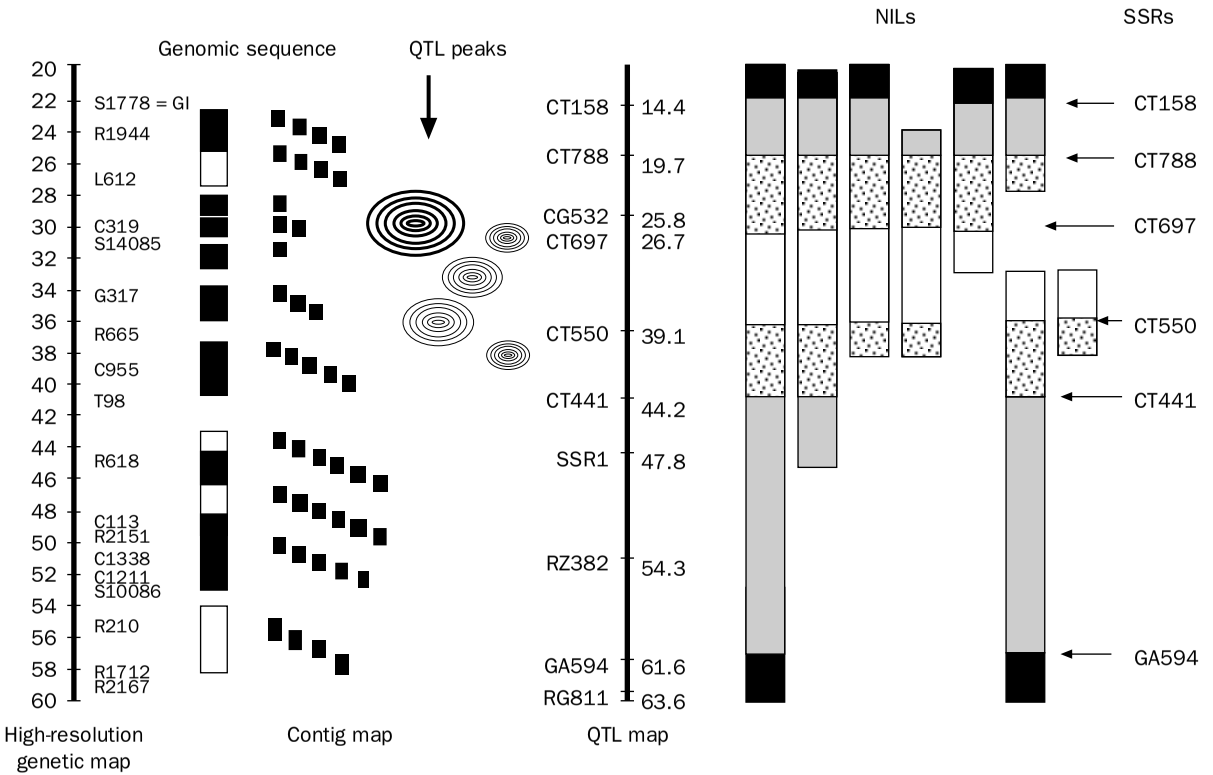


Fig. 6. Use of SSRs to link genotype and phenotype via the high-resolution genetic map, genomic sequence, contiged BAC and PAC clones, a QTL map targeting regions of peak LOD score, and the development of near-isogenic lines (NILs) for use in phenotypic evaluation.

provide a bridge for moving rapidly and efficiently from phenotype to sequence information and vice versa, and providing a foundation for understanding the evolutionary consequences of sequence variation.

The International Rice Microsatellite Initiative (IRMI) is a consortium of public and private research groups that are working together to generate a high-resolution set of microsatellite markers for rice. With the recent release of more than 7,000 uncharacterized SSR sequences (http://www.riceresearch.org) and with new genomic sequences emerging all the time from both public and private rice sequencing efforts, opportunities for developing, characterizing, and using SSR markers in high-resolution studies in rice have opened up dramatically. IRMI aims to make the primer sequences, PCR conditions, polymorphism information, and map positions of these markers publicly available over the Internet (www.gramene.org). To facilitate the use of SSRs in automated analyses, all primer pairs are being tested using a standard PCR protocol and panel of rice genotypes. Results of IRMI's decentralized effort are coordinated in a central database in which each contributor's input is recognized.

It is of great interest not only to harness SSRs as genetic markers but also to understand the role they play in the biology and evolution of organisms and as contributors to the rapidly evolving repetitive fraction of the rice genome. The power of SSR markers resides in their abundance, co-dominance, hypervariability, and ease and economy of use in both manual and automated systems. Groups interested in the rice genome can use them for evolutionary studies, to examine patterns of variation in specific regions of the genome, compare recombination frequencies across populations, fingerprint varieties, analyze pedigree relationships, construct high-resolution genetic and physical maps, discover and isolate genes and QTLs, screen mutant populations, and in molecular breeding programs. By establishing a core set of publicly available semiautomated SSR markers that can be readily used by independent groups worldwide, a large amount of data bridging sequence, genotype, and phenotype can be accumulated and readily integrated into a single database that integrates information across studies. Within the next few years, all SSRs in rice will likely be identified and their genomic locations known. This will provide the rice community with a rich source of material for addressing a wide array of scientific questions of both basic and applied interest.

References

Akagi H, Yokozeki Y, Inagaki A, Fujimura T. 1996. Microsatellite DNA markers for rice chromosomes. Theor. Appl. Genet. 94:61-67.

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD. 1997. Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor. Appl. Genet. 94:773-781.

Bao, JS, Zheng XW, Xia YW, He P, Shu QY, Lu X, Chen Y, Zhu LH. 2000. QTL mapping for the paste viscosity characteristics in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100:280-284.

Blair MW, McCouch SR. 1997. Microsatellite and sequenced-tagged site markers diagnostic for the rice bacterial leaf blight resistance gene *xa-5*. Theor. Appl. Genet. 95:174-184.

Blair MW, Hedetale V, McCouch SR. 2001. Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). Theor. Appl. Genet. (In press.)

Bligh HFJ, Till RI, Jones CA. 1995. A microsatellite sequence closely linked to the *Waxy* gene of *Oryza sativa*. Euphytica 86:83-85.

Bligh HFJ, Roach PS, Jones CA, Larkin PD, Fu H, Park WD. 1998. Use of alternate splice sites in granule bound starch synthase mRNA from low amylose rice varieties. Plant Mol. Biol. 38:407-415.

Brar DS, Abbasi FM, McNally R, Talag J, Corral M, Molinawe J, Li Z, Khush GS. 2000. Molecular characterization of introgression and alien gene tranfer in rice. In: Abstracts. International Rice Genetics Symposium, International Rice Research Institute, Los Baños, Philippines, Oct. 2000. p 56.

Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor. Appl. Genet. 95:553-567.

Chin EC, Senior ML, Shu H, Smith JSC. 1996. Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866-873.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100:713-721.

Coburn JR, Temnykh SV, McCouch SR. 2001. Chromosome-specific panels of microsatellite markers for semi-automated genotyping of rice (*Oryza sativa* L.). Crop Sci. (Submitted.)

Daly MJ, Lincoln SE, Lander ES. 1991. "PRIMER," unpublished software, Whitehead Institute/MIT Center for Genome Research. Available at http://www-genome.wi.mit.edu/ftp/pub/software/primer.0.5, and via anonymous ftp to ftp-genome.wi.mit.edu, directory/pub/software/primer.0.5.

Doyle JJ, Morgante M, Tingey SV, Powell W. 1998. Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). Mol. Biol. Evol. 15:215-218.

Goldstein DB, Clark AG. 1995. Microsatellite variation in North American populations of *Drosophila melanogaster*. Nucleic Acids Res. 23:3882-3886.

Guiderdoni E, Galinato E, Luistro J, Vergara G. 1992. Anther culture of tropical *japonica* × *indica* hybrids of rice (*Oryza sativa* L.). Euphytica 62:219-224.

Hahn WJ, Grifo FT. 1996. Molecular markers in plant conservation genetics. In: Sobral BWS, editor. The impact of plant molecular genetics. New York, N.Y. (USA): Birkhaeuser Boston, Inc.

Harrington S. 2000. A survey of genetic diversity of eight AA genome species of *Oryza* using microsatellite markers. MS thesis. Cornell University, Ithaca, NY.

He Y, Zhou P, Tan Y, Xu C, Zhang Q. 2000. Improvement of cooking and eating quality of Zhenshan 97, an elite maintainer line of hybrid rice, by molecular marker-assisted selection. In: Abstracts. International Rice Genetics Symposium, International Rice Research Institute, Los Baños, Philippines, Oct. 2000. p 94.

Hua JP, Xing YZ, Xu CG, Sun ZL, Zhang Q. 2000. Analyzing genetic basis of heterosis using an immortalized F_2 population. In: Abstracts. International Rice Genetics Symposium, International Rice Research Institute, Los Baños, Philippines, Oct. 2000. p 13.

Huang N, McCouch SR, Mew MT, Parco A, Guiderdoni E. 1994. Development of a RFLP map from a double haploid population in rice. Rice Genet. Newsl. 11:134-137.

Ishii T, McCouch SR. 2000. Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theor. Appl. Genet. 100:1257-1266.

Ishii T, Xu Y, McCouch SR. 2001. Nuclear and chloroplast microsatellite variation in A-genome species of rice (*Oryza* spp.). Genome. (In press.)

Ji HS, Koh HJ, Park SU, McCouch SR. 1998. Varietal identification in Japonica rice using microsatellite DNA markers. Korean J. Breed. 30:350-360.

Kwon SJ, Ahn SN, Yang CI, Hong HC, Kim YK, Suh JP, Hwang HG, Moon HP, Choi HC. 2000. Genetic diversity of Korean *Japonica* rice cultivars. In: Abstracts. International Rice Genetics Symposium, International Rice Research Institute, Los Baños, Philippines, Oct. 2000. p 41.

Larkin PL, Park WD. 1999. Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule bound starch synthase is temperature-sensitive and controlled by a single nucleotide polymorphism. Plant Mol. Biol. 40:719-727.

Lorieux M, Ndjiondjop M-N, Ghesquiere A. 2000. A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite-based genetic map. Theor. Appl. Genet. 100:591-601.

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M. 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol. Biol. 35:89-99

Mitchell SE, Kresovich S, Jester CA, Hernandez CJ, Szewc-McFadden AK. 1997. Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Sci. 37:617-624.

Moncada P, Martinez CP, Borrero J, Chatel M, Gauch Jr. H, Guimaraes E, Tohme J, McCouch SR. 2000. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. 102:441-452.

Ni J, Colowit PM, Oster JJ, Mackill DJ. 2001. Evaluation of genetic diversity in rice subspecies using SSLP markers. Crop Sci. (Submitted.)

Olufowote JO, Xu Y, Chen X, Park WO, Beachell HM, Dilday RH, Goto M, McCouch SR. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40:370-378.

Panaud O, Chen X, McCouch SR. 1995. Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome 38:1170-1176.

Panaud O, Chen X, McCouch SR. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol. Gen. Genet. 252:597-607.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. 1996. A comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2:225-238.

Provan J, Corbett G, Waugh R, McNicol JW, Morgante M, Powell W. 1996. DNA fingerprints of rice (*Oryza sativa*) obtained from hypervariable chloroplast simple sequence repeats. Proc. R. Soc. Lond. B 263:1275-1281.

Provan J, Corbett G, McNicol JW, Powell W. 1997. Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. Genome 40:104-110.

Pupko T, Graur D. 1999. Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. J. Mol. Evol. 48:313-316.

Semon M, Coburn J, Cadalen T, Jones MP, McCouch SR. 2001. Microsatellite variation in *O. glaberrima* Steud. accessions from West Africa. (Abstracts: Poster P3b_33.) Proceedings Plant and Animal Genome IX Conference, San Diego, Calif., 13-17 Jan. 2001. (http://www.intl-pag.org/pag/9/abstracts/)

Talag JD, Brar DS, Li ZK, Jones M, Khush GS. 2000. Molecular characterization of introgression from *O. glaberrima* into *O. sativa* using microsatellite markers. In: Abstracts. International Rice Genetics Symposium, International Rice Research Institute, Los Baños, Philippines, Oct. 2000. p 68.

Tautz D, Ranz M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 12:4127-4138.

Temnykh S, Cartinhour S, Park W, Ayres N, Hauck N, Lipovich L, Cho YG, McCouch SR. 2000. Mapping and genome organization of microsatellites in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100:697-712.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch SR. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. (Submitted.)

Vaughan LK, Ottis V, Prazak AM, Conaway-Bormans CA, Sneller CH, Chandler JM, Park WD. 2001. Is all red rice found in commercial rice really *Oryza sativa*? Weed Sci. (In press.)

Weber JL. 1990. Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics 7:524-530.

Weber JL, May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44:388-396.

Wu K, Tanksley SD. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. 241:225-235.

Xiao J, Li J, Yuan L, McCouch SR, Tanksley SD. 1996. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor. Appl. Genet. 92:637-643.

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899-909.

Yang GP, Saghai-Maroof MA, Xu CG, Zhang Q, Biyashev RM. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Genet. 245:187-194.

Yu SB, Zhong DB, Sanchez A, Xu JL, Domingo J, Khush GS, Li ZK. 2000. An integrated molecular linkage map and genomic regions with clustered QTLs detected in the LT RI population. In: Abstracts. International Rice Genetics Symposium, International Rice Research Institute, Los Baños, Philippines, Oct. 2000. p 116.

Zhang Q, Liu KD, Yang GP, Saghai-Maroof MA, Xu CG, Zhou ZQ. 1997. Molecular marker diversity and hybrid sterility in *indica-japonica* rice crosses. Theor. Appl. Genet. 95:112-118.

Zou JH, Pan XB, Chen ZX, Xu JY, Lu JF, Zhai WX, Zhu LH. 2000. Mapping quantitative trait loci controlling sheath blight resistance into rice cultivars (*Oryza sativa* L.). Theor. Appl. Genet. 101:569-573.

Notes

Authors' addresses: S.R. McCouch, S. Temnykh, A. Lukashova, J. Coburn, S. Harrington, M. Thomson, E. Septiningsih, M. Semon, P. Moncada, Jiming Li, Department of Plant Breeding, Cornell University, Ithaca, NY 14853-1901, USA; G. DeClerck, S. Cartinhour, USDA Center for Bioinformatics and Comparative Genomics, Cornell University, Ithaca, NY 14853-1901, USA.

Acknowledgments: We gratefully acknowledge financial support from the Rockefeller Foundation (RF95001#315, RF98001#663, and RF99001#726), RiceTec, Inc., and the United States Department of Agriculture (NRI #93-37300-8703).

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Molecular mapping and marker-assisted selection for major-gene traits in rice

D.J. Mackill and Junjian Ni

Since the late 1980s, many loci controlling both qualitative and quantitative traits of rice have been mapped using DNA markers. This chapter focuses on the mapping of major genes in rice. The distinction between major and minor genes has become unclear because of advances in molecular-marker mapping. A major gene can be involved in determining both qualitative and quantitative traits. Here we define a major gene as a locus that results in discrete phenotypes in a segregating population or one that controls more than 50% of the phenotypic variation for a continuously distributed trait. The latter is common for major genes controlling traits that are highly influenced by environment. Resistances to diseases and insects have received the most attention for mapping because of their importance as breeding objectives. Molecular mapping of major genes is important in determining the allelism of genes conferring identical phenotypes, use as a selectable marker in a breeding program, and positional cloning of genes. Molecular markers will be most useful for selection when (1) the phenotype is difficult or expensive to measure directly, (2) genes of similar phenotype are being pyramided into a single line, or (3) markers are being used to select against the donor genome in a backcrossing program.

The use of genetic markers for identifying and manipulating plant genes is not a new concept (Tanksley 1983). Morphological and isozyme variants have been used on a limited scale to map genes of economic importance. However, the insufficient number and variability of such markers prevented their wide-scale adoption for breeding purposes. This situation changed dramatically with the discovery that differences in endonuclease restriction sites or insertions/deletions between adjacent sites could be used as markers to construct genetic maps (Botstein et al 1980). The first restriction fragment length polymorphism (RFLP) map of rice was constructed in the 1980s at Cornell University (McCouch et al 1988) and high-density maps were subsequently developed (Causse et al 1994, Kurata et al 1994). Rice geneticists have now widely applied these maps to map genes controlling qualitatively and quantitatively inherited

traits of economic importance (Mackill 1999). In this chapter, we discuss the application of molecular markers to the mapping and manipulation of major genes in rice.

Major- versus minor-gene traits

Plant traits can generally be classified as qualitative or quantitative. In the former, the phenotype of plants in segregating populations can be classified easily into discrete classes, whereas, in the latter, the data usually show continuous variation. Qualitative traits are almost always controlled by one or two major genes segregating in Mendelian fashion. Quantitative traits are usually controlled by several genes with smaller effects that result in a continuous variation in segregating populations. These traits are typically sensitive to environmental factors, and the genetic effects are diluted by this environmental variation and by the interaction between the two.

In practice, there is a range of situations between the two extremes (Table 1). For example, some major genes are responsible for semiquantitative traits that segregate in a more or less discrete fashion. A common example is the semidwarfing gene such as *sd1* and the photoperiod-sensitivity gene *Se1*. In some cases, major genes can be responsible for quantitative traits showing continuous segregation; an example is the submergence tolerance gene *Sub1*. In the mapping of quantitative trait loci (QTL), the percent of the phenotypic variation under the control of each locus is generally used to assess the effects of the specific locus on the trait. For the purposes of this discussion, we designate QTL responsible for more than 50% of the phenotypic variation as "major genes," whereas loci controlling 25–50% of the variation would be "major QTL." For any particular QTL, the percent of the phenotype explained value will vary depending on the environment under which the trait is measured and the genetic population used.

Major-gene traits mapped in rice

Since the first RFLP map of rice was constructed (McCouch et al 1988), hundreds of genes and QTL have been mapped. As QTL in rice has been discussed in several previous reviews (McCouch and Doerge 1995, Yano and Sasaki 1997) and in this

Table 1. Classification of major- and minor-gene traits.

Trait	Segregation	Percent of phenotypic variation explained	Examples	Classification
Qualitative	Discrete	100	Purple leaf, blast resistance	Major gene
Semiquantitative	Discrete	100	Semidwarfism, *sd1*	Major gene
Quantitative	Continuous	>50	Submergence tolerance gene, *Sub1*	Major gene
Quantitative	Continuous	25–50	Stem rot resistance	Major QTL
Quantitative	Continuous	<25	QTL for most agronomic and physiological traits	QTL

symposium (Li, this volume), here we will focus on major-gene traits. Table 2 lists the major genes tagged in rice and reported in the available literature to date. Approximately half of all tagged genes are resistances to biotic factors. This likely reflects not only the economic importance of these characters and the applicability of marker-assisted selection (see below) but also the fact that most resistances are conferred by major genes.

Upon reviewing the literature of gene-tagging research in rice, several generalizations can be made:

1. While genes have been mapped on all rice chromosomes, it is interesting to note the large number of genes mapped on chromosomes 11 and 12 relative to their size in cM.
2. For molecular markers, polymerase chain reaction (PCR)-based markers such as RAPD (random amplified polymorphic DNA), AFLP (amplified fragment length polymorphism), and microsatellite or SSR (simple sequence repeat) are increasingly being applied in gene mapping.
3. For the mapping populations, the permanent populations such as DH (doubled haploid) and RI (recombinant inbred) populations are now more often used owing to their inherent advantages, although F_2 populations and F_3 families are still widely applied to locate interesting loci. Furthermore, although intersubspecific crosses are most frequently used to generate mapping populations, intrasubspecific and interspecific crosses have received more attention recently. The intrasubspecific crosses can avoid sterility and linkage drag that result from wide hybridization and will be favored in terms of practical application in plant breeding. Interspecific crosses are especially useful to introgress some specific traits from a wild source into elite cultivars.

Application of markers in breeding and genetics

Three major applications of molecular markers to rice breeding and genetics are (1) identifying allelism of genes that confer identical phenotypes (for example, blast resistance genes identified in different countries with different isolates), (2) use in marker-assisted selection (MAS) or marker-assisted backcrossing (MAB) programs, and (3) use in positional cloning of genes. This chapter focuses on their use in selection programs.

Advantages of MAS or MAB

A general rule when deciding about whether to employ MAS is that, if the gene of interest can be scored easily without markers, it will usually not be economical to employ them. Markers will be more appropriate when (1) the trait is difficult or expensive to score, (2) several genes are being selected that confer similar phenotypes, or (3) background markers are being used to select against genes from the donor parent.

Traits difficult to measure. With MAS, selection can be made without phenotype data, selection will not be influenced by the environment, and selection is possible

Table 2. Tagging and mapping of some major genes in rice using molecular markers.

Trait	Gene	Chromosome	Linked markers	Reference
1. Disease resistance				
Blast	*Pi1*	11	RZ536	Yu et al (1996)
	Pi2	6	RG64	Yu et al (1991)
	Pi4	12	RG869	Yu et al (1991)
	Pi5	4	RZ788	Wang et al (1994)
	Pi7	11	RG16	Wang et al (1994)
	Pi10	5	RG13	Naqvi et al (1995)
	Pi11	8	RG181B/BP127	Zhu et al (1993)
	Pi12	12	RG869/RG81/RZ397	Zheng et al (1996)
	Pi18	11	RZ536	Ahn et al (1996)
	Pi20	12	XNpb88	Imbe et al (1997)
	Pi21	4	G271/G317	RGN14:98-99
	Pi44	11	CDO520	Chen et al (1999)
	Pib	2	RZ123	Miyamoto et al (1996)
	Pita-2, Pita	12	XNpb088	Rybka et al (1997)
	Pik^{m}	11	L190/R1506	Kaji and Ogawa (1996)
	Pb1	11	S723/CDO226/C189	Fujii et al (1999)
Bacterial blight				
	Xa1	4	XNpb235	Yoshimura et al (1992)
	Xa3	11	XNpb181	Yoshimura et al (1995)
	Xa4	11	XNpb181	Yoshimura et al (1995)
	xa5	5	RZ390/RG550/RG207	Yoshimura et al (1995)
	Xa10	11	RG303	Yoshimura et al (1995)
	xa13	8	RG136	Zhang et al (1996)
	Xa21	11	RG103	Ronald et al (1992)
	Xa22(t)	11	R543/RZ536	Lin et al (1996)
Yellow mottle virus				
	RYMV	12	RG341/RG869	Ghesquière et al (1997)
Tungro (RTSV)				
	RTSV	4	RZ262	Sebastian et al (1996)
Rice stripe				
	Stv-b(i)	11	XNpb220	Hayano Saito et al (1998)
2. Insect resistance				
Gall midge				
	Gm2	4	RG329	Mohan et al (1994)
	Gm4(t)	8	R1813	Mohan et al (1997b)
Brown planthopper				
	Bph1	12	XNpb248	Hirabayashi and Ogawa (1995)
	Bph10	12	RG457	Ishii et al (1994)
	Bph(t)	9	RZ404/UCH170	Mei et al (1996)
Green leafhopper				
	GLH	4	RZ262	Sebastian et al (1996)
	Grlp3	3	XNpb144	Fukuta et al (1998)
	Grlp11	11	G1465	Fukuta et al (1998)
	Grh1	5	R566	Tamura et al (1999)
Whitebacked planthopper				
	WBPH	6	R1954/L668	Yamasaki et al (1999)
	WBPH	11	RG103/RG167	Kadirvel et al (1999)

continued on next page

Table 2 continued.

Trait	Gene	Chromosome	Linked markers	Reference
3. Abiotic stresses				
Submergence tolerance				
	Sub1	9	RZ698	Xu and Mackill (1996)
Salt tolerance				
	Salt	7	RG4	Zhang et al (1995)
	OSA3	12	RG457/Y12817R	Zhang et al (1999)
4. Heterosis and wide compatibility				
	tgms1.2	8	RZ562	Wang et al (1995)
	tms2	7	R643A/R1440	Yamaguchi et al (1997)
	tms3	6	RZ144	Subudhi et al (1997)
	tgms	9	RM257	Reddy et al (2000)
	tgms-vn1 (tms4)	2	RM27	Dong et al (2000)
	pms1	7	RG477	Zhang et al (1994)
	pms2	3	RG191	Zhang et al (1994)
	pms3	12	C751/RZ261	Mei et al (1999)
	ms-h(t)	9	RG451/RZ404	Koh et al (1999)
	Rf-1	10	OSRRF	Akagi et al (1996)
	Rf?	10	C1361	Tan et al (1998)
	Rf-2	2	CDO686/RZ58	Yang et al (1997)
	Rf3	1	RG532	Zhang et al (1997)
	Rf5	1	RG374/RG394	Shen et al (1998)
	Rfu	10	C4003	Yao et al (1997)
	Rf?	10	RM258	Huang et al (1999)
Hybrid breakdown				
	Hwd1	10	C701/R2309	Fukuoka et al (1998)
	Hwd2	7	C796B/R1382/C492/C145	Fukuoka et al (1998)
Wide compatibility				
	S5	6	RG213	Yanagihara et al (1995)
5. Grain quality				
Grain aroma				
	Fgr	8	RG28	Ahn et al (1992)
Cooked-kernel elongation				
	KNE	8	RZ323	Ahn et al (1993)
Amylose				
	Wx	6	wx	Wang et al (1992)
6. Other traits				
Photoperiod sensitivity				
	Se1	6	RG64	Mackill et al (1993)
Semidwarf gene				
	sd1	1	XNpb363	Ogi et al (1993)
	sdg	5	RZ182	Liang et al (1994)
Shattering-resistance gene				
	Sh2	1	XNpb174	Ogi et al (1993)
	Sh4	3	R250	Fukuta and Yagi (1998)
	Sht	4	R1427/C107	Sobrizal et al (1999)

even at the juvenile stage. Several traits, although being controlled by a single gene, are quite laborious to measure. Most traits that fall under this category are considered QTL that affect agronomic traits. Some types of disease resistance are quantitative or partial in nature and require replicated testing to measure accurately. Examples include resistance to blast (Wang et al 1994) and stem rot (Ni et al 2001). Although resistance conferred by one or two loci is not complete, it can be highly effective under field conditions in reducing damage from the disease. For some abiotic resistance traits such as salinity tolerance and submergence tolerance, selection in traditional breeding programs can only be done when the specific stress is present. In many cases, the trait must be measured in special screening nurseries grown from progeny seed. MAS has made it possible for breeders to perform indirect selection with the molecular markers closely linked with the genes. Although disease resistance is often easy to score in screening nurseries, scoring is not possible for recessive resistance genes during backcrossing unless an additional selfing generation is performed.

Traits that can only be measured after the reproductive stage would be good candidates for marker-assisted selection. For example, amylose content is currently measured after harvest using chemical methods or sophisticated equipment. A microsatellite repeat that is part of the *Wx* gene (Ayres et al 1997) can be effective in selecting for this trait. PCR-based markers proved 85% accurate for identifying the thermosensitive male sterility gene *tms3* in the juvenile stage (Lang et al 1999). The ultimate example of this concept is genes that control traits that can only be observed in the progeny of test crosses of individual plants. Two examples are the wide compatibility allele $S5_n$ (Liu et al 1997) and restorer genes for cytoplasmic male sterility (Akagi et al 1996, Ichikawa et al 1997, Yao et al 1997). The ability to identify such genes at the seedling stage during a backcrossing program would offer a remarkable savings in time and effort to transfer these genes into a specific genetic background.

Pyramiding multiple genes. Gene pyramiding is considered a viable approach to attaining durable resistance to rice diseases or insect pests. Different resistance genes often confer resistance to different isolates, races, or biotypes. Combining these resistances broadens the number of races or biotypes that a variety can resist, and there is evidence that multiple resistance genes make it more difficult for virulent races to evolve. Furthermore, combining major-gene and minor-gene resistance may lead to increased durability (Wang et al 1994). When partial resistance is present in breeding lines, scoring for major-gene resistance can also be more difficult (Kelly and Miklas 1998). In some cases, multiple pathogen races or insect biotypes can be used to detect plants with more than one resistance gene, but in practice this may be difficult or impossible. Molecular markers can be used to select for these multiple resistances simultaneously. When hybrid crops are the goal, additional options for pyramiding different resistance-gene combinations into different parents exist (Witcombe and Hash 2000).

Hittalmani et al (2000) used marker-assisted selection to combine three blast resistance genes, *Pi1* on chromosome 11, *Piz-5* on chromosome 6, and *Pita* on chromosome 12, in a single genotype. For *Piz-5*, a single marker was used, whereas flanking

markers were used for the other two. The authors confirmed that the markers were efficient in developing gene pyramids and that the line containing all three resistance genes had a broader resistance spectrum than lines with individual genes.

Markers have been used to pyramid several bacterial blight resistance genes. Huang et al (1997) pyramided four resistance genes, *Xa4, xa5, xa13*, and *Xa21*, using PCR-based markers. Sanchez et al (2000) transferred three bacterial blight resistance genes into three susceptible rice lines possessing desirable agronomic characteristics. Two RFLP markers, RG556 and RG207, linked to the recessive gene *xa5* were converted to sequence tagged site (STS) markers based on their DNA sequences. The RFLP marker RG136 was converted to an STS marker for selection for the resistance gene *xa13*. The sequence of the genomic clone RAPD248 was used to develop an STS marker for *Xa21*. This work showed the effectiveness of using markers linked to recessive genes in a backcrossing program, particularly in the presence of a dominant resistance gene. In an F_2 population, selection efficiency was as high as 95% for *xa5* and 96% for *xa13*.

Selecting against the donor genotype during backcrossing. To expand the number of genes available for rice breeding, exotic cultivars or even wild species with advantageous traits are being employed as parents more frequently in breeding programs. Unfortunately, these new genes are often associated with unfavorable alleles from the donor. High-density molecular maps and the graphical genotype (Young and Tanksley 1989) can be used to determine the genotype of different progenies throughout the entire genome and make it possible to select the individuals with the most favorable genetic constitution from the progenies of backcrosses.

There are two ways to use markers to assist in backcrossing a gene into a recurrent parent: (1) use markers to select for the recurrent parent markers on noncarrier chromosomes and (2) use markers to select against markers linked to the locus of interest to avoid linkage drag. When linkage drag is not a problem, selection against nonlinked markers can be performed easily. This procedure has been shown in simulation studies to reduce the number of backcrosses from six to three to transfer a gene of interest (Frisch et al 1999a). In most cases, even after six backcrosses, the size of the introgressed chromosomal segment can be quite large (Stam and Zeven 1981). Selecting for recombination on either side of the target gene requires a marker that cosegregates with the gene and two markers that flank the gene on either side. After the first backcross, BC_1F_1 plants recombinant for the gene and one of the flanking markers are selected. In the BC_2F_1 plants, recombinants for the other flanking marker are selected. The genetic distance between the flanking markers and the gene will determine the population size necessary to obtain the desirable recombinant. As the genetic distance decreases, the number of BCF_1 plants needed becomes prohibitively large (Frisch et al 1999b). For example, if the flanking markers are 5 cM from the target gene, 100 BC_nF_1 individuals would need to be assayed to obtain a recombinant between the target gene and one of the markers at the 0.99 level of probability. If one of the flanking markers is homozygous for the recurrent parent allele, the number rises to 192 for obtaining a recombinant with the other flanking marker (Frisch et al 1999b).

Chen et al (2000) transferred the bacterial blight resistance gene *Xa21* into a widely used parent for hybrid rice production in China, Minghui 63. They used one marker that is part of the *Xa21* gene to select for resistance. Markers flanking the gene at 0.8 and 3.0 cM on both sides were used to select for recombinants containing only a small fragment of the donor chromosome. Another 128 RFLP markers distributed throughout the genome were used to select for chromosomal fragments derived from Minghui 63. BC_1F_1 individuals containing *Xa21* and recombinant for one of the flanking markers were backcrossed to Minghui 63. BC_2F_1 individuals were selected for recombination with the other flanking marker. *Xa21*-containing plants in the BC_3F_1 generation were screened with markers covering the genome, and those homozygous for Minghui 63 alleles were selfed. The improved version of Minghui 63 was shown to be identical with the original except for its resistance to bacterial blight.

Practical considerations in marker-assisted selection

Molecular marker technology is starting to be integrated more and more into plant breeding programs. However, several factors should be taken into consideration when determining how suitable this approach will be.

Choice of molecular markers. Application of markers in selection schemes depends on identifying a closely linked marker or flanking markers near the gene or genes of interest. The suitability of any particular marker would depend on several factors:

1. *DNA quality and quantity required.* Markers such as RFLP, RAPD, and AFLP require high-quality DNA, necessitating more laborious extraction protocols. Also, compared with PCR-based methods, RFLPs need a large quantity of DNA.
2. *Difficulty of assay*. RFLP markers are the most laborious, requiring preparation of filters and Southern blots and development of film. AFLP markers do not require blotting, but the assay is relatively laborious, requires higher skills, and must be detected on polyacrylamide gels. Simple PCR-based markers such as RAPD, microsatellite, and CAPS (cleaved amplified polymorphic sequence) or STS are the simplest to detect, although microsatellite markers often require polyacrylamide gels.
3. *Degree of polymorphism*. Ideally, a marker should be useful within the germplasm pool to be used by breeders. Thus, although RFLP and AFLP markers are generally suitable for indica-japonica crosses, they are not highly polymorphic within a subspecies. Microsatellite markers are by far the most polymorphic and can be used within subspecies.
4. *Reliability*. RAPD markers have the reputation of being the most unreliable, although they are still used because of their simplicity. The other types of markers tend to be more reliable; however, none could be said to be error-free.

Many new types of markers are becoming available. As the rice genome sequence is completed, markers based on specific sequence differences, such as CAPS and SNPs (single nucleotide polymorphisms), will become the markers of choice. Now, microsatellite markers are certainly the best choice for most purposes. These markers are highly polymorphic, reliable, and abundantly available (see McCouch et al, this

volume). Those SSRs that can be scored using agarose gels will be particularly useful in marker-assisted breeding (Gupta et al 1999).

False-positives in MAS. If the distance between the linked marker and the target gene is not small enough, a crossover will result in false-positives during MAS. Another reason for the false-positive screening is incorrect results of gene mapping. Fine mapping or high-resolution mapping of the gene and the discovery of the more closely linked markers using larger populations will reduce the occurrence of false-positives. Furthermore, phenotypic evaluation should be performed with more reliable methods, with multiple replications, and under different environments in order to locate genes more precisely. New efficient gene/QTL mapping strategies and quantitative genetic analysis methods should also be proposed and adopted.

Expense of MAS. The relatively high expense is another factor limiting the development and application of MAS. The expense includes not only the materials and supplies but also less definable costs such as quality of technical support, lab space, and radioisotope permits (Mohan et al 1997a). However, advances in technologies will result in a decrease in the costs of MAS. PCR-based markers such as microsatellites and AFLP are amenable to automation. In addition, DNA extraction methods have been improved. Not only have rapid DNA extraction methods for rice been developed (Williams and Ronald 1994, Zheng et al 1995, Lange et al 1998), but it is also possible to isolate DNA directly from the seeds before sowing (Chunwongse et al 1993). With these developments, DNA marker technology without electrophoresis should come into use in the future.

Restriction of number of genes in the screening program. The number of genes (loci) involved in the MAS program is another factor that should be considered. For example, with only four or five loci being selected, the population sizes and number of F_1 seeds needed for a MAS program will be considerable and any further addition will lead to an exponential increase (Mackill et al 1999). This indicates that only the most important traits or loci should be identified and selected in a MAS program. More importantly, marker-assisted selection should be considered as a complement to conventional breeding rather than a replacement for it.

Conclusions

Molecular marker technology has changed the way plant genetics research is conducted. Genetics studies in rice should include placing newly identified genes on the molecular map. The technology is already leading to the cloning of important genes. However, the application of these tools in conventional breeding programs has been limited. The challenge for the future is to integrate these tools into ongoing rice improvement programs to accelerate the development of cultivars possessing unique combinations of genes in a productive and adapted genetic background.

Aside from its direct application to rice breeding programs, molecular mapping of major genes will ultimately allow the identification of the DNA sequence and resulting protein product responsible for the observed phenotype. Map-based cloning of genes has been a laborious undertaking, but, with the availability of the complete

genome sequence of rice, this will become easier. Although functional genomic approaches to gene identification will result in isolation of thousands of new genes, final determination of function for agronomically important genes will usually require association of the DNA sequence with the map position of the locus. Identification of gene sequences will also result in more accurate markers that can be used for marker-assisted selection.

References

Ahn SN, Bollich CN, McClung AM, Tanksley SD. 1993. RFLP analysis of genomic regions associated with cooked-kernel elongation in rice. Theor. Appl. Genet. 87:27-32.

Ahn SN, Bollich CN, Tanksley SD. 1992. RFLP tagging of a gene for aroma in rice. Theor. Appl. Genet. 84:825-828.

Ahn SN, Kim YK, Han SS, Choi HC, Moon HP, McCouch SR. 1996. Molecular mapping of a gene for resistance to a Korean isolate of rice blast. Rice Genet. Newsl. 13:74-76.

Akagi H, Yokozeki Y, Inagaki A, Nakamura A, Fujimura T. 1996. A codominant DNA marker closely linked to the rice nuclear restorer gene, *Rf-1*, identified with inter-SSR fingerprinting. Genome 39:1205-1209.

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD. 1997. Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. Theor. Appl. Genet. 94:773-781.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314-331.

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu KS, Xiao JH, Yu ZH, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251-1274.

Chen DH, dela Vina M, Inukai T, Mackill DJ, Ronald PC, Nelson RJ. 1999. Molecular mapping of the blast resistance gene, *Pi44*(t), in a line derived from a durably resistant rice cultivar. Theor. Appl. Genet. 98:1046-1053.

Chen S, Lin XH, Xu CG, Zhang QF. 2000. Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci. 40:239-244.

Chunwongse J, Martin GB, Tanksley SD. 1993. Pre-germination genotypic screening using PCR amplification of half-seeds. Theor. Appl. Genet. 86:694-698.

Dong NV, Subudhi PK, Luong PN, Quang VD, Quy TD, Zheng HG, Wang B, Nguyen HT. 2000. Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. Theor. Appl. Genet. 100:727-734.

Frisch M, Bohn M, Melchinger AE. 1999a. Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci. 39:1295-1301.

Frisch M, Bohn M, Melchinger AE. 1999b. Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci. 39:967-975.

Fujii K, Hayano-Saito Y, Sugiura N, Hayashi N, Saka N, Tooyama T, Izawa T, Shumiya A. 1999. Gene analysis of panicle blast resistance in rice cultivars with rice stripe resistance. Breed. Res. 1:203-210.

Fukuoka S, Namai H, Okuno K. 1998. RFLP mapping of the genes controlling hybrid breakdown in rice (*Oryza sativa* L.). Theor. Appl. Genet. 97:446-449.

Fukuta Y, Tamura K, Hirae M, Oya S. 1998. Genetic analysis of resistance to green rice leafhopper (*Nephotettix cincticeps* UHLER) in rice parental line, Norin-PL6, using RFLP markers. Breed. Sci. 48:243-249.

Fukuta Y, Yagi T. 1998. Mapping of a shattering resistance gene in a mutant line SR-5 induced from an indica rice variety, Nan-jing11. Breed. Sci. 48:345-348.

Ghesquière A, Albar L, Lorieux M, Ahmadi N, Fargette D, Huang N, McCouch SR, Notteghem JL. 1997. A major quantitative trait locus for rice yellow mottle virus resistance maps to a cluster of blast resistance genes on chromosome 12. Phytopathology 87:1243-1249.

Gupta PK, Varshney RK, Sharma PC, Ramesh B. 1999. Molecular markers and their applications in wheat breeding. Plant Breed. 118:369-390.

Harrington SE, Bligh HFJ, Park WD, Jones CA, McCouch SR. 1997. Linkage mapping of starch branching enzyme III in rice (*Oryza sativa* L.) and prediction of location of orthologous genes in other grasses. Theor. Appl. Genet. 94:564-568.

Hayano Saito Y, Tsuji T, Fujii K, Saito K, Iwasaki M, Saito A. 1998. Localization of the rice stripe disease resistance gene, *Stv-b*(i), by graphical genotyping and linkage analyses with molecular markers. Theor. Appl. Genet. 96:1044-1049.

Hirabayashi H, Ogawa T. 1995. RFLP mapping of *Bph-1* (brown planthopper resistance gene) in rice. Breed. Sci. 45:369-371.

Hittalmani S, Parco A, Mew TV, Zeigler RS, Huang N. 2000. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor. Appl. Genet. 100:1121-1128.

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95:313-320.

Huang Q, He Y, Jin R, Huang H, Zhu Y. 1999. Tagging of the restorer gene for rice HL-type CMS using microsatellite markers. Rice Genet. Newsl. 16:75-77.

Ichikawa N, Kishimoto N, Inagaki A, Nakamura A, Koshino Y, Yokozeki Y, Oka M, Samoto S, Akagi H, Higo K, Shinjyo C, Fujimura T, Shimada H. 1997. A rapid PCR-aided selection of a rice line containing the *Rf-1* gene which is involved in restoration of the cytoplasmic male sterility. Mol. Breed. 3:195-202.

Imbe T, Oba S, Yanoria MJT, Tsunematsu H. 1997. A new gene for blast resistance in rice cultivar IR24. Rice Genet. Newsl. 14:60-62.

Ishii T, Brar DS, Multani DS, Khush GS. 1994. Molecular tagging of genes for brown planthopper resistance and earliness introgressed from *Oryza australiensis* into cultivated rice, *O. sativa*. Genome 37:217-221.

Kadirvel P, Maheswaran M, Gunathilagaraj K. 1999. Molecular mapping of quantitative trait loci (QTLs) associated with whitebacked planthopper in rice. Int. Rice Res. Notes 24(3):12-14.

Kaji R, Ogawa T. 1996. RFLP mapping of blast resistance gene *Pik-m* in rice. Int. Rice Res. Notes 21(2-3):47.

Kelly JD, Miklas PN. 1998. The role of RAPD markers in breeding for disease resistance in common bean. Mol. Breed. 4:1-11.

Koh HJ, Son YH, Heu MH, Lee HS, McCouch SR. 1999. Molecular mapping of a new genic male-sterility gene causing chalky endosperm in rice (*Oryza sativa* L.). Euphytica 106:57-62.

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y. 1994. A 300-kilobase interval genetic map of rice including 883 expressed sequences. Nat. Genet. 8:365-372.

Lang NT, Subudhi PK, Virmani SS, Brar DS, Khush GS, Li ZK, Huang N. 1999. Development of PCR-based markers for thermosensitive genetic male sterility gene *tms3*(t) in rice (*Oryza sativa* L.). Hereditas 131:121-127.

Lange DA, Penuela S, Denny RL, Mudge J, Concibido VC, Orf JH, Young ND. 1998. A plant DNA isolation protocol suitable for polymerase chain reaction-based marker-assisted breeding. Crop Sci. 38:217-220.

Liang CZ, Gu MH, Pan XB, Liang GH, Zhu LH. 1994. RFLP tagging of a new semidwarfing gene in rice. Theor. Appl. Genet. 88:898-900.

Lin XH, Zhang DP, Xie YF, Gao HP, Zhang QF. 1996. Identifying and mapping a new gene for bacterial blight resistance in rice based on RFLP markers. Phytopathology 86:1156-1159.

Liu KD, Wang J, Li HB, Xu CG, Liu AM, Li XH, Zhang QF. 1997. A genome-wide analysis of wide compatibility in rice and the precise location of the S_5 locus in the molecular map. Theor. Appl. Genet. 95:809-814.

Mackill DJ. 1999. Genome analysis and breeding. In: Molecular biology of rice. New York (USA): Springer-Verlag. p 17-41.

Mackill DJ, Nguyen HT, Zhang JX. 1999. Use of molecular markers in plant improvement programs for rainfed lowland rice. Field Crops Res. 64:177-185.

Mackill DJ, Salam MA, Wang ZY, Tanksley SD. 1993. A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice. Theor. Appl. Genet. 85:536-540.

McCouch SR, Doerge RW. 1995. QTL mapping in rice. Trends Genet. 11:482-487.

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

Mei MH, Chen L, Zhang ZH, Li ZY, Xu CG, Zhang QF. 1999. *pms3* is the locus causing the original photoperiod-sensitive male sterility mutation of 'Nongken 58S'. Sci. China Ser. C-Life Sci. 42:316-322.

Mei M, Zhuang C, Wan R, Wu J, Hu W, Kochert G. 1996. Genetic analysis and tagging of genes for brown planthopper resistance in indica rice. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Manila (Philippines): International Rice Research Institute. p 590-595.

Miyamoto M, Ando I, Rybka K, Kodama O, Kawasaki S. 1996. High resolution mapping of the indica-derived rice blast resistance genes. 1. *Pi-b*. Mol. Plant-Microbe Interact. 9:6-13.

Mohan M, Nair S, Bentur JS, Rao UP, Bennett J. 1994. RFLP and RAPD mapping of the rice *Gm2* gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). Theor. Appl. Genet. 87:782-788.

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T. 1997a. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed. 3:87-103.

Mohan M, Sathyanarayanan PV, Kumar A, Srivastava MN, Nair S. 1997b. Molecular mapping of a resistance-specific PCR-based marker linked to a gall midge resistance gene (*Gm4t*) in rice. Theor. Appl. Genet. 95:777-782.

Naqvi NI, Bonman JM, Mackill DJ, Nelson RJ, Chattoo BB. 1995. Identification of RAPD markers linked to a major blast resistance gene in rice. Mol. Breed. 1:341-348.

Naqvi NI, Chattoo BB. 1996. Molecular genetic analysis and sequence characterized amplified region-assisted selection of blast resistance in rice. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Manila (Philippines): International Rice Research Institute. p 570-576.

Ni J, Colowit PM, Oster JJ, Xu K, Mackill DJ. 2001. Molecular markers linked to stem rot resistance in rice. Theor. Appl. Genet. 102:511-516.

Ogi Y, Kato H, Maruyama K, Saito A, Kikuchi F. 1993. Identification of RFLP markers closely linked to the semidwarfing gene at the *sd-1* locus in rice. Jpn. J. Breed. 43:141-146.

Reddy OUK, Siddiq EA, Sarma NP, Ali J, Hussain AJ, Nimmakayala P, Ramasamy P, Pammi S, Reddy AS. 2000. Genetic analysis of temperature-sensitive male sterilty in rice. Theor. Appl. Genet. 100:794-801.

Ronald PC, Albano B, Tabien R, Abenes L, Wu K-S, McCouch S, Tanksley SD. 1992. Genetic and physical analysis of the rice bacterial blight disease resistance locus, *Xa21*. Mol. Gen. Genet. 236:113-120.

Rybka K, Miyamoto M, Ando I, Saito A, Kawasaki S. 1997. High resolution mapping of the indica-derived rice blast resistance genes. 2. *Pi-ta*(2) and *Pi-ta* and a consideration of their origin. Mol. Plant-Microbe Interact. 10:517-524.

Sanchez AC, Brar DS, Huang N, Li Z, Khush GS. 2000. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci. 40:792-797.

Sebastian LS, Ikeda R, Huang N, Imbe T, Coffman WR, McCouch SR. 1996. Molecular mapping of resistance to rice tungro spherical virus and green leafhopper. Phytopathology 86:25-30.

Shen YW, Guan ZQ, Lu J, Zhuang JY, Zheng KL, Gao MW, Wang XM. 1998. Linkage analysis of a fertility restoring mutant generated from CMS rice. Theor. Appl. Genet. 97:261-266.

Sobrizal, Ikeda K, Sanchez PL, Yoshimura A. 1999. RFLP mapping of a seed shattering gene on chromosome 4 in rice. Rice Genet. Newsl. 16:74-75.

Stam P, Zeven AC. 1981. The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227-238.

Subudhi PK, Borkakati RP, Virmani SS, Huang N. 1997. Molecular mapping of a thermosensitive genetic male sterility gene in rice using bulked segregant analysis. Genome 40:188-194.

Tamura K, Fukuta Y, Hirae M, Oya S, Ashikawa I, Yagi T. 1999. Mapping of the *Grh1* locus for green rice leafhopper resistance in rice using RFLP markers. Breed. Sci. 49:11-14.

Tan XL, Vanavichit A, Amornsilpa S, Trangoonrung S. 1998. Genetic analysis of rice CMS-WA fertility restoration based on QTL mapping. Theor. Appl. Genet. 97:994-999.

Tanksley SD. 1983. Molecular markers in plant breeding. Plant Mol. Biol. Rep. 1:3-8.

Wang AY, Yu WP, Juang RH, Huang JW, Sung HY, Su JC. 1992. Presence of three rice sucrose synthase genes as revealed by cloning and sequencing of cDNA. Plant Mol. Biol. 18:1191-1194.

Wang B, Xu WW, Wang JZ, Wu W, Zheng HG, Yang ZY, Ray JD, Nguyen HT. 1995. Tagging and mapping the thermo-sensitive genic male-sterile gene in rice (*Oryza sativa* L.) with molecular markers. Theor. Appl. Genet. 91:1111-1114.

Wang GL, Mackill DJ, Bonman JM, McCouch SR, Champoux MC, Nelson RJ. 1994. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136:1421-1434.

Williams CE, Ronald PC. 1994. PCR template-DNA isolated quickly from monocot and dicot leaves without tissue homogenization. Nucleic Acids Res. 22:1917-1918.

Witcombe JR, Hash CT. 2000. Resistance gene deployment strategies in cereal hybrids using marker-assisted selection: gene pyramiding, three-way hybrids, and synthetic parent populations. Euphytica 112:175-186.

Wu KS, Martinez C, Lentini Z, Tohme J, Chumley FG, Scolnik PA, Valent B. 1996. Cloning a blast resistance gene by chromosome walking. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Los Baños (Philippines): International Rice Research Institute. p 669-674.

Xu K, Mackill DJ. 1996. A major locus for submergence tolerance mapped on rice chromosome 9. Mol. Breed. 2:219-224.

Yamaguchi Y, Ikeda R, Hirasawa H, Minami M, Ujihara A. 1997. Linkage analysis of thermosensitive genic male sterility gene, *tms-2*, in rice (*Oryza sativa* L.). Breed. Sci. 47:371-373.

Yamasaki M, Tsunematsu H, Yoshimura A, Iwata N, Yasui H. 1999. Quantitative trait locus mapping of ovicidal response in rice (*Oryza sativa* L.) against whitebacked planthopper (*Sogatella furcifera* Horvath). Crop Sci. 39:1178-1183.

Yanagihara S, McCouch SR, Ishikawa K, Ogi Y, Maruyama K, Ikehashi H. 1995. Molecular analysis of the inheritance of the *S-5* locus, conferring wide compatibility in indica/japonica hybrids of rice (*O. sativa* L.). Theor. Appl. Genet. 90:182-188.

Yang DC, Magpantay GB, Mendoza M, Huang N, Brar DS. 1997. Construction of a contig for a fertility restorer gene, *Rf2*, in rice using BAC library and its sequence with *Rf2* gene of maize. Rice Genet. Newsl. 14:116-117.

Yano M, Sasaki T. 1997. Genetic and molecular dissection of quantitative traits in rice. Plant Mol. Biol. 35:145-153.

Yao FY, Xu CG, Yu SB, Li JX, Gao YJ, Li XH, Zhang QF. 1997. Mapping and genetic analysis of two fertility restorer loci in the wild-abortive cytoplasmic male sterility system of rice (*Oryza sativa* L.). Euphytica 98:183-187.

Yoshimura S, Yoshimura A, Iwata N, McCouch SR, Abenes ML, Baraoidan MR, Mew TW, Nelson RJ. 1995. Tagging and combining bacterial blight resistance genes in rice using RAPD and RFLP markers. Mol. Breed. 1:375-387.

Yoshimura S, Yoshimura A, Saito A, Kishimoto N, Kawase M, Yano M, Nakagahra M, Ogawa T, Iwata N. 1992. RFLP analysis of introgressed chromosomal segments in three near-isogenic lines of rice for bacterial blight resistance genes, *Xa-1*, *Xa-3* and *Xa-4*. Jpn. J. Genet. 67:29-37.

Young ND, Tanksley SD. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theor. Appl. Genet. 77:95-101.

Yu ZH, Mackill DJ, Bonman JM, McCouch SR, Guiderdoni E, Notteghem JL, Tanksley SD. 1996. Molecular mapping of genes for resistance to rice blast (*Pyricularia grisea* Sacc.). Theor. Appl. Genet. 93:859-863.

Yu ZH, Mackill DJ, Bonman JM, Tanksley SD. 1991. Tagging genes for blast resistance in rice via linkage to RFLP markers. Theor. Appl. Genet. 81:471-476.

Zhang G, Angeles ER, Abenes MLP, Khush GS, Huang N. 1996. RAPD and RFLP mapping of the bacterial blight resistance gene *xa-13* in rice. Theor. Appl. Genet. 93:65-70.

Zhang G, Bharaj TS, Lu Y, Virmani SS, Huang N. 1997. Mapping of the *Rf-3* nuclear fertility-restoring gene for WA cytoplasmic male sterility in rice using RAPD and RFLP markers. Theor. Appl. Genet. 94:27-33.

Zhang GY, Guo Y, Chen SL, Chen SY. 1995. RFLP tagging of a salt tolerance gene in rice. Plant Sci. 110:227-234.

Zhang JS, Xie C, Li ZY, Chen SY. 1999. Expression of the plasma membrane H^+-ATPase gene in response to salt stress in a rice salt-tolerant mutant and its original variety. Theor. Appl. Genet. 99:1006-1011.

Zhang Q, Shen BZ, Dai XK, Mei MH, Saghai Maroof MA, Li ZB. 1994. Using bulked extremes and recessive class to map genes for photoperiod-sensitive genic male sterility in rice. Proc. Natl. Acad. Sci. USA 91:8675-8679.

Zheng K, Subudhi PK, Domingo J, Magpantay G, Huang N. 1995. Rapid DNA isolation for marker assisted selection in rice breeding. Rice Genet. Newsl. 12:255-258.

Zheng KL, Zhuang JY, Lu J, Qian HR, Lin HX. 1996. Identification of DNA markers tightly linked to blast resistance genes in rice. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Los Baños (Philippines): International Rice Research Institute. p 565-569.

Zhu LH, Chen Y, Xu YB, Xu JC, Cai HM, Ling ZZ. 1993. Construction of a molecular map of rice and gene mapping using a double haploid population of a cross between Indica and Japonica varieties. Rice Genet. Newsl. 10:132-134.

Notes

Authors' address: USDA-ARS, Department of Agronomy and Range Science, University of California, Davis, CA 95616, USA.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

QTL mapping in rice: a few critical considerations

Zhi-Kang Li

Advances in DNA markers and molecular linkage maps have stimulated a new area of molecular quantitative genetics through mapping of quantitative trait loci (QTLs). Several important questions regarding the types and number of QTLs, QTL × environment (QE) interactions, molecular dissection of trait correlation, and gene actions of QTLs are addressed based on results from many previous QTL mapping studies in rice and examples from two well-studied mapping populations and their related progenies. Two major types of QTLs, main-effect QTLs and epistatic QTLs, are recognized. Many QTLs are found to affect specific quantitative traits and they are widely distributed in the genome, though only a limited number of these loci are detectable in a mapping population largely because of epistasis and genotype × environment (GE) interactions. The effects (both main and epistatic) of individual QTLs affecting specific phenotypes may vary considerably. Most QTLs appear to be epistatic and complementary interaction appears to be the most common form of epistasis. Most QTLs tend to show varied degrees of QE interactions as a result of differential gene expression to biotic and abiotic stresses in different environments. QTLs differ greatly in their GE interactions and epistasis plays an important role in QE interactions. QTLs showing different gene actions appear to belong to different groups of genes and those exhibiting both additive and nonadditive gene actions are few. A new strategy is proposed for simultaneous QTL identification and transfer, and allele discovery through the development of introgression lines and use of DNA markers.

Most important traits dealt with by plant breeders are quantitative in nature. The classical multiple-factor hypothesis considered the continuous variation of quantitative traits as the collective effects of many genes, each with a small effect (Nilsson-Ehle 1909, Mather and Jinks 1982). Although numerous quantitative genetic studies have revealed the relative importance of genes with different actions on quantitative traits, the methodology itself does not allow resolution of continuous trait variation into individual underlying Mendelian factors. Advances in DNA markers since the late

1980s have had far-reaching effects on many areas of biological sciences. DNA markers are defined as linear landmarks in DNA molecules or chromosomes where genotypic differences arising from point mutations, insertions or deletions, transpositions, etc., can be detected and visualized by various molecular tools. Several major types of DNA markers are available, classified largely according to the molecular techniques by which the DNA differences are detected. These include RFLP (restriction fragment length polymorphism), RAPD (randomly amplified polymorphic DNA), CAPS (cleaved amplified polymorphic sequences), STS (sequence tagged sites), AFLP (amplified fragment length polymorphism), microsatellites or simple sequence repeats (SSRs), and single nucleotide polymorphisms (SNPs). The most common use of DNA markers is to develop comprehensive molecular genetic linkage maps, in which the linear orders and relative genetic distances of linked DNA markers on individual chromosomes and the whole genome in an organism are determined genetically and represented graphically. In rice, several high-density rice molecular linkage maps have been constructed (Causse et al 1994, Kurata et al 1994, Harushima et al 1998, Shomura et al 1997, Xiong et al 1997, Cho et al 1997). Establishment of these molecular linkage maps has greatly facilitated efforts in genome mapping of rice. One of the most important applications of DNA markers and molecular linkage maps is to dissect the genetic variation of quantitative traits into individual Mendelian factors through quantitative trait loci (QTL) mapping analyses.

QTL mapping can be defined as the marker-facilitated genetic dissection of variation of complex phenotypes through appropriate experimental design and statistical analyses of segregating materials. In QTL mapping, genes controlling genetic variation of quantitative traits in segregating populations are resolved into individual Mendelian factors by detecting marker-trait associations. The primary objective of a QTL mapping experiment is to understand the genetic basis of specific quantitative traits by determining the number, locations, gene effects, and actions of loci involved and their interactions with other loci (epistasis) and with environments (QTL $\times$ environment, or QE, interactions). Another major purpose of QTL mapping is to identify DNA markers diagnostic for particular phenotypes of interest so that marker-aided selection (MAS) can be used to efficiently manipulate progenies carrying alleles for target traits grown under nontarget environments.

In the past 10 years, many rice QTL mapping studies have been conducted and QTLs affecting a wide range of phenotypes have been identified. In this chapter, results from these QTL mapping studies are reviewed with regard to three critical questions: the types and number of QTLs involved in specific phenotypes, the behavior or gene action of QTLs, and QE interactions.

Types and number of QTLs

To date, QTLs identified in rice can be classified into two major types: main-effect QTLs (M-QTLs) and epistatic QTLs (E-QTLs), based largely on the presence or absence of epistasis. Distinction of the two types of QTLs is critical to our understanding of the genetic basis of quantitative trait variation in rice.

M-QTLs

M-QTLs are defined as single Mendelian factors at which effects (additive and/or dominance) on a given phenotype arise from allelic substitution and are detected by marker-trait associations using single-factor ANOVA or interval mapping models (Lander and Botstein 1989, Zeng 1994, Li 1997). M-QTLs in rice appear to include two groups of genes. The first group includes major genes of very large effects on highly heritable traits, which are typically detected with very large LOD scores (>10.0), and each explains a large portion of the total trait variation in a mapping population. Examples of this type are *sd-1* for semidwarf stature in rice, *Xa4* for bacterial blight resistance, *Ta9* for tiller angle, *Hd-1*, *Hd-3*, and *QHd3* for heading date, etc. (Table 1).

The second group includes the typical M-QTLs, which represent most (more than 90%) QTLs reported to date. These typical M-QTLs tend to have relatively small effects. There are two general results regarding the M-QTLs from previous studies. First, the number of detected M-QTLs for a specific trait in a population evaluated in a specific environment is relatively small. Based on the 324 cases (trait/population/environment combinations) in the previous rice QTL mapping studies involving 46 mapping populations and 71 phenotypes, the average number of detected M-QTLs per trait/population/environment is 3.7 ± 1.2 and, surprisingly, this number does not differ between high- and low-heritability traits. For example, the average detectable number of M-QTLs per population/environment is 3.7 and 4.1 for two highly heritable traits, days to heading and plant height, respectively, and 3.3 and 3.4 for the low-heritability traits, grains per panicle and grain yield, respectively. However, when a comparison is made across mapping populations and environments, many more M-QTLs are detected for each trait, and these are widely distributed on the 12 rice chromosomes (Table 2). This underestimation of M-QTL number in most QTL mapping studies is due largely to epistasis and QE interactions, which will be discussed later. Second, the accuracy of the estimated M-QTL genetic parameters, such as their effects, genomic locations, and gene actions, varies considerably, depending largely on errors in phenotyping and the statistical methods for parameter estimation. In the

Table 1. Some major genes affecting quantitative traits detected as main-effect QTLs in rice.

M-QTL	Trait	Chromosome	LOD	Reference
sd-1	Plant height	1	17.5	Huang et al (1996)
Xa4	Bacterial blight resistance	11	66.5	Li et al (1999b)
Sub1	Submergence tolerance	9	58.4	Sripongpangkul et al (2000), Xu and Mackill (1995), Nandi et al (1997)
Ta9	Tiller angle	9	32.3	Li et al (1999a)
Qlne1	Internode elongation	1	21.5	Sripongpangkul et al (2000)
QHd3	Heading date	3	24.5	Li et al (1995a)
Hd-1	Heading date	6	44.2	Yano et al (1997)
Hd-3	Heading date	6	64.4	Yano et al (1997)
QFll3b	Flag leaf length	3	11.1	Li et al (1998)
QSh2	Grain shattering	2	16.4	Zhong et al (1999), Fukuta et al (1996)

Table 2. The number of detected main-effect QTLs in rice.

Trait	Number of cases[a]	Mean	Range	Chromosome distribution
Heading date	29	3.7	1–8	All
Plant height	32	4.1	1–2	All
Lodging resistance traits	10	4.6	1–7	All but 5
Seedling vigor	10	3.6	2–8	1,2,3,5,6,7,9
Tiller and leaf angles	3	5.7	5–6	1,2,3,5,6,7,8,9
Leaf traits	14	2.3	0–5	All but 11
Resistances to abiotic stresses	25	4.8	0–11	All
Resistances to biotic stresses	25	5.1	3–9	All
Other morphological traits	5	5.4	4–7	All but 7,10
Anther culturability	9	2.8	1–5	All but 11
Grain traits	11	4.2	1–7	All but 6,8,9
Panicle traits	36	3.1	0–7	All
Grain quality and nutrients	19	2.0	1–4	All but 10,11
Panicles per plant	16	2.0	1–4	All but 6,8
Spikelet sterility	13	3.0	0–8	All
Grains per panicle	13	3.6	1–8	All
1,000-grain weight	18	4.4	1–10	All
Grain yield per plant	15	3.4	1–10	All but 10
Average	–	3.8		

[a]Each case represents a trait/population/environment combination in a total of 303 cases.

former case, the nature and size of mapping populations and use of replications in phenotyping play a key role in reducing phenotyping errors. In the latter case, control of background genetic variation and inclusion of epistasis and QE interactions in the statistical models are vital to obtaining more reliable parameter estimates (Zeng 1994, Li 1997, Wang et al 1999, Li 1999).

E-QTLs

The second type of QTLs is epistatic QTLs, or E-QTLs. E-QTLs are defined as loci at which trait values are determined by interactions between alleles at two or more loci and are detected by associations between trait values and multilocus marker genotypes using epistatic models (Li 1997, Wang et al 1999, Kao et al 1999). In other words, trait values (phenotypes) are associated with specific alleles at single loci for M-QTLs, but with multilocus genotypes for E-QTLs, as shown in Figure 1.

Historically, epistasis has been recognized as an important genetic basis underlying complex phenotypes (Wright 1932, 1951, Allard 1988) and founder-effect models of speciation (Templeton 1980). Recent results from an experiment and several mapping studies (Tanksley and Hewitt 1988, Doebley et al 1995, Lark et al 1995, Li et al 1997a,b, Yu et al 1997) have provided strong evidence suggesting that epistasis is an important genetic component determining complex traits. Li et al (1997a) showed that most of the early QTL mapping studies preferentially identify M-QTLs that either have large effects and/or act independently, and the inability to detect epistasis in most QTL mapping studies is due largely to the lack of appropriate methodology

(genetic/statistical models and corresponding software). Recently, the mixed linear model approaches with inclusion of digenic epistasis and QE interactions and comprehensive control of background genetic variation developed by Wang et al (1999) and by Kao et al (1999) have allowed more accurate detection and quantification of epistasis affecting complex quantitative traits.

Table 3 shows the relative contributions of M-QTLs and E-QTLs to total phenotypic variation of some quantitative traits detected in five related rice mapping populations from the cross between Lemont (japonica) and Teqing (indica). These include a set of recombinant inbred lines (RILs), two BCF_1 (the RILs × the parents, Lemont and Teqing), and two testcross F_1 populations (the RILs × two testers, Zhong413 and IR64) (Li et al 2001a, Luo et al 2001). Both M-QTLs and E-QTLs affecting several agronomic traits were mapped. Obviously, E-QTLs account for a much greater portion of the total phenotypic variation than M-QTLs for all traits except grains per panicle. Similar results were obtained in the IR64/Azucena doubled-haploid (DH) population (Li et al 2001b) except for plant height, in which the segregation of a major gene, *sd-1* (for semidwarf plant stature), accounted for more than 30% of the total trait variation. The relative importance of E-QTLs versus M-QTLs is further supported by the magnitude of epistatic effects of E-QTLs versus the main effects of M-QTLs estimated in these mapping populations (Table 4). Again, for complex grain yield and its components, the mean epistatic effects of the identified E-QTLs are equivalent to the M-QTL main effects.

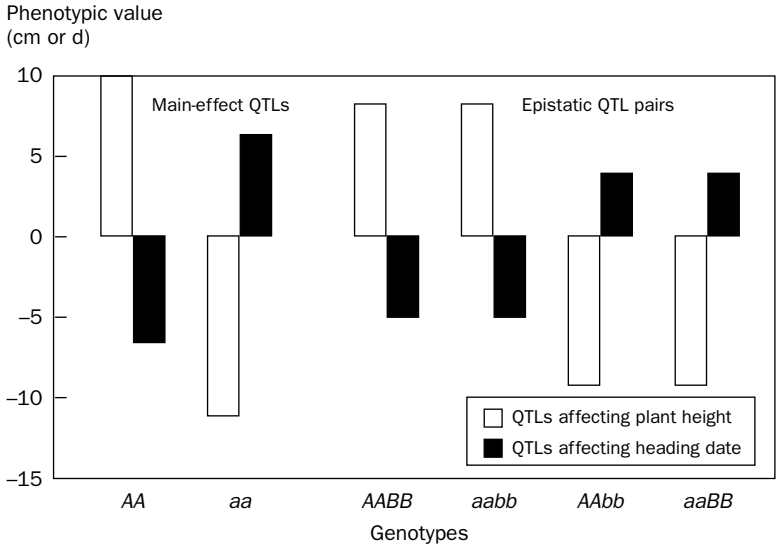


Fig. 1. Comparison between two main-effect QTLs, *QPh2* (plant height) and *QHd3a* (heading date), and two epistatic QTL pairs, between RG13 (chromosome 5) and RG103 (chromosome 11) (plant height), and between *Pgi1* (chromosome 3) and RZ66 (chromosome 8) (heading date). *A* is the Lemont (japonica) allele and *a* represents the Teqing (indica) allele.

Table 3. Relative importance of main-effect QTLs (M-QTLs) and epistatic QTLs (E-QTLs) in five related rice mapping populations from the Lemont/Teqing cross.

Trait	Lemont/Teqing RI[a] population				Four BC/testcross F_1 populations			
	M-QTLs		E-QTL pairs		M-QTLs		E-QTL pairs	
	N	R^2 (%)	N	R^2 (%)	N	R^2 (%)	N	R^2 (%)
Heading date	3.0	31.7	4.0	33.2	2.8	26.9	4.0	33.6
Plant height	4.5	39.1	4.0	30.9	2.3	20.9	4.3	37.7
Panicle length	1.0	5.2	4.5	51.0	1.5	17.0	2.8	27.7
Florets per panicle	2.0	20.0	5.5	49.0	0.8	10.2	5.5	48.3
Fertility	1.5	16.5	6.0	56.4	1.8	16.5	4.3	38.2
Panicles per plant	2.0	27.6	5.0	36.0	1.3	19.4	6.3	48.9
Grains per panicle	3.0	35.7	7.0	58.6	2.8	39.0	5.3	38.6
1,000-grain weight	2.0	17.7	6.0	43.6	1.3	15.5	7.3	43.2
Biomass	3.0	29.9	6.0	48.5	1.8	17.9	6.4	51.1
Grain yield	3.0	29.9	7.0	58.6	1.8	17.2	6.3	51.3
Mean	2.5	25.3	5.5	46.6	1.8	20.1	5.3	41.9

[a]RI = recombinant inbred, BC = backcross. N = number of M-QTLs detected or E-QTL pairs.

Table 4. Comparison of the magnitude of mean QTL effects between main-effect QTLs (M-QTLs) and epistatic QTLs (E-QTLs) detected in five related rice mapping populations from the Lemont/Teqing cross.

Trait	M-QTLs (mean ± SD)[a]	E-QTLs (mean ± SD)
Biomass (t ha^{-1})	0.75 ± 0.24	0.75 ± 0.25
Grain yield (t ha^{-1})	0.40 ± 0.13	0.45 ± 0.14
Panicles per plant	0.76 ± 0.13	0.84 ± 0.18
Grains per panicle	0.46 ± 0.15	0.58 ± 0.22
1,000-grain weight (g)	0.83 ± 0.22	0.95 ± 0.28

[a]SD = standard deviation.

Table 5 summarizes results on 423 E-QTL pairs affecting 13 quantitative traits in the five related mapping populations. First, three types of epistasis can be recognized. Type I is well described in classic quantitative genetics theory (Mather and Jinks 1982), in which two M-QTLs are involved in epistasis and affect the same phenotype. Type II involves interactions between alleles at an M-QTL and a background (or modifying) locus. Type III represents epistasis between two complementary loci that do not have detectable main effects (Li 1997). Of the total of 423 E-QTL pairs detected in the five mapping populations, types I, II, and III account for 3.2%, 28.5%, and 68.3%, respectively. Second, the frequencies of epistasis types I and II are greater in the RI population than in the backcross and testcross populations. This is expected since M-QTLs and E-QTLs are interchangeable depending on genetic background. An E-QTL can be detected as an M-QTL when alleles at the other locus with which it interacts become fixed in the backcrosses or testcrosses. Third, the predominance of type III epistasis or the interactions between complementary loci are important, indi-

Table 5. The percentage of three types of epistatic QTL pairs affecting 13 traits in the five related rice mapping populations from the Lemont/Teqing cross.

Trait	Recombinant inbred population				Four backcross/testcross F_1 populations			
	n	Type I	Type II	Type III	n	Type I	Type II	Type III
Heading date	8	12.5	50.0	37.5	21	4.8	28.6	66.7
Plant height	8	0.0	12.5	87.5	20	10.0	45.0	45.0
Flag leaf length	8	0.0	25.0	75.0	14	7.1	35.7	57.1
Flag leaf width	4	0.0	25.0	75.0	18	5.6	27.8	66.7
Panicle length	9	0.0	11.1	88.9	15	0.0	33.3	66.7
Floret density	12	0.0	16.7	83.3	25	0.0	28.0	72.0
Spikelets per panicle	12	0.0	41.7	58.3	27	0.0	44.4	55.6
Spikelet sterility	12	0.0	16.7	83.3	19	5.3	21.1	73.7
Panicles per plant	12	8.3	8.3	83.3	25	8.0	20.0	72.0
Grains per panicle	12	0.0	25.0	75.0	21	0.0	42.9	57.1
1,000-grain weight	12	0.0	16.7	83.3	29	0.0	24.1	75.9
Biomass	12	0.0	16.7	83.3	28	7.1	50.0	42.9
Grain yield	14	0.0	35.7	64.3	26	15.4	38.5	46.2
Mean	10.4	1.6	23.2	75.2	22.2	4.9	33.8	61.3
SD[a]	2.8	4.0	12.5	14.4	4.9	4.8	9.8	11.5

[a]SD = standard deviation.

cating that the trait values are more properties of the digenic genotypes than different alleles at the two loci. In cases of type III epistasis, both alleles at an E-QTL pair can be favorable depending on which allele is fixed at the other locus. This type of incompatible interaction between alleles at uncomplementary loci has been shown to be an important genetic basis underlying hybrid sterility and inbreeding depression in the progenies from the Lemont/Teqing cross (Li et al 1997a,b, 2001a, Luo et al 2001).

QE interactions

Genotype × environment (GE) interactions are a common property of most quantitative traits and the subject of extensive investigation. However, results from numerous classical GE studies provide little information regarding the QE interactions underlying GE interactions. To address this important issue, QTL mapping was performed on a DH population from the IR64/Azucena cross, evaluated in nine diverse environments. These environments cover a wide geographical range from 13.5° to 31.5° N and from 76° to 121.5° E at seven locations in four Asian countries (Philippines, China, India, and Thailand) and two different growing seasons at two of the locations (Li et al 2001b). Tables 6–8 summarize several important results on the QE interactions affecting plant height and heading date in the DH population.

First, GE interactions of quantitative traits are reflected in two aspects: inconsistent QTL detection across environments and the presence of significant QE effects. In the former case, some undetectable QTLs appear to result from non- or weak gene expression in certain environments, as suggested by variation in QTL main effects and their corresponding test statistics (Tables 6 and 7).

Table 6. Some nonenvironment-specific main-effect QTLs (M-QTLs) for plant height (cm) and heading date (days) and their interactions with environments detected in the IR64/Azucena doubled-haploid population.

M-QTLs	Chromosome	Marker interval	Parameter[a]	E1[b]	E2	E3	E4	E5	E6	E7	E8	E9	Mean ± SD
Sd-1	1	RG810–RZ801	*A*	–9.3	–12.9	–8.2	–16.0	–14.7	–14.4	–9.9	–13.4	–18.3	–13.0 ± 3.3
			AE	2.5	–	6.1	–1.8	–1.2	–	–	–	–2.8	0.6 ± 3.7
QPh3c	3	CD087–RG418a	*A*	–7.0	–6.6	–6.7	–7.7	–7.9		–6.3	–5.6	–3.7	–6.4 ± 1.3
			AE	–	–	–	–	–	–	–	–	–	
QPh4a[c]	4	**RG908–RG190**	*A*	3.1	4.5	4.6	4.4	4.7	5.3	7.1	–	7.5	5.2 ± 1.5
			AE	–	–2.1	–2.2	–	2.8	–	2.9	–2.4		–0.2 ± 2.8
QPh9a	9	**RZ206–RZ422**	*A*	**–3.3**	**–2.5**	**–3.9**	–3.3	–5.9		**–1.8**	**–3.1**	**–5.0**	–3.6 ± 1.3
			AE	–	–	–	–	–1.6	–1.9	1.3	–	–	–0.7 ± 1.8
QPh3b	3	**RZ394–RZ284**	*A*	–	**3.3**	**–2.2**	**3.3**	4.4	**5.4**	6.4	**4.4**	**5.4**	3.8 ± 2.6
			AE	–2.7	–	–2.2	–	1.6	–	2.8	–	2.7	0.4 ± 2.7
QPh5b	5	**CD0105–RZ649**	*A*	**5.2**	**2.9**	**4.5**	3.1	2.3	**6.6**	**3.2**	**8.5**	**5.2**	4.6 ± 2.0
			AE			**–2.8**					**3.9**		0.6 ± 4.7
QHd2	2	RZ123–RZ213	*A*	–2.6	–1.6	–2.0	–0.8	–1.3		–2.2	–1.1	0.9	–1.34 ± 1.08
			AE	–1.7	–	0.9	–	–	–	–	–	–	–0.40 ± 1.84
QHd3a	3	RG104–RG348	*A*	–4.0	–3.1	–3.3	–1.9	–1.5	–3.2	–2.1		–1.1	–2.53 ± 1.02
			AE	–	–3.1	–1.6	–	0.8	–	–	1.6	1.3	–0.20 ± 2.05
QHd7	7	**RZ488–RG477**	*A*	3.6	1.5	–	2.4	4.1	3.1	2.2	1.8	3.4	2.76 ± 0.92
			AE	–	**–0.9**	–4.3	**–0.9**	1.3	–	–	–	0.7	–0.82 ± 2.18
QHd8	8	RG978–RG1	*A*	2.0	1.4	–2.0	1.1	1.9	1.8	**1.2**	2.1	–1.5	0.89 ± 1.54
			AE	–	–	–	–	1.5	–0.8	–	–	1.2	0.63 ± 1.25

[a]*A* and *AE* are additive QTL and additive × environment interaction effects. [b]E1 = IRRI, wet season 1994 at Los Baños, Philippines; E2 = IRRI, dry season 1994-95 at Los Baños, Philippines; E3 = China National Rice Research Institute, Hangzhou, China, 1995; E4 = University of Agricultural Sciences, GKVK, Bangalore, India, 1995; E5 = South China Agricultural University, China, early season, 1995; E6 = South China Agricultural University, China, late season, 1995; E7 = Indian Agricultural Research Institute, New Delhi, 1995; E8 = Punjab Agricultural University, Ludhiana, India, 1995; E9 = Rice Research Institute, Bangkok, Thailand, 1995. [c]QTLs in bold letters are involved in epistasis and the bold parameters are obtained in epistatic models, – represents statistically insignificant QTL parameters. SD = standard deviation.

Table 7. Some environment-specific main-effect QTLs (M-QTLs) for plant height (cm) and heading date (days) and their interactions with environments detected in the IR64/Azucena doubled-haploid population.

M-QTLs	Chromosome	Marker interval	Parameter[a]	E1[b]	E2	E3	E4	E5	E6	E7	E8	E9	Mean ± SD[c]
QPh3a	3	RG348–RZ329	*A*	–2.7	–3.2								–3.0 ± 0.4
			AE		–2.0								–2.0
QPh4c	4	RG449–RG788	*A*				–4.4						–4.4
			AE			3.2			1.1	–2.0		**1.1**	0.9 ± 2.1
QPh5c	5	RZ70–RZ225	*A*									–3.6	–3.6
			AE				–1.4						–1.4
QPh7a	7	RG769–RG511	*A*			2.8							2.8
			AE		–2.7		2.3						–0.2 ± 3.5
QPh7b	7	RZ488–RG477	*A*	2.3						4.8			3.6 ± 1.8
			AE	–1.6	–3.0		1.9			1.8			–0.2 ± 2.5
QPh8	8	TGMS1.2–AG8	*A*	2.8	3.8		2.7			**2.2**			2.9 ± 0.7
			AE		2.6		0.8	–2.0		–1.1	**–2.2**		–0.4 ± 2.0
QPh11	11	RG247–RG167	*A*			–3.1						2.5	–0.3 ± 4.0
			AE			–2.6						2.6	0.0 ± 3.7
QHd1c	1	RZ801–RG331	*A*		–1.1	**–2.1**	–2.2	–2.3	–3.3	**–0.7**		–1.3	–1.86 ± 0.88
			AE					–1.2	–1.5		1.5		–0.40 ± 1.65
QHd3b	3	RZ678–RZ574	*A*	2.5				2.2	1.9	**2.2**		0.9	1.94 ± 0.62
			AE		–1.6			0.8	–0.8				–0.53 ± 1.22
QHd3c	3	RZ448–*Pgi1*	*A*	**1.1**	1.8			3.0	**3.1**	**1.9**	**1.9**	**2.5**	2.19 ± 0.72
			AE		**–0.8**			1.9					0.55 ± 1.91
QHd4a	4	RG218–RZ262	*A*	**2.7**									2.70
			AE										
QHd4b	4	RG908–RG190	*A*				1.0	0.9		1.0		1.6	1.13 ± 0.32
			AE								–1.1		–1.10
QHd5a	5	RG556–RZ556	*A*	**2.3**	**1.7**	1.3						1.4	1.68 ± 0.45
			AE										
QHd9	9	RZ206–RZ422	*A*	–2.6	**–1.0**		**–1.5**	**–1.9**	–1.5			**–1.6**	–1.68 ± 0.53
			AE			**1.2**					1.2		1.20 ± 0.00
QHd11	11	G44–RG247	*A*				–2.0			–1.5			–1.75 ± 0.35
			AE										

[a]*A* and *AE* are additive QTL and additive × environment interaction effects. [b]E1 = IRRI, wet season 1994 at Los Baños, Philippines; E2 = IRRI, dry season 1994-95 at Los Baños, Philippines; E3 = China National Rice Research Institute, Hangzhou, China, 1995; E4 = University of Agricultural Sciences, GKVK, Bangalore, India, 1995; E5 = South China Agricultural University, China, early season, 1995; E6 = South China Agricultural University, China, late season, 1995; E7 = Indian Agricultural Research Institute, New Delhi, 1995; E8 = Punjab Agricultural University, Ludhiana, India, 1995; E9 = Rice Research Institute, Bangkok, Thailand, 1995. [c]SD = standard deviation.

On average, QTL main effects were detectable (under the minimum threshold of $P < 0.05$) in 53.8% (50.0% for plant height, PH, and 57.6% for heading date, HD) of the environments for the 37 identified M-QTLs and in 55.0% of the cases (67.5% for plant height and 42.4% for heading date) for the 33 identified E-QTL pairs. This would have dropped to 32.3% (41.7% for PH and 22.9% for HD) for the M-QTLs and 40.3% (57.6% for PH and 32.3% for HD) for the E-QTLs if a more stringent threshold of $P < 0.005$ (LOD = 2.4) had been used. In other words, the experiment-wide type II error would have reached 20.4% for the M-QTLs and 24.4% for the E-QTLs in this experiment under such a threshold. In the latter situation, altered gene expression appeared to be responsible for the significant QE effects. The QE effects were detectable in 25.7% (27.8% for PH and 23.5% for HD) of the QTL/environment combinations for the M-QTLs and in 22.8% of the cases (19.2% for PH and 26.5% for HD) for the E-QTLs. The average magnitudes of *AE* and *AAE* effects were 74.5% and 85.9% of the QTL main effects (*A* and *AA*) for HD, but only 42.2% and 39.4% for PH, respectively (Table 8).

Second, different traits may show different levels of QE interaction, consistent with the levels of their GE interactions. In this study, heading date showed a much greater degree of GE interaction than plant height. This is reflected by more environment-specific M-QTLs, greater magnitude of significant QE effects, and greater involvement of epistasis for heading date than for plant height. For example, the average magnitude of the QE effects was 80.8% of the mean QTL main effect for heading date but only 40.8% for plant height. Also, there were 21 detectable E-QTL pairs for heading date but only 11 for plant height (Table 8).

Third, different M-QTLs affecting the same phenotype may differ greatly in their interactions with the environments. Some M-QTLs and E-QTLs were detectable across all nine environments but others were detected in only one or two environments (see Tables 6 and 7). When detectable in multiple environments, the main effects of most M-QTLs tend to be in the same direction and show relatively small variation in magnitude across the environments. In seven cases (*QPh1*, *QPh3b*, *QPh11*, *QHd1a*, *QHd2*, *QHd8*, and *QHd12*), the two alleles at M-QTLs had opposite effects in different environments. This immediately suggests that the two alleles at these loci responded dif-

Table 8. Magnitudes of QTL main effects (*A* and *AA*) and QTL × environment effects (*AE* and *AAE*) associated with plant height (PH) and heading date (HD) detected in the IR64/Azucena doubled-haploid population.

Trait	M-QTLs[a]					E-QTLs				
	A			*AE*		*AA*			*AAE*	
	N	*n*	Mean ± SD	*n*	Mean ± SD	*N*	*n*	Mean ± SD	*n*	Mean ± SD
PH	20	90	5.29 ± 3.18	52	2.23 ± 0.90	11	70	4.82 ± 1.64	19	1.91 ± 0.69
HD	17	88	1.88 ± 0.80	36	1.40 ± 0.84	21	78	1.63 ± 0.74	50	1.40 ± 0.73

[a] *N* is the total number of detected M-QTLs or E-QTLs and *n* is the total number of significant QTL parameters (effects) of the M-QTLs or E-QTLs detected across the nine environments. SD = standard deviation.

ferently to the environments. Similarly, five E-QTL pairs for heading date showed opposite *AAE* effects in different environments (Li et al 2001b). Thus, the power (LOD score) by which a QTL is detected and the magnitude of its effect obtained in a single environment provide little information regarding its performance in another. All of these results strongly justify the advantage of marker-aided selection for QTLs of different behavior over phenotypic selection, even for highly heritable traits.

Finally, epistasis plays an important role in GE interactions. In addition to the above two pieces of evidence, 28.3% of the QTL main effects (27.8% for plant height and 28.9% for heading date) were detectable only in epistatic models (Tables 6 and 7). In other words, nearly 30% of the undetected M-QTLs in a single environment were due to epistasis.

Differential gene expression to biotic and abiotic stresses is largely responsible for GE interactions of quantitative traits. The greater level of QE interactions observed for QTLs affecting rice heading date is not surprising since the flowering time of rice plants is known to be affected by many environmental factors, such as daylength, temperature, soil fertility, drought, etc. Strong evidence for the presence of epistatic interactions between and among different M-QTLs for rice heading date and their differential responses to daylength has been clearly demonstrated using near-isogenic lines (Lin et al 2000). Table 9 shows another more extreme case of QE interactions affecting plant height under different submergence conditions. Under nonstress and submergence stresses, different sets of M-QTLs affecting plant height (elongation) were detected and the expression of most M-QTLs was much stronger under the more stressful condition (under the submergence of muddy water of the field) than under the clean water submergence in the greenhouse. It is also interesting to note that the major gene *Sub1* for submergence tolerance was detected as a small QTL when the plants were submerged under the clear water condition (test 1). Similarly, it was clearly shown that the bacterial blight resistance gene *Xa4* acts as a major resistance gene against the avirulent races of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*), but as a resistance M-QTL against the virulent races of *Xoo* (Li et al 1999b, 2001c).

Molecular dissection of trait correlation

Determination of the genetic basis of trait correlation has been a major challenge in quantitative genetics and it has important implications for plant and animal improvement. It is well known that trait correlation may arise from linkage, pleiotropy, and epistasis. In addition to its genetic determinants, trait correlation may have physiological and environmental bases (Falconer 1983). In this respect, QTL mapping can help in gaining insights into this problem by identifying clustered QTLs mapped for correlated traits and comparing their genetic parameters (locations and effects).

Table 10 shows some examples of clustered M-QTLs affecting correlated traits in rice. Cases 1 and 2 represent examples of pleiotropy, which is often responsible for positive correlation between a complex phenotype (spikelets per panicle) and its component traits (number of primary and secondary branches per panicle). Cases 3 and 4 present a situation of QTLs affecting developmentally related traits. For *QLusi12*

Table 9. Main-effect QTLs affecting plant elongation and submergence tolerance in the IR74/Jalmagna recombinant inbred population.

QTL	Trait[a]	Chromosome	Marker	Stress under clean water LOD	*A*[b]	G^2	Stress under muddy water LOD	*A*	G^2	Tolerance allele
QPh1a	IPH	1	*sd-1*	10.1	–3.8		4.1	–3.2		
QPh1b	IPH	1	P3M7-6				6.6	–3.5		
QPh3	IPH	3	P2M1-12				4.6	3.4		
Qlne1	PHI	1	*sd-1*	10.5	–13.6		21.5	–23.2	23.2	Jalmagna
Qlne1	INI			7.0	–8.7		18.9	–18.2		
Qlne2	PHI	2	P2M9-8				7.4	–14.2		
	INI						5.1	–9.6		
Qlne4	PHI	4	P3M1-5				12.6	21.8		
	INI						10.7	21.8		
	LLI			6.2	–6.2				19.7	Jalmagna
QLe6	LLI	6	P2M5-17	2.5	4.9					
QLe7	LLI	7	P1M10-5				3.7	–3.5	5.2	Jalmagna
QSubt3	ST	3	P1M3-5						15.7	IR74
QSubt5	ST	5	P1M6-9			7.2			9.4	Jalmagna
QSubt8	ST	8	P3M7-3			6.9			11.9	Jalmagna
Sub1	ST	9	P2M1-15			5.1			58.3	IR74
QSubt9	ST	9	P2M5-18			6.6				Jalmagna
QSubt10	ST	10	P3M1-3						13.0	Jalmagna

[a]IPH is initial plant height (before stress), PHI, INI, LLI, and ST are plant height increment, internode increment, leaf length increment, and submergence tolerance after stress. G^2 is the likelihood ratio chi-square statistic associated with submergence tolerance. The significant values of G^2 at $P = 0.05$, 0.01, and 0.001 are 3.79, 6.63, and 11.60, respectively. [b]*A* = additive QTL effect.
Source: Sripongpangkul et al (2000).

(case 3), one may infer that this gene is expressed when the plant starts to elongate. Its expression apparently reaches its peak in the middle phase of panicle differentiation, as suggested by its strongest effect on the upper third internode. *QNnei3* (case 4) is even more interesting. The mapping data suggest that this gene expresses strongly at the early vegetative growth stage, indicated by its large effect on the number of nonelongated internodes, which contributes strongly to heading date but weakly to height. *QGl3* and *QGw5* (cases 5 and 6) represent the indirect effects on grain weight of two genes affecting grain length and width. In addition, the opposite effects of the two genes on grain density provide an appropriate explanation for the negative correlation between grain length and/or width and grain density. This is a good example for the clustered QTLs arising from correlated traits having physiological (causal) relationships (Xu et al 2001). It should be pointed out that QTLs affecting correlated traits tend to be mapped in clusters in most reported QTL mapping studies. On the other hand, it is common that QTLs for uncorrelated traits can also be mapped together in the same experiment. In these cases, caution should be taken in interpreting the results because of the typical low resolution (5–15 cM) of most mapped QTLs even with the use of more advanced analytical methods (Wang et al 1999, Kao et al 1999).

Table 10. Some main-effect QTLs (M-QTLs) affecting correlated traits identified in the Lemont/Teqing recombinant inbred population.

Case	M-QTL	Trait	Marker interval	LOD	Effect
1	*QPbn3b*	Primary branch number	RM227–RM85	12.77	–0.54
		Secondary branch number		19.83	–4.89
		Spikelets per panicle		20.26	–0.82
2	*QPbn4*	Primary branch number	RM303–RM317	8.30	0.48
		Secondary branch number		3.54	2.14
		Spikelets per panicle		2.94	0.34
3	*QLusi12*	Length of upper 2nd internode (cm)	RG20q–RG91q	6.70	–0.69
		Length of upper 3rd internode (cm)		12.34	–0.92
		Length of upper 4th internode (cm)		12.03	–0.68
		Plant height (cm)		5.81	–2.37
4	*QNnei3*	Nonelongated internode number	RG348a–C636x	13.47	–1.02
		Elongated internode number		11.23	–0.21
		Heading date (d)		15.23	–3.60
		Length of upper 4th internode (cm)		4.11	–0.42
		Plant height (cm)		4.98	–1.96
5	*QGl3*	Grain length (mm)	RD3.5–RD3.7	15.31	3.04
		Grain volume (mm^3)		3.97	11.80
		Grain density (mg mm^{-3})		5.43	–0.74
		1,000-grain weight (g)		8.34	0.80
6	*QGw5*	Grain width (mm)	Y1049–R569a	12.87	–1.07
		Grain thickness (mm)		7.20	–0.36
		Grain volume (mm^3)		7.56	–1.85
		Grain density (mg mm^{-3})		7.93	1.03
		1,000-grain weight (g)		3.81	–0.62

Gene action of QTLs

In QTL mapping studies, determination of gene action associated with QTLs has been one of the major objectives, which is required for the application of MAS of QTLs to breeding programs. Early QTL mapping studies using F_2 or backcross populations have generated some results on gene action of mapped QTLs in rice (Li et al 1995a,b, Lin et al 1996, Xiao et al 1995). However, the information of mapped QTLs showing nonadditive gene action has been particularly lacking largely because of the shifting trend of using permanent homozygous populations such as RI/DH populations and introgression lines for QTL mapping. Here, a striking result from the five related mapping populations was obtained, which appears to be unexpected from the classical quantitative genetics theory.

Table 11 shows that, for most traits, QTLs showing additive gene action and QTLs having nonadditive gene action appear to belong to different groups of genes and few loci show both additive and nonadditive (partial or complete dominance) gene action. For example, most QTLs affecting the trait performance of the backcross and testcross F_1 hybrids showed either additive gene action (32.5% for M-QTLs and 17.5% for E-QTLs) or overdominance action (57.9% for M-QTLs and 78.5% for E-QTLs), and few loci (10.5% for M-QTLs and 3.9% for E-QTLs) showed partial or complete dominance (Table 11). The only exception was the QTLs affecting heading date where QTLs of partial or complete dominance accounted for 40.0% and 14.3% of the total QTLs identified, respectively. These results, in conjunction with those in previous sections, have led us to conclude that epistasis and overdominance are the primary genetic basis of heterosis for most yield-related traits in rice (Li et al 2001a, Luo et al 2001).

Table 11. Number of main-effect QTLs (M-QTLs) and epistatic QTLs (E-QTLs) showing different gene action identified in the five related rice mapping populations of the Lemont/Teqing cross.

Trait	M-QTLs			E-QTLs		
	Additive[a]	with complete or partial dominance	Over-dominance	Additive	with complete or partial dominance	Over-dominance
Heading date	5 (2)	4	1	4 (8)	3	14
Plant height	3 (4)	2	4	2 (8)	1	14
Panicles plant^{-1}	2 (4)	0	3	5 (11)	0	20
Spikelets panicle^{-1}	3 (4)	0	4	2 (12)	0	24
Spikelet sterility	4 (2)	1	4	6 (12)	1	12
Panicle length	3 (1)	0	4	3 (9)	4	8
Grains panicle^{-1}	2 (4)	1	8	2 (12)	0	19
Grain weight	1 (3)	0	4	11 (12)	0	19
Biomass	1 (4)	0	6	4 (12)	0	24
Grain yield	1 (4)	0	6	1 (12)	0	25
Mean	2.5 (3.2)	0.8	4.4	4.0 (10.8)	0.9	17.9
Standard deviation	1.4 (1.1)	1.3	1.9	2.9 (1.8)	1.4	5.7

[a]The numbers in parentheses are results from the recombinant inbred population and the remaining are the mean data from two backcross and two testcross F_1 populations.

Simultaneous QTL introgression and identification

One of the most important objectives of QTL mapping is to apply MAS for genetic improvement of quantitative traits. Although marker-aided backcrossing has been successfully used to transfer specific target genes into an otherwise desirable genotype (Huang et al 1997, Sanchez et al 2000), few efforts on MAS for QTLs have been reported (Tanksley and Hewitt 1988, Shen et al 2001). Theoretically, the accuracy of the genetic parameters (effect and position) of a QTL is highly correlated with the magnitude of its effect (Wang et al 1999). Thus, it is expected that MAS for QTLs with very large main effects should be effective even though it is advisable to use markers that closely flank the target region. However, marker-aided introgression of a target QTL affecting a low-heritability trait requires the transfer of a large donor segment of ~10 cM flanked by two informative markers in the target QTL region given a relatively precise estimation of the QTL position (Visscher et al 1996).

As discussed in the above sections, epistasis and QE interactions are properties associated with most QTLs, which, unfortunately, have been largely ignored in most QTL mapping studies. Also, because most parents in the previous mapping experiments are not the parental lines used in common breeding programs, accurate MAS for QTLs cannot be practiced without information on QTL epistasis and QE interactions, both of which are difficult and very expensive to obtain. Thus, information from previous mapping studies may not be used directly in designing MAS experiments for QTLs in breeding programs. Other than epistasis and QE interactions, many important questions regarding QTLs remain unknown. For instance, for most traits, the number of polymorphic QTLs and the number of functional alleles at each of these loci in the gene pool are generally unknown. This is because almost all previous QTL mapping studies were conducted using biparental materials. Cross-population comparison can only provide limited information. In other words, the following questions remain to be answered before any MAS experiment for QTL transfer can be used: (1) Do, and to what extent, the mapping studies provide reliable information about the parameters (locations, effects, and linked markers) of QTLs for target traits to be used in breeding populations? (2) How many loci and alleles at each locus are there and how great are their individual and combined effects for each of the target traits? (3) Will the QTLs have the same phenotypic effects in the new genetic backgrounds (recipients) as the effects that were detected in the original mapping populations or parents (presence of epistasis)? (4) Are detected QTLs associated with other undesirable traits (genetic drag) and to what extent (genetic basis of genetic drag)?

A new strategy of molecular breeding has been proposed and is being used at IRRI for simultaneous QTL identification and introgression (Li et al 1999c). Briefly, this strategy is defined as a comprehensive backcross breeding program involving well-designed and organized breeding activities of international collaboration, and integration and use of DNA markers with phenotypic and genotypic (progeny testing) selection. The technical components of the strategy involve (1) inclusion of maximum genetic diversity in the parental lines by selecting two core sets of the best germplasm from the primary gene pool; (2) molecular and phenotypic characteriza-

tion of the parental lines; (3) enforced massive gene flow through maximum introduction of diversity from the primary rice gene pool into elite genetic backgrounds by backcrossing and phenotypic selection and developing many sets of near-isogenic introgression lines (NIILs) in the elite genetic backgrounds; (4) simultaneous identification, mapping, and transfer of large numbers of desirable QTLs for target traits using DNA markers and the NIILs; (5) development of new improved cultivars and large numbers of near-isogenic line sets for the core rice germplasm by using marker-aided QTL pyramiding; and (6) establishment of a molecular and phenotypic database for core rice germplasm.

Future perspectives

We conclude that QTLs are real and may include different kinds of genes or groups of genes that have varied and appreciable effects on complex phenotypes. Information on QTL epistasis and QE interactions is particularly lacking, thus limiting the application of the results from past QTL mapping efforts to MAS for genetic improvement of quantitative traits. In addition, major challenges remain regarding what genes underlie biochemical pathways and the physiology causing quantitative trait variation, even though a couple of successful cases in QTL cloning have been reported (Yano et al 2000, Zamir et al 2000). Fortunately, several recent developments in rice genome research will provide tremendous opportunities to solve these problems. These include complete physical maps of the rice genome, DNA sequence data of almost all rice expressed sequence tags (genes), and the complete genomic sequence data of individual rice chromosomes and the whole rice genome. The immediate results of these developments are new high-throughput functional genomic tools such as microarray and gene chips for large-scale gene expression studies and the availability of a virtually unlimited number of PCR-based DNA markers (SSRs and SNPs) covering the whole rice genome. All of these developments prophesy a golden era of rice functional genomics to connect DNA sequences to gene functions and to phenotypes, and applying this to rice improvement. Nevertheless, fine-QTL mapping and accurate phenotyping using unique genetic stocks such as near-isogenic lines will remain vital elements for the discovery and full understanding of the genes underlying biochemical pathways and physiology related to complex phenotypes.

References

Allard RW. 1988. Future directions in plant population genetics, evolution and breeding. In: International Symposium for Population Genetics and Germplasm Resources in Crop Improvement. Suppl. Vol.:9-18.

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251-1274.

Cho YG, McCouch SR, Kuiper M, Kang MR, Pot J, Groenen JTM, Eun MY. 1998. Integrated map of AFLP, SSLP, and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). Theor. Appl. Genet. 97:370-380.

Doebley J, Stec A, Gustus C. 1995. *teosinte branchedl* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 141:333-346.

Falconer DS. 1983. Introduction to quantitative genetics. Second edition, Longman Group Limited, London and New York.

Fukuta Y, Harushima Y, Yano M. 1996. Using quantitative trait locus analysis for studying genetic regulation of shattering. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Los Baños (Philippines): International Rice Research Institute. p 657-662.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

Huang N, Courtois B, Khush GS, Lin H, Wang G, Wu P, Zheng K. 1996. Association of quantitative trait loci for plant height with major dwarfing genes in rice. Heredity 77:130-137.

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95:313-320.

Kao CH, Zeng ZB, Teasdale RD. 1999. Multiple interval mapping for quantitative trait loci. Genetics 152:1203-1216.

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuta A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y. 1994. A 300-kilobase interval genetic map of rice including 883 expressed sequences. Nat. Genet. 8:365-372.

Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lark KG, Chase K, Adler F, Mansur LM, Orf JH. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc. Natl. Acad. Sci. USA 92:4656-4660.

Li ZK. 1997. Molecular analysis of epistasis affecting complex traits. In: Paterson AH, editor. Molecular analysis of complex traits. Boca Raton, Fla. (USA): CRC Press. p 119-130.

Li ZK. 1999. DNA markers and QTL mapping in rice. In: Ito O, O'Toole J, Hardy B, editors. Genetic improvement of rice for water-limited environments. Proceedings of the Workshop on Genetic Improvement of Rice for Water-Limited Environments, 1-3 December 1998. Los Baños (Philippines): International Rice Research Institute. p 173-182.

Li ZK, Pinson SRM, Stansel JW, Park WD. 1995a. Identification of quantitative trait loci (QTL) for heading date and plant height in rice using RFLP markers. Theor. Appl. Genet. 91:374-381.

Li ZK, Pinson SRM, Marchetti MA, Stansel JW, Park WD. 1995b. Characterization of quantitative trait loci (QTL) contributing to sheath blight (*Rhizoctonia solani*) in rice. Theor. Appl. Genet. 91:382-388.

Li ZK, Pinson SRM, Paterson AH, Park WD, Stansel JW. 1997a. Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145:453-465.

Li ZK, Pinson SRM, Paterson AH, Park WD, Stansel JW. 1997b. Genetics of hybrid sterility and hybrid breakdown in an inter-subspecific rice (*Oryza sativa* L.) population. Genetics 145:1139-1148.

Li ZK, Pinson SRM, Paterson AH, Stansel JW. 1998. Genetic dissection of the source-sink relationship in rice. Mol. Breed. 4:419-426.

Li ZK, Paterson AH, Pinson SRM, Stansel JW. 1999a. RFLP facilitated analysis of tiller and leaf angles in rice (*Oryza sativa* L.). Euphytica 109(2):79-84.

Li ZK, Luo L, Tabien R, Mei H, Paterson AH et al. 1999b. A 'defeated' resistance gene acts as a QTL against a virulent strain of *Xanthomonas oryzae* pv. *oryzae*. Mol. Gen. Genet. 261:58-63.

Li ZK, Shen LS, Courtois B, Lafitte R. 1999c. Development of near isogenic introgression line (NIIL) sets for QTLs associated with drought tolerance in rice. Proc. Int. Workshop for Molecular Approaches for the Genetic Improvement of Cereals for Stable Production in Water-Limited Environments. CIMMYT, El Batán, Mexico.

Li ZK, Luo LJ, Mei HW, Shu QY, Wang DL, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH. 2001a. Overdominant epistatic loci are the primary basis of inbreeding depression and heterosis in rice: I. Biomass and grain yield. Genetics. (In press.)

Li ZK, Yu SB, Huang N, Courtois B, Hittlemani S, McLaren G, Khush GS, Shashidhar HE, Zhuang JY, Zheng KL, Liu GF, Wang GC, Singh VP, Sidhu JS, Srivantaneeyakul S. 2001b. QTL × environment interactions in rice: I. Plant height and heading date. Genetics. (Submitted.)

Li ZK, Sanchez A, Angeles E, Singh S, Domingo J, Huang N, Khush GS. 2001c. Are the dominant and recessive plant resistance genes similar? A case study of rice R genes and *Xanthomonas oryzae* pv. *oryzae* races. Genetics. (In press.)

Lin HX, Qian HR, Zhuang JY, Lu J, Min SK, Xiong ZM, Huang N, Zheng KL. 1996. RFLP mapping of QTLs for yield and related characters in rice (*Oryza sativa* L.). Theor. Appl. Genet. 92:920-927.

Lin HX, Yamamoto T, Sasaki T, Yano M. 2000. Characterization and detection of epistatic interactions of 3 QTLs, Hd1, Hd2, and Hd3, controlling heading date in rice using nearly isogenic lines. Theor. Appl. Genet. 101:1021-1028.

Luo LJ, Li ZK, Mei HW, Shu QY, Wang DL, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH. 2001. Overdominant epistatic loci are the primary basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics. (In press.)

Mather K, Jinks JL. 1982. Biometrical genetics. 3rd ed. London (UK): Chapman and Hall.

Nandi S, Subudhi PK, Senadhira D, Manigbas NL, Sen-Mandi S, Huang N. 1997. Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. Mol. Gen. Genet. 255:1-8.

Nilsson-Ehle H. 1909. Kreuzunguntersuchungen an Hafer und Weizen, Lund.

Sanchez AC, Brar DS, Huang N, Yano M, Li ZK, Sasaki T, Khush GS. 2000. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci. 40:792-797.

Shen L, Courtois B, McNally K, Robin S, Li ZK. 2001. Evaluation of near-isogenic lines of rice introgressed with QTLs for root traits through marker-aided selection. Theor. Appl. Genet. (In press.)

Shomura A, Yano M, Harushima Y, Sato M, Lin SY et al. 1997. The current molecular map of rice. Proc. Int. Conf. Plant & Animal Genome V. San Diego, Calif. (USA): Scherago International Inc. p 87.

Sripongpangkul K, Posa GBT, Senadhira DW, Brar D, Huang N, Khush GS, Li ZK. 2000. Genes/QTLs affecting flood tolerance in rice. Theor. Appl. Genet. 101:1074-1081.

Tanksley SD, Hewitt JD. 1988. Use of molecular markers in breeding for soluble solids in tomato: a re-examination. Theor. Appl. Genet. 75:811-823.

Templeton AR. 1980. The theory of speciation via the founder principle. Genetics 94:1011-1038.

Visscher PM, Haley CS, Thompson R. 1996. Marker-aided introgression in backcross breeding programs. Genetics 144:1923-1932.
Wang DL, Zhu J, Li ZK, Paterson AH. 1999. Mapping QTLs with epistatic effects and genotype × environment interactions by mixed linear model approaches. Theor. Appl. Genet. 99:1255-1264.
Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. VI. Int. Congr. Genet. 1:356-366.
Wright S. 1951. The genetic structure of populations. Ann. Eugen. 15:323-354.
Xiao JH, Li J, Yuan L, Tanksley SD. 1995. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745-754.
Xiong L, Liu ZKD, Dai XK, Wang SW, Xu CG, Zhang DP, Maroof MAS, Sasaki T, Zhang Q. 1997. A high density RFLP map based on the F_2 population of a cross between *Oryza sativa* and *O. rufipogon* using Cornell and RGP markers. Rice Genet. Newsl. 14:110-115.
Xu JL, Zhong DB, Yu SB, Luo LJ, Khush GS, Li ZK. 2001. Molecular dissection of the primary sink size in rice (*Oryza sativa* L.). Theor. Appl. Genet. (Submitted.)
Xu K, Mackill D. 1995. RAPD and RFLP mapping of a submergence tolerance locus in rice. Rice Genet. Newsl. 12:244-245.
Yano M, Marushima Y, Nagamura Y, Kurata N, Minobe Y, Sasaki T. 1997. Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. Theor. Appl. Genet. 95:1025-1032.
Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Sasaki T. 2000. *Hd1*, major photoperiod-sensitivity QTL in rice, encodes a protein with the structure of zinc finger transcription factor. Abstracts of VIII International Conference on the Status of Plant and Animal Genome Research. San Diego, California, p 41.
Yu SB, Li JX, Xu CG, Tan YF, Gao Y, Li XH, Zhang Q, Saghai Maroof MA. 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. USA 94:9226-9231.
Zamir D, Fridman E, Pleban T. 2000. Map based cloning of BRIX9-2-5: a wild species QTL that increases sugar content in tomato. Abstracts of VIII International Conference on the Status of Plant and Animal Genome Research. San Diego, California. p 4.
Zeng ZB. 1994. The precision mapping of quantitative trait loci. Genetics 136:1457-1468.
Zhong DB, Yu SB, Xu JL, Luo LJ, Li ZK. 1999. QTLs for grain shattering in the Lemont/Teqing RI population. Rice Genet. Newsl. 16:71-73.

Notes

Author's address: Plant Breeding, Genetics, and Biochemistry Division, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Genetic and molecular basis of heterosis in rice

Qifa Zhang, Jinping Hua, Sibin Yu, Lizhong Xiong, and Caiguo Xu

We have conducted a series of studies to elucidate the biological mechanism of heterosis using rice as the model system. We analyzed the genetic basis of heterosis using an $F_{2:3}$ population and an "immortalized" F_2 population derived from a cross between Zhenshan 97 and Minghui 63. Our results demonstrated the involvement of large numbers of two-locus interactions, or epistasis, underlying the genetic basis of quantitative traits and heterosis. We assessed the relationship between gene expression and heterosis by assaying the patterns of differential gene expression in hybrids relative to their parents in a diallel cross. The analysis revealed that differentially expressed fragments occurring in only one parent of the cross were positively correlated with heterosis, and fragments detected in F_1s but not in the respective parents were negatively correlated with heterosis. For further analysis, 384 fragments were recovered from gels and arrayed onto nylon membranes. Hybridization with RNAs from seedling and flag leaf tissues detected an overall elevated level of gene expression in the hybrid compared with the parents. Several fragments showed much higher expression in the highly heterotic hybrid than in other hybrids. Many of these fragments were sequenced and mapped to the rice linkage map, which provided insights into the understanding. We believe that analyses combining genetic and molecular approaches will eventually lead to the characterization of the biological mechanisms of heterosis.

The use of heterosis has become one of the major strategies for increasing crop production. Hybrid varieties have contributed greatly worldwide to the production of many crop species, including the most important food crops such as maize and rice. There has also been considerable interest in the genetic basis of heterosis. Two hypotheses, the dominance hypothesis (Davenport 1908) and the overdominance hypothesis (East 1908, Shull 1908), were proposed early in the 20th century to explain the genetic basis of heterosis. Although many investigators favored one hypothesis over the other, data allowing for critical assessments of the hypotheses remained largely

unavailable until recently with the advent of molecular marker technology and high-density molecular linkage maps. Genetic analyses based on high-density molecular maps have been reported recently in maize and rice (Stuber et al 1992, Xiao et al 1995, Yu et al 1997).

In the past ten years, our group has conducted a series of genetic and molecular studies with the long-term goal of elucidating the biological mechanism of heterosis using rice as the model system. This series includes characterization of the relationship between molecular marker heterozygosity and hybrid performance and heterosis in a wide range of rice germplasm (Zhang et al 1994, 1995, 1996, Saghai Maroof et al 1997, Zhao et al 1999). The general finding of such analyses is that the correlation between molecular marker heterozygosity and heterosis varies with the germplasm analyzed, which suggested the need for a detailed understanding of the genetic basis of heterosis. To characterize the genetic basis of heterosis, we have focused our efforts on analyzing genetic materials derived from Shanyou 63, a cross between Zhenshan 97 and Minghui 63, the most widely used hybrid in rice production in China. The hybrid of this cross is highly heterotic and was planted with an annual area of 6.7 million hectares in its peak period, accounting for approximately 25% of the rice production of China. The most noticeable finding of the analysis is the existence of a large number of epistatic interactions in the genome that play an important role in the genetic basis of heterosis (Yu et al 1997). Another component of our work is the assessment of the relationship between gene expression and heterosis, with the objective of identifying genes involved in the manifestation of heterosis.

In this chapter, we will report on our studies in characterizing the genetic basis of heterosis and the analyses of the relationship between gene expression and heterosis.

Analyzing the genetic basis of heterosis using an $F_{2:3}$ population

Most of this analysis has been published previously (Yu et al 1997). We will give a brief description here for the purpose of comparison.

The experimental population and data collection

The experimental population consisted of 250 F_3 families, each derived from selfed (bagged) seeds of a single F_2 plant, from a cross between Zhenshan 97 and Minghui 63. The F_3 families, two parents, and F_1 were transplanted to a bird-net-equipped field on the experimental farm of Huazhong Agricultural University in the 1994 and 1995 rice-growing seasons in Wuhan, China. Traits examined were yield per plant, tillers per plant, grain weight, and grains per panicle. The 250 families were assayed using 150 molecular markers that detected polymorphisms between the parents.

Single-locus QTLs for yield and yield component traits

A molecular marker linkage map was constructed using Mapmaker 3.0 (Lincoln et al 1992a), which spanned a total of 1,842 cM in length and well integrated the markers from the two high-density restriction fragment length polymorphism (RFLP) linkage maps (Causse et al 1994, Kurata et al 1994). The QTLs (quantitative trait loci) given

in Table 1 for the four traits were determined using Mapmaker/QTL 1.1 (Lincoln et al 1992b) with LOD 2.4. QTLs with LOD exceeding 2.4 in one year but slightly below this threshold in another year are also included in the list. A total of 34 QTLs were detected for the four traits; 10 were observed in both years and the remaining 24 were detected in only one year.

Two features of the QTLs detected for yield and yield component traits were revealed. First, overdominance was evident at six of the ten QTLs detected for yield, whereas partial dominance was prevalent among the QTLs detected for other traits. Second, the additive effects of the QTLs indicated that, for all the traits, alleles from the parent of higher value contributed to the increase in trait scores at some QTLs, whereas, at other QTLs, alleles from the parent of lower value contributed to the increase in performance (Table 1).

Table 1. QTLs detected for yield and yield component traits by analyzing data of the $F_{2:3}$ population from the cross between Zhenshan 97 and Minghui 63 tested in 1994 and 1995 (adapted from Yu et al 1997).

Trait	QTL[a]	Flanking markers	LOD	Var (%)	*A*[f]	D	IL
1994							
Yield (g)	*yd1*	R753-C161	3.5	9.7	–0.3	3.2	0
	yd1b	RG101-C922	2.6	5.8	–1.7	–0.3	3
	yd4	R514-C2807	2.4	7.4	–1.8	1.4	4
	yd5	G193-RZ649	3.3	11.7	–0.5	3.4	3
	yd8	C483-R1629	3.2	6.6	0.8	2.4	1
Tillers plant^{-1}	*tp4*	C820-C56	3.1	6.1	–0.4	–0.1	2
	tp5	G1458-G193	2.8	5.5	–0.4	0.2	2
	tp10	C677-G4003	2.7	5.3	–0.2	–0.5	4
Grains panicle^{-1}	*gp1a*	R753-C161	3.1	6.1	–5.2	4.4	0
	gp1b	RG173-RG532	4.2	10.6	–7.9	–3.5	4
	gp3	R1966-G144	8.3	16.9	10.0	–3.6	1
	gp5[b]	G193-RZ649	2.4	8.9	–1.3	10.3	1
	gp6[b]	RG653-G342	2.5	4.9	4.1	–5.6	4
Grain weight (g)	*gw1*[b]	R753-C161	3.5	12.1	0.7	1.5	1
	gw3[c,d]	R19-RZ403	11.4	23.6	–1.8	–0.4	8
	gw5[c]	RG360-C734	8.5	15.7	1.3	0.1	1
	gw6	R1952-C226	2.6	5.2	0.7	0.5	3
	gw7[b,d,e]	RG128-C1023	2.3	8.6	–1.0	0.2	3
	gw9	R2638-RG570	4.1	8.9	1.2	–0.3	0
	gw11[e]	RM4-RG98	2.2	5.6	–0.6	0.7	1
1995							
Yield (g)	*yd1*	R753-C161	2.6	6.2	–1.1	2.1	1
	yd5a	C624-C246a	2.7	10.2	–1.4	2.9	0
	yd5b	RG360-R1674	2.5	4.8	1.5	0.2	4
	yd6	R565-R902	2.8	5.2	1.7	0.4	4
	yd7[b]	RG128-C1023	3.5	9.4	–1.8	2.3	2
	yd11[b]	C950-G389b	3.2	7.0	–0.5	2.7	1
Tillers plant^{-1}	*tp4*	C820-C56	2.6	5.5	–0.3	–0.3	0
	tp7	RZ471-MX2	3.2	11.3	0.5	–0.5	6
Grains panicle^{-1}	*gp1a*	RM1-R753	6.0	15.5	–8.1	3.4	10
	gp1b[b]	RG173-RG532	7.1	17.8	–9.1	–2.8	13
	gp3[b]	RZ403-C269	9.0	16.0	8.7	–0.5	4

continued on next page

Table 1 continued.

Trait	QTL[a]	Flanking markers	LOD	Var (%)	*A*[f]	D	IL
	gp5	G193-RZ649	2.3	8.6	–0.2	8.8	3
	gp6b	R1014-G200	2.5	5.5	5.0	0.9	0
	gp7	C1023-R1440	4.7	9.5	–5.2	6.5	2
	gp11	C950-G389b	3.1	5.9	–0.7	7.3	0
Grain weight (g)	*gw1*	R753-C161	2.6	9.7	0.5	1.3	0
	gw3a	C944-C746	9.5	16.8	–1.4	0.2	3
	gw3b	R1966-G144	11.4	22.5	–1.7	–0.7	9
	gw5	RG360-R1674	12.9	24.1	1.6	–0.3	0
	gw6b	R565-R902	4.7	8.6	1.0	0.3	5
	gw7	RG128-C1023	5.2	17.7	–1.4	–0.4	9
	gw8	C1121-RG333	2.7	6.7	0.7	–0.7	5
	gw9	R2638-RG570	4.0	8.5	1.1	–0.4	1
	gw11	RM4-RG98	3.5	8.6	–0.9	0.6	0

[a]Numbers following the two letters represent the chromosomal locations of the QTLs. Var = variation explained by each QTL, A = additive effect, D = dominance effect, IL = number of loci with which the QTL interacts. [b, c, d, e]Pairs of interacting loci within a trait. [f]Positive values of the additive effect indicate that alleles from Zhenshan 97 increase the trait scores and negative values indicate that alleles from Minghui 63 increase the score.

Correlations between overall heterozygosity of the F_2 individuals and the F_3 family means were very small for all four traits (data not shown). They were significant at $P \leq 0.05$ only for grain weight (correlation coefficients 0.17 in 1994 and 0.16 in 1995). Thus, the overall heterozygosity made a very small contribution to trait expression.

Digenic interactions between loci in the entire genome

The entire genome was searched for digenic interactions for each trait with two-way analyses of variance using all possible two-locus combinations of the marker genotypes. The interaction, often referred to as epistasis, was further partitioned into four terms each specified by a single degree of freedom: additive (first locus) × additive (second locus) (AA), additive × dominance (AD), dominance × additive (DA), and dominance × dominance (DD). Table 2 presents the number of two-locus combinations showing significant ($P < 0.01$) interactions resulting from testing all possible two-locus combinations and also the number of various types of interactions determined by orthogonal contrast tests. Only data sets formed of two codominant loci with each class of the marker genotypes containing five or more individuals were included in the calculation, resulting in 7,585 tests for 1994 and 7,681 for 1995. With $\alpha = 0.01$ for individual tests, the 99% confidence intervals for the expected numbers of spurious interactions (false positive) would be 75.85 ± 20.02 for 1994 and 76.81 ± 20.14 for 1995. It is clear from Table 2 that the number of interactions detected in both years for all the traits far exceeded expectations based on spurious interactions, indicating the existence of real interactions in the genome of this population.

Significant interactions were detected simultaneously in both years for many of the two-locus combinations (Table 2). These interactions, referred to as common interactions for ease of description, can be taken as the minimum of statistically significant

Table 2. A summary of the significant ($P \leq 0.01$) interactions identified in 1994 and 1995 in the $F_{2:3}$ population by searching all possible two-locus combinations (adapted from Yu et al 1997).

Trait	Interaction	1994	1995	Common
Yield	Positive pairs[a]	105	165	10
	AA	60	91	8
	AD (DA)	51	73	3
	DD	4	18	0
Tillers plant^{-1}	Positive pairs[a]	105	141	17
	AA	79	105	17
	AD (DA)	28	42	1
	DD	10	6	0
Grains panicle^{-1}	Positive pairs[a]	99	160	15
	AA	52	80	9
	AD (DA)	56	74	10
	DD	4	16	0
Grain weight	Positive pairs[a]	125	164	49
	AA	84	102	27
	AD (DA)	47	71	19
	DD	15	16	9
Number of tests[b]		7,585	7,681	

[a]Number of two-locus combinations showing significant interactions ($P \leq 0.01$).
[b]Number of possible two-locus combinations tested.

interactions. The proportion of common interactions was the largest for grain weight. They made up approximately 40% and 30% of the significant interactions detected in 1994 and 1995, respectively. Common interactions ranged from 6% to 17% of the significant interactions for the remaining three traits in the two years (Table 2).

The number of three different types of interaction (AA, AD/DA, and DD) determined using the orthogonal contrast test is also given in Table 2. All three types of interaction were found in all the traits and the types of interaction were also very consistent between the two years. The loci involved in the interactions, as indicated by the molecular markers, were distributed in all 12 chromosomes. Overall, AA interactions were the most frequent and DD interactions were the least frequent. In general, each of the interactions had only a small effect on the trait by explaining on average 3.0% of the genotypic variation of the trait. In some cases, two different types of interaction were detected for the same two-locus pairs. In other cases, the same two-locus pair showed interaction effects on two or more traits, which may be referred to as the pleiotropic effect of epistasis.

The effects of epistasis on single-locus QTLs

Interactions between QTLs were detected in grains per panicle and grain weight in 1994 and in yield and grains per panicle in 1995 (Table 1). The most noteworthy case was that of grain weight in 1994, in which significant interactions were detected in four pairs of QTLs. Interactions of QTLs with loci in the rest of the genome were also detected using markers flanking the QTLs and markers located in

the remaining parts of the linkage map. The majority of the QTLs were found to interact with at least one other locus and many of the QTLs interacted with three or more loci (Table 1).

Analyzing the genetic basis of heterosis using an "immortalized" F_2 population

It is well known that the F_2 generation theoretically provides the most complete and most informative population for genetic analysis (Allard 1956). However, it is very difficult to use the F_2 for genetic analysis of quantitative traits, as each genotype is represented by only a single individual, which makes it difficult to get replicated measurements of the same genotype. Also, the population is in a transient state; thus, the experiment cannot be repeated. Although genetic analyses using the $F_{2:3}$ type of populations can produce a considerable amount of useful information on the genetic components underlying heterosis, such analyses suffer from several shortcomings that are inherent in the data generated from the populations. First, F_3 families are genetically heterogeneous; thus, it is impossible to have exact replications in the field trials. Second, an additional cycle of meiosis would result in gene combinations that are different from those in the F_2 generation; thus, the genotypes of $F_{2:3}$ families do not correspond exactly with those of F_2 individuals. And, third, one generation of self-fertilization theoretically reduces the level of heterozygosity by half; thus, data from the $F_{2:3}$ type of populations underestimate dominant types of genetic effects, such as dominance and overdominance at the single-locus level and dominant types of interaction at the two-locus level. To overcome these shortcomings, we designed a new mating scheme that generated an "immortalized" F_2 population, using a population of 240 recombinant inbred lines (RILs) produced by the single-seed descent method from the Zhenshan 97/Minghui 63 cross.

Experimental design and data collection

In this design, crosses were made between the RILs chosen by random permutations of the 240 RILs. In each round of permutation, the 240 RILs were randomly divided into two groups and the lines in the two groups were paired at random without replacement to provide parents for 120 crosses. This procedure can be repeated as many times as desired and each round of permutation will pair the parents for 120 crosses. Such a mating scheme would result in a population consisting of crosses in multiples of 120.

For heterosis study, such a population possesses several distinct advantages. First, the genotypes and their proportions in this population are similar to those of an F_2 population (i.e., 1:2:1 for single-locus genotypes and 1:2:1:2:4:2:1:2:1 for two-locus combinations), thus allowing for more complete and direct analyses of the genetic components. Second, instead of only one individual per genotype represented in an F_2 population, each genotype in this population is represented by as many plants as desired, thus permitting replicated trials. The whole population can be recreated when needed, either in exactly the same way or by different permutation schemes. Thus, the field experiment can be repeated in multiple environments. Third, plants in each of the lines are

hybrids rather than progenies of self-fertilization. Fourth, the molecular marker data need to be collected from only the 240 RILs no matter how many crosses are included in the population. Additionally, this population provides opportunities for genetic mapping of heterosis per se rather than analyses based on performance measurements of the trait if the hybrids and the parents for each cross are planted side by side in the field, which allows for measuring the level of heterosis for every cross.

The field trials of 360 crosses were conducted in the rice-growing seasons of 1998 and 1999 following a randomized complete block design with two replications within each year. Each block consisted of four rows with 12 plants per row: two rows of the hybrid and one row for each of the parents. Hybrid plants of the crosses were identified by morphological comparison with the parents and SSR (simple sequence repeats) assay. Data for agronomic performance were scored in the same manner as in the $F_{2:3}$ analysis.

A total of 217 molecular markers, including 171 RFLPs and 46 SSRs, were used for assaying the 240 RILs. A linkage map was constructed based on this data set that provided the framework for QTL mapping. We have currently completed part of the analysis for the data collected in 1998, which will be presented briefly in the following sections.

Single-locus QTLs for yield and yield component traits

A total of 28 QTLs were detected for yield and the three traits that were components of yield (Table 3). This number is much larger than the numbers detected for a single year in the $F_{2:3}$ analysis. The most pronounced differences were the number of QTLs for tillers per plant and for grain weight; in both cases the numbers detected in this study were substantially greater than in the $F_{2:3}$ analysis. Overdominance was detected at three of the five QTLs for yield and two of the five QTLs for grains per panicle, but was not detected in the other two traits. This trend is also similar to the level of dominance observed in the $F_{2:3}$ study.

Digenic interactions

The data set was also subject to the two-way ANOVA for digenic interactions, as occurred in the $F_{2:3}$ analysis. A search of the 23,239 possible two-locus pairs that were possible with the data set resulted in 96 to 224 two-locus pairs that showed significant interactions at the 0.001 probability level (Table 4). Partitioning of the interactions using orthogonal contrasts revealed that all three types of interaction (i.e., AA, AD or DA, and DD) were involved in the interactions. AA interactions occurred much more frequently than AD ones, which in turn were more frequent than DD interactions. This trend is also similar to the results of $F_{2:3}$ analysis. It should be noted that, although the screening was performed statistically at a much higher stringency than was done in the $F_{2:3}$ analysis, the number of epistatic interactions that were detected was approximately of the same order of magnitude as those detected in the $F_{2:3}$ analysis.

In summary, the analyses of the data from both the $F_{2:3}$ and immortalized F_2 populations detected the existence of a large number of two-locus epistatic interactions in the rice genome. Epistasis clearly plays a significant role in the inheritance of quantitative traits as well as in the genetic basis of heterosis. Thus, the relationship between

Table 3. QTLs detected for yield and yield component traits in the analysis of the "immortalized" F_2 population for the field trial of 1998.

Trait	QTL	Flanking markers	LOD	Var (%)	A	D
Yield (g)	*yd6*	R2147-C751a	4.76	13.9	3.37	–2.75
	yd7	R1440-G678	2.60	3.6	–1.96	0.21
	yd9a	G570-G667	3.20	4.5	–1.23	3.45
	yd9b	RM201-RM215	3.05	12.5	–1.21	5.25
	yd11	R543a-RZ536	2.46	3.5	0.07	2.72
Tillers plant^{-1}	*tp1*	G236-C112	3.40	6.3	–0.80	–0.45
	tp3	C316-C63	3.73	5.2	0.79	0.27
	tp5a	C734b-RZ649	3.31	5.6	–0.86	–0.20
	tp5b	RM26-C1447	5.13	7.1	1.08	0.20
	tp6a	RZ398-R1014	3.00	4.3	0.78	–0.78
	tp6b	P-G200	3.81	5.3	–0.29	–1.01
Grains panicle^{-1}	*gp1*	RM237-C922	3.08	5.5	0.99	8.22
	gp3	RZ403-C1087	3.96	6.3	6.38	0.45
	gp6	R2147-C751a	4.51	8.8	7.42	–3.73
	gp7	C1023-R1440	4.53	6.6	–6.38	2.22
	gp12	R887-G1314b	3.19	4.5	4.79	5.46
Grain weight (g)	*kgw1a*	G359-G532	4.52	7.1	0.75	0.24
	gw1b	C86-G236	2.68	4.5	0.59	0.07
	kgw3a	G393-G144	10.04	13.4	–1.01	–0.33
	kgw3b	RZ403-C1087	17.88	26.0	–1.40	–0.76
	kgw5	G360-C734b	8.33	13.0	1.03	–0.08
	kgw7	G528-G128	4.71	6.9	–0.73	0.24
	kgw7	G678-RZ471	2.79	4.4	–0.58	–0.41
	kgw8	R1394-R727	2.99	4.8	0.41	–0.61
	kgw9	RM242-G570	3.22	4.8	0.70	–0.05
	kgw10	C1633-C677	3.31	5.3	0.55	0.53
	kgw11	G44-G257	2.42	3.7	–0.65	0.36
	kgw12	G1128a-R887	3.57	5.0	0.64	0.45

Table 4. Significant interactions identified for performance of yield and yield component traits by searching all possible (23,239 in total) two-locus combinations.

Trait	Pairs[a]	AA[b]	AD (DA)[b]	DD[b]
Yield	134	116	70	16
Tillers plant^{-1}	96	86	50	7
Grains panicle^{-1}	100	79	51	14
Grain weight	224	185	78	16

[a]Significant at the 0.001 probability level. [b]Significant at the 0.01 probability level.

traits and genes in the manifestation of heterosis is much more complex than has commonly been expected on the basis of dominance and overdominance hypotheses.

Gene expression and heterosis

We have also been interested in the last several years in the relations between gene expression and heterosis, with the belief that the manifestation of heterosis is the direct outcome of gene expression.

Correlation between gene expression in leaves and heterosis

Eight commonly used parental lines of hybrid rice were employed in this study including three restorer lines (Ce 64-7, Minghui 63, and Teqing) and maintainers of five male sterile lines (Guang B, Maxie, Qing Si Ai, Xian Gai, and Zhenshan 97). These lines were intermated in all possible nonreciprocal combinations to form a diallel set of 28 crosses, which were previously examined for yield and several agronomic characters in replicated field trials. This is the same diallel set of parents and hybrids that Zhang et al (1994, 1995) used to assess the relationship between genetic diversity of molecular markers and heterosis.

Differential display (Liang and Pardee 1992) of the mRNA was conducted using flag leaf tissue from each of the parents and hybrids. The details of the analysis were described previously (Xiong et al 1998). Six of the most informative primer combinations, displaying the largest number of differential bands, were used to survey the level of differential expression among the various crosses of the diallel set. Among a total of approximately 350 bands that could be repeatedly amplified using the six primer pairs, 135 fragments represented sequences that were differentially expressed among the parents and F_1s in at least one cross. The analysis revealed several patterns of differential expression including (1) bands present in one parent and F_1 but absent in the other parent (DMPT), (2) bands observed in both parents but not in the F_1 (ABF_1), (3) bands occurring in only one parent but not in the F_1 or the other parent (UNP), and (4) bands detected only in the F_1 but in neither of the parents (UNF_1). Both banding patterns of ABF_1 and UNF_1 signified the uniqueness of the F_1 compared with the parents (DF_1).

We calculated the correlations of the various banding patterns with the performance and heterosis of the hybrids (Table 5). In general, the correlation between differential expression and hybrid performance is low. However, two types of differential expression showed highly significant correlations with heterosis: UNF_1 was negatively correlated with the heterosis of all traits, whereas UNP was positively correlated with heterosis.

We also calculated the correlation between differential gene expression and molecular marker heterozygosity (Table 6). UNF_1 was correlated with general heterozygosity that referred to the overall level of heterozygosity of the hybrid (Zhang et al 1994), and was negatively correlated with the specific heterozygosity of all six traits

Table 5. Correlations of various patterns of differential expression with heterosis of six traits (adapted from Xiong et al 1998).

Population	Plant height[a]	Biomass	Tillers plant^{-1}	Grains panicle^{-1}	Grain weight	Yield
ABF_1	0.255	0.307	0.070	0.375	0.271	0.362
UNF_1	–0.454*	–0.538**	–0.449*	–0.429*	–0.517**	–0.547**
DF_1	–0.197	–0.229	–0.360	–0.065	–0.242	–0.187
DMP_T	–0.144	–0.128	–0.041	–0.153	–0.208	–0.135
UNP	0.437*	0.486**	0.370	0.288	0.577**	0.377

[a]*, ** = statistically significant at the 0.05 and 0.01 probability levels, respectively.

Table 6. Correlations of various patterns of differential expression with general and specific heterozygosity for each of the six traits (adapted from Xiong et al 1998).

Population	General heterozygosity[a]	Specific heterozygosity					
		Plant height	Biomass	Tillers plant^{-1}	Grains panicle^{-1}	Grain weight	Yield
ABF_1	0.196	0.195	0.243	0.294	0.156	0.125	0.170
UNF_1	–0.618**	–0.509**	–0.551**	–0.508**	–0.457*	–0.427*	–0.526**
DF_1	–0.404*	–0.304	–0.300	–0.213	–0.289	–0.289	–0.341
DMPT	–0.149	–0.173	–0.195	–0.239	–0.040	0.049	–0.125
UNP	0.191	0.520**	0.527**	0.006	0.440*	–0.052	0.528**

[a]*, ** statistically significant at the 0.05 and 0.01 probability levels, respectively.

analyzed. In contrast, UNP was positively correlated with specific heterozygosity for most of the traits.

Analysis of the differentially expressed fragments

The differentially expressed fragments were excised from the gels, reamplified with polymerase chain reaction, and cloned into plasmid vectors. A total of 384 clones were recovered and arrayed onto nylon membranes. RNA samples were isolated from three-leaf stage seedling and flag-leaf tissues of the hybrid Shanyou 63 and its two parents, Zhenshan 97 (male sterile line) and Minghui 63 (restorer). Probes were prepared from the RNA tissues by reverse transcription labeled with α-^{32}P and hybridized with the nylon membranes. The intensity of the hybridization signal was scanned to determine the level of expression of the cloned fragments in the respective tissues.

Unlike the presence vs. absence type of difference revealed by differential display, the differences in hybridization signal among the clones were quantitative rather than qualitative. In flag-leaf tissue, a large number of clones (114) showed a higher expression in Zhenshan 97 than in Minghui 63, whereas only a small number of clones (36) showed a higher expression in Minghui 63 than in Zhenshan 97 (Table 7). The reverse was the case in the seedling tissue. Many more clones showed a higher expression in the hybrid than in both parents in both tissues, with the possible exception of the comparison of Zhenshan 97 with the hybrid in flag-leaf tissue, in which approximately the same number of clones showed a higher expression in the hybrid as the number showing a higher expression in Zhenshan 97. Thus, collectively, the hybrid had an obvious advantage in terms of the number of genes showing elevated expression compared with the parents.

We further calculated the level of heterosis in gene expression, treating the reading of the signal intensity as a quantitative trait. At the seedling stage, a much larger number of clones showed positive heterosis in expression than the number showing negative heterosis. At the flag-leaf stage, a slightly larger number of clones showed negative heterosis than those showing positive heterosis (Table 8). Again, the hybrid had an advantage in the level of gene expression over the parents, especially at the seedling stage.

Table 7. Numbers of cDNAs showing various levels of differential expression in parents and hybrids of the Shanyou 63 cross.

Group	Relative signal intensity	Zhenshan 97/ Minghui 63			Zhenshan 97/ hybrid			Minghui 63 hybrid		
		FL[a]	SD[b]	C[c]	FL[a]	SD[b]	C[c]	FL[a]	SD[b]	C[c]
Low ratio	0–0.2	12	28	6	18	48	8	17	28	5
	0.2–0.5	13	52	1	20	92	5	29	34	6
	0.5–0.67	11	62	4	31	71	9	60	64	8
	Subtotal	36	142	11	69	211	22	106	126	19
About equal	0.67–1	70	94	21	116	76	24	147	143	86
	1–1.5	134	50	27	102	28	11	83	55	17
	Subtotal	204	144	48	218	104	35	230	198	103
High ratio	1.5–2	53	18	7	35	6	1	5	6	0
	2–5	46	15	7	19	14	3	4	6	0
	>5	15	10	4	14	18	2	12	13	2
	Subtotal	114	43	18	68	38	6	21	25	2
Total		354	329	77	355	353	63	357	349	124

[a]FL = flag leaf. [b]SD = seedling. [c]C = in common between flag leaf and seedling.

Table 8. Number of cDNAs with various levels of expression of heterosis in the Shanyou 63 cross.

Groups	Heterosis (%)	Flag leaf	Seedlings	Common
Negative heterosis	<–100	18	14	2
	–100– –50	59	9	1
	Subtotal	77	23	3
No heterosis	–50– –15	9	21	8
	–15– +15	126	52	24
	15–50	85	90	20
	Subtotal	220	163	52
Positive heterosis	50–100	39	79	19
	100–200	14	51	5
	>200	11	19	3
	Subtotal	64	149	27
Total		361	335	82

Forty-eight of the clones showing large expression differences in the Shanyou 63 cross were selected to assay the expression in three hybrids and their parents, Zhenshan 97/Minghui 63, Zhenshan 97/Teqing, and Guang B/Teqing, representing highly heterotic, intermediate, and nonheterotic crosses, respectively, as reported previously (Zhang et al 1995). Twenty-seven of the 48 clones showed a higher expression in the Zhenshan 97/Minghui 63 hybrid than in the other two hybrids. Thirty-six of the 48 clones showed a higher heterosis in expression in the Zhenshan 97/Minghui 63 hybrid than in the other two hybrids.

We sequenced 35 of the clones and searched the sequences against the databases. Homology was found in the databases for 12 of the clones; 3 were homologous with chloroplast sequences and 9 were homologous with sequences with known functions.

The functions of these sequences need further study. We also mapped 12 of these clones to their respective chromosomal locations and the correspondence between the map locations and the single-locus QTLs and epistatic loci remains to be established.

Perspectives

The tools of genetic mapping and functional genomics that are available today have made it possible to precisely characterize the mechanism of heterosis. As demonstrated in our studies, genetic mapping using molecular markers can determine the genomic locations of loci (single-locus QTLs and also epistatic interactions) that control the performance and heterosis of the traits. Such analyses can also provide information regarding the magnitudes of the effects of the QTLs and epistatic loci on the trait. Gene expression analysis, using techniques such as differential display, subtractive cDNA cloning, or cDNA microarray, can lead to the identification of genes that are differentially expressed between the parents and the hybrid, or between highly heterotic hybrids and nonheterotic hybrids. There is also a considerable need for physiological and biochemical characterization of the nature of the genes involved in heterosis as well as the processes underlying heterosis. We believe that comprehensive analysis combining the differentially expressed genes, genetically determined QTLs and epistatic loci, and physiologically characterized biochemical processes will eventually lead to complete elucidation of the biology of heterosis.

References

Allard RW. 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235-278.

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu KS, Xiao JH, Yu ZH, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251-1274.

Davenport CB. 1908. Degeneration, albinism and inbreeding. Science 28:454-455.

East EM. 1908. Inbreeding in corn. Reports of the Connecticut Experiment Station for years 1907-1908. p 419-428.

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y. 1994. A 300-kilobase interval genetic map of rice including 883 expressed sequences. Nat. Genet. 8:365-372.

Liang P, Pardee AB. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967-971.

Lincoln S, Daly M, Lander E. 1992a. Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1. Whitehead Institute Technical Report, 2nd ed. Cambridge, Mass., USA.

Lincoln S, Daly M, Lander E. 1992b. Constructing genetic maps with Mapmaker/Exp 3.0. Whitehead Institute Technical Report, Cambridge, Mass., USA.

Saghai Maroof MA, Yang GP, Zhang Q, Gravois KA. 1997. Correlation between molecular marker distance and hybrid performance in U.S. southern long grain rice. Crop Sci. 37:145-150.

Shull GH. 1908. The composition of a field of maize. Am. Breed. Assoc. 4:296-301.

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander E. 1992. Identification of genetic factors contributing to heterosis in a hybrid from elite maize inbred lines using molecular markers. Genetics 132:823-839.

Xiao J, Li J, Yuan L, Tanksley SD. 1995. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745-754.

Xiong LZ, Xu CG, Saghai Maroof MA, Zhang Q. 1998. Patterns of cytosine methylation in an elite rice hybrid and its parental lines detected by a methylation-sensitive amplification polymorphism technique. Mol. Gen. Genet. 261:439-446.

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Saghai Maroof MA. 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. USA 94:9226-9231.

Zhang Q, Gao YJ, Yang SH, Ragab RA, Saghai Maroof MA, Li ZB. 1994. A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor. Appl. Genet. 89:185-192.

Zhang Q, Gao YJ, Saghai Maroof MA, Yang SH, Li JX. 1995. Molecular divergence and hybrid performance in rice. Mol. Breed. 1:133-142.

Zhang Q, Zhou ZQ, Yang GP, Xu CG, Liu KD, Saghai Maroof MA.1996. Molecular marker heterozygosity and hybrid performance in indica and japonica rice. Theor. Appl. Genet. 93:1218-1224.

Zhao MF, Li XH, Yang JB, Xu CG, Hu RY, Liu DJ, Zhang Q. 1999. Relationship between molecular marker heterozygosity and hybrid performance in intra- and inter-subspecific crosses of rice. Plant Breed. 118:139-144.

Notes

Authors' address: National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China.

Acknowledgment: The research presented in this chapter was supported by a grant from the National Natural Science Foundation of China and a grant from the Rockefeller Foundation.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Structural and functional genomics

The International Rice Genome Sequencing Project: progress and prospects

T. Sasaki, T. Matsumoto, T. Baba, K. Yamamoto, J. Wu, Y. Katayose, and K. Sakata

The rice genome sequencing project has been pursued as a national project in Japan since 1998. At the same time, a desire to accelerate the sequencing of the entire rice genome led to the formation of the International Rice Genome Sequencing Project (IRGSP), initially comprising five countries. The sequencing strategy is the conventional clone-by-clone shotgun method using P1-derived artificial chromosome/bacterial artificial chromosome (PAC/BAC) libraries from rice variety Nipponbare as a common template resource. As of September 2000, ten countries from this international collaboration had already contributed about 30 Mb of the rice genome sequence. Analysis of the rice genome should facilitate a better understanding of the concept of inheritance in the rice plant and the development of new research endeavors in physiology and biochemistry. Crucial information from nucleotide sequences will be useful for improving breeding technology as one of the ultimate goals of rice genome research.

Rice is undoubtedly a dominant and important staple worldwide. The global production of rice reached 600 million tons in 1999, which is roughly equivalent to the production volume achieved in maize and wheat. However, to meet the expected increase in the population (8 billion by 2020) that depends largely on rice as a direct or indirect source of food, rice production needs to be sustained for at least the next 50 years. One way to address this problem is to ensure the improvement of current rice yields and expand agricultural area for rice production. Conventional breeding, however, has been the sole or principal approach for crop improvement, particularly in the development of many elite rice varieties. The actual situation is far from ideal. The population increase in Asia, Africa, and Latin America is expected to continue beyond modest estimates and breeding of rice varieties that are tolerant of biotic and abiotic stresses on those continents remains a feasible way to meet this challenge.

The boom in rice genome analysis is closely linked to the facts described above. Refinements in genome research technology have progressed well during the past 10 years such that nucleotide sequencing is now 20-fold faster. Simply put in a more concrete concept, sequencing the entire rice genome is a goal that can be realistically accomplished. The development of basic tools such as a fine molecular genetic map with 2,300 markers (Harushima et al 1998), a physical map with YACs (yeast artificial chromosomes) covering 60% of the rice genome (Saji et al 2001), and 10,000 independent ESTs (expressed sequence tags) (Yamamoto and Sasaki 1997) has greatly facilitated the launching of the whole-genome sequencing of rice. Without these tools and information, genome sequences are only meaningless digitized codes that are irrelevant to the study of inheritance. For example, How can genetically unassigned sequences be used to improve rice quality? or How can genes controlling the targeted phenotype be identified? Only by combining genetically reliable data can genome sequence information provide a clue for advanced studies to overcome biotic or abiotic stresses in rice plants.

In 1997, the Japanese government decided to start a new program beginning in fiscal year 1998 to tackle all of the rice genome sequence as an indispensable resource for advanced genomics such as gene discovery or gene expression profiling with an ultimate goal of improving rice quality by using a novel breeding strategy. This announcement stimulated other countries that had a strong interest in genome sequencing, not only of rice but also of other cereal crops in view of the significant synteny shared by these closely related species. Compared with that of rice (430 Mb), the genomic sizes of other important cereal crops are huge: 3,200 Mb for maize and 17,200 Mb for wheat, which make them less attractive as targets for whole-genome sequencing. The utility of the syntenic relationship among cereals has been proven in the discovery of common genes such as the dwarf phenotype in maize (*D8*) and wheat (*Rht1*) based on genomic information derived from rice (Peng et al 1999). During the 4th International Symposium of Plant Molecular Biology in 1997 in Singapore, a workshop on rice genome sequencing was held to organize the foundation for an international collaboration, the International Rice Genome Sequencing Project (IRGSP), aimed at accelerating completion of the sequencing of the entire rice genome (http://rgp.dna.affrc.go.jp/Seqcollab.html). In February 1998, the first working group meeting was held in Tsukuba with five countries establishing basic guidelines governing the sequencing collaboration such as sequence accuracy and policies for sequence release. Later in this chapter, sequencing progress at IRGSP will be described.

The sequencing strategy at RGP

Japonica rice variety Nipponbare was chosen as a template for genome sequencing in the Rice Genome Research Program (RGP) because it has been used as one of the parents of the F_2 population used for constructing the linkage map, the DNA resource of the YAC physical map, and as a source of mRNAs used to construct cDNA libraries to obtain ESTs. This variety is also used as a common resource in the IRGSP. In

the RGP, a PAC (P1-derived artificial chromosome) library was used as a vector to establish a genomic library for sequencing. Our library is composed of about 70,000 clones of *Sau*3A1 fragments of DNA with an average insert size of 112 kb (Baba et al 2000). Another type of Nipponbare library was constructed using *Mbo*I as a restriction enzyme that has the same specificity as *Sau*3A1 but different sensitivity for methylated cytosine in the target sequence, GATC. In the rice genome, it is known that guanine and cytosine are rich in expressed genomic regions that might influence the distribution of fragment sizes and sequence characteristics obtained by both restriction enzymes. Both libraries should be complementary to make a perfectly reconstructed rice genome using DNA fragments.

About half of the clones in the *Sau*3A1 library are now being used to construct a sequence-ready PAC physical map by polymerase chain reaction (PCR) screening. The PCR primers are designed from the sequences of sequence tagged sites (STS)/EST markers aligned on the linkage map or YAC physical map. The STS markers are derived from 1,350 restriction fragment length polymorphism (RFLP) markers, of which 1,250 originated from rice cDNAs. On the other hand, locating cDNAs on the YAC physical map using a three-dimensional PCR screening method generates the EST markers (Wu et al 1999). Positive YACs obtained by each EST are aligned consistently along the linkage map using the computer software SEGMAP with the help of RFLP markers. In the RGP, 5,200 EST markers have been generated on the rice genome and these are used for ordering PAC clones. Because the positions of EST markers are not defined genetically, the ordering of markers is sometimes ambiguous. A fingerprinting strategy of ordered PACs with *Hin*dIII, *Eco*RV, and *Bgl*II is used to overcome this limitation and to estimate the degree of overlapping among the positive PAC clones extracted by EST markers.

Screening with STS/EST markers is invalid for identifying PACs located in genomic regions where no such markers are available. There are two main methods for filling these gaps: one is end-walking and the other is the STC (sequence tagged connector) method. End-walking uses the end-specific sequence derived from the clone flanking the gap as a PCR primer and hence requires the use of PCR against pooled PAC DNAs in several steps. This method is essential for efficient and specific screening of correctly adjacent PACs. STC, on the other hand, requires a data set of partial sequences of both ends of each PAC or BAC clone within a library. After the core PAC is sequenced, its overlapped sequences are computationally searched within STCs and the degree of overlap is estimated by fingerprinting. This method requires a database of STCs for a target library and, so far, a total of 105,000 STCs derived from two types of Nipponbare BAC libraries (*Hin*dIII and *Eco*RI fragments) constructed by Rod Wing of Clemson University and 55,000 STCs from RGP's *Mbo*I BAC library are available. These numbers indicate a ratio of one STC in every 2.7 kb of the rice genome. Considering the average size of BACs (130 kb) in these libraries, 40 points are expected within each BAC and this density seems sufficient for locating adjacent BACs. The overlap of selected BACs must be examined by fingerprinting because a similar sequence, especially a repeat sequence or transposons, might interfere with the specific identification of correctly positioned BACs among picked BACs.

Selected PAC/BAC clones are cultured and their DNAs are purified by a Qiagen Largeconstruct Kit™ to eliminate DNA from the host bacteria and then they are sheared by ultrasonic treatment. The fragmented DNAs are blunt-ended by T4 DNA polymerase and electrophoresed on an agarose gel for size fractionation to obtain about 2-kb and 5-kb fragments. The eluted DNA fragments are ligated into plasmid vector pUC18 and then transformed using *E. coli* DH5α to construct 2-kb and 5-kb libraries. Usually, about 1,000 clones from both libraries are used as templates for sequencing by the dye-terminator Sanger method from both ends to generate 4,000 sequences with about 500 bp. Using an ABI3700 automated capillary sequencer, it takes one week to collect this amount of data for one PAC/BAC. These data are analyzed using a computer (UNIX) software, Phred/Phrap, to check the sequence quality and then assembled in a correctly oriented manner from one end to the other. If gaps exist, trials to fill them must be performed first by finding a clone spanning the gap and then sequencing it. Another problem is how to raise the sequence quality when the first assemblage fails to give the threshold quality value. These genomic regions often comprise GC-rich, AT-repeat, or inverted repeat sequences. The sequencing technology now available does not necessarily solve all of these problems and, in some cases, unsequenced regions remain. The final finishing step requires consistent size distribution of *Eco*RI and *Hin*dIII fragments with expected fragment sizes from obtained nucleotide sequences.

The finished sequence is then computationally analyzed to extract meaningful information from the sequence. This procedure is called annotation, in which mainly information on the existence of expressed genes is obtained and the expected gene names (identification) or functions are annotated after searching for similarities with nucleotide or amino acid sequence databases. Similarity search tools such as BLASTX and BLASTN are usually used against databases with nonredundant translated amino acid sequences and with rice ESTs, respectively. In addition, a gene-prediction tool, GENSCAN, is used for predicting the plausible initiation and stop codons along the sequenced nucleotide. Although this tool or its derivative is only well fitted to *Arabidopsis* or maize genomic sequences, genes in the rice genome are predictable by combining all of the results obtained by these three tools. The predicted results are both available in the form of tables and graphical view.

To show the annotated results described above with the positional information of the sequenced PAC/BAC, we developed a database named INE, INtegrated Rice Genome Explorer, with a graphical view tool that is designed to function with Java applet (http://dna.affrc.go.jp/giot/INE.html) (Sakata et al 2000). This makes it feasible to track the target PAC/BAC basically by genetic markers along the linkage map. In addition, ESTs mapped on YACs are shown on the sequenced PAC/BAC after confirmation of their existence within the YAC sequence. The idea that sequence information should be shown basically by a map-view aims to facilitate gene identification when phenotypes such as mutants or components of QTL are genetically tagged by DNA markers and mapped on a genetic map. The present version of INE will soon be revised for easier use.

Progress of sequencing in the RGP

We are currently equipped with 14 units of ABI3700 capillary type-sequencers used for massive routine sequencing and 10 ABI377 96-well slab-gel-type sequencers used mainly for gap filling and PAC-end sequencing. Robotic machines are employed for plasmid isolation, sequencing reaction, colony picking, or clone transfer from one titer plate to another. For the first assembly of 4,000 shotgun sequence data, each chromatogram file is assembled with computer software, Phred/Phrap or its equivalent, Tracetuner. The assembled data are then visualized by another computer software, Consed. If sequence gaps are detected after assembly or if the score of each nucleotide is too low to satisfy a reliable threshold value even after assembly, sequences in these regions must be reanalyzed. In case of gaps, first, a shotgun clone bridging the gap must be located using the Phred/Phrap data, and then this clone should be sequenced after preparing subclones by a shotgun or deletion method. Another choice is to use primer walking by designing PCR primers for a Dye-terminator sequencing reaction using the bridging clone as a template. The former method is cheaper and saves more time than the latter.

The RGP is now focusing on chromosome 1 to complete its sequence. The size of this chromosome is estimated at 52 Mb based on its total genetic distance. PCR screening of the 34,560 PAC clones with about 400 STS/EST markers and a computational search of the flanking sequences of the sequenced PAC/BAC among 120,000 BAC-STCs resulted in 52 contigs covering 27 Mb of chromosome 1 as of August 2000. As mentioned above, the strategy adopted in the RGP to make a sequence-ready physical map largely depends on STS/EST markers; therefore, the genomic region without such markers must be filled by a combination of STC search and fingerprints. The disadvantage of the latter method is the length of time required to obtain sequence information on PAC/BAC to be used for the STC search (about three weeks from the beginning of PAC/BAC culturing to the end of the first shotgun sequencing). However, this is the only method that can be used for filling gaps between genomic regions without any markers and much effort must be given to this approach.

By the end of September 2000, 48 PACs on chromosome 1, 12 PACs on chromosome 6, 2 PACs on chromosome 2, 1 PAC on chromosome 3, and 1 PAC on chromosome 8 had been completely sequenced and annotated. Following these clones, 25 PACs and 2 BACs on chromosome 1, 7 PACs on chromosome 6, and 1 PAC on chromosome 2 had also been completely sequenced and are under annotation. The total sequenced length is 13.0 Mb including the overlapped regions within a contig. These sequences have been registered in the DNA Databank of Japan (DDBJ) and the information is freely available on the Web.

We observed several characteristics of the completed and annotated sequences. The average predicted gene density is one gene in every 5 kb. This ratio suggests the presence of about 80,000 genes within the rice genome, assuming their even distribution. The total number of expressed rice genes was previously estimated at about 20,000–30,000 based on the matching frequency of genomic and expressed genes so far cloned. This large discrepancy in the estimated total gene number could be due to

an overestimation of the gene number and uneven gene distribution within the rice genome. The former seems to be the case because, as in *Arabidopsis*, a total of 30,000 genes is predicted from the whole-genome sequence. We cannot offer any reasonable explanation regarding the large discrepancy in estimating the total gene number considering that rice is a monocot, whereas *Arabidopsis* is a dicot. The most plausible reason for this discrepancy may be the uneven gene distribution because the genomic regions sequenced so far have been selected using PCR primers designed by nucleotide sequences from expressed genes. In addition, the restriction enzyme *Sau*3A1 used for constructing the PAC library is sensitive to methylated cytosine, and intergenic or heterochromatin regions may be resistant to cutting into small fragments with this enzyme. This speculation will be proven by using the BAC library constructed using *Mbo*I, which has the same specificity as *Sau*3A1 but is insensitive to methylation. This *Mbo*I BAC library of Nipponbare has been used so far to generate 50,000 STCs, which are used to screen clones to fill gaps between PACs picked by ESTs.

Information on repeated sequences is also available from sequenced PACs. Many types of repeats in the rice genome can be evaluated with the software Miropeats. If the most abundant two-base repeat is TA, then the exact number of repeats cannot be confirmed experimentally. Also, a one-base repeat of G or C and a two-base repeat of GC may be present with short lengths. These types of repeats can seriously interfere with sequencing using any chemistry of the dideoxy nucleotide method.

Transposable elements are annotated and are shown in INE. Several types of elements, such as *Gypsy*-like, *Copia*-like, and non-LTR types, are found. The frequency of existence of these transposable elements is one element in every 100 kb among the PACs we have sequenced so far. This frequency is unexpectedly low and for the same reason as to why gene density is higher within these PACs. This situation is very different from that observed in the maize genome, which is rich with transposable elements around the *Adh1* gene (San Miguel et al 1998). The difference in the genome size of both species may reflect the contents of such elements between genes.

A similarity search by BLASTX of the NCBI nonredundant protein database frequently matches a hypothetical or putative *Arabidopsis* gene as the most similar counterpart of the predicted rice gene. Although this reflects the progress of genome sequencing of *Arabidopsis* and huge amounts of genome sequence data from other plants, no evidence of clear similarity has been shown in the gene order between sequences of rice chromosome 1 and almost all of the five chromosomes of *Arabidopsis*. A similarity search with BLASTN of rice ESTs from rice variety Nipponbare (the same variety used for genome sequencing) revealed that more than one-third of the predicted genes matched ESTs with a significant score. In many cases, the same predicted gene redundantly hit ESTs.

IRGSP: International Rice Genome Sequencing Project

Rice has a genome size of 430Mb, the second largest species so far targeted for genome sequencing, and its importance as a staple food for about half of the world population is well recognized. This motivated researchers having a strong interest in

rice genome sequencing to organize an international collaborative project to decode the sequence in an accelerated manner encompassing strict sequencing standards common to all participating members (Sasaki and Burr 2000). At the beginning of IRGSP, five countries, Japan, the United States, the United Kingdom, China, and Korea, joined together to discuss and agree upon guidelines on sequence standards, chromosome sharing, and so on. An in-depth description of the rules governing the collaboration is available on the Web at http://rgp.dna.affrc.go.jo/Seqcollab.html. The main points of these guidelines are summarized as follows: (1) japonica rice variety Nipponbare is sequenced as a common DNA resource, (2) the sequencing strategy is clone-by-clone using a PAC/BAC library shared among IRGSP members, (3) accuracy of the sequence must be 99.99%, (4) each member has a mandate to sequence at least 1 Mb per year, and (5) sequence data will be released immediately to the public after the completion of sequencing. As of September 2000, a total of 102 PACs/BACs had been completely sequenced in the IRGSP, covering about 15 Mb of the rice genome. An additional 100 PACs/BACs covering another 15 Mb have also been sequenced although several gap regions remain to be filled.

Thus far, at most 10% of the rice genome has been sequenced. How can we accelerate complete sequencing under limited financial support? Sequencing can be enhanced with funding support from sources other than IRGSP. In April 2000, a private company, Pharmacia/Monsanto, announced the delivery of its rice genomic sequences, BAC clones, and other related data to IRGSP. The data have been generated for a project to produce a rough draft sequence of the rice genome using rice variety Nipponbare. Two key points are expected to complement the IRGSP sequencing strategy: one is to construct a sequence-ready BAC/PAC physical map; the other is to supply shotgun sequence data for each target BAC. An agreement to combine both sets of data from IRGSP and Pharmacia/Monsanto is under negotiation. Although key issues surrounding the merging of data remain to be resolved, a blueprint of the rough draft sequence of the rice genome should enhance early completion of the sequencing project.

References

Baba T, Katagiri S, Tanoue H, Tanaka R, Chiden Y, Saji S, Hamada M, Nakashima M, Okamoto M, Hayashi M, Yoshiki S, Karasawa W, Honda M, Ichikawa Y, Arita K, Ikeno M, Ohta T, Umehara Y, Matsumoto T, de Jong PJ, Sasaki T. 2000. Construction and characterization of rice genomic libraries: PAC library of japonica variety Nipponbare and BAC library of indica variety Kasalath. Bull. NIAR 14:24-36.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP. 1999. 'Green Revolution' genes encode mutant gibberellin response modulators. Nature 400:256-261.

Saji S, Umehara Y, Antonio BA, Yamane H, Tanoue H, Baba T, Aoki H, Ishige N, Wu J, Koike K, Matsumoto T, Sasaki T. 2001. A physical map with yeast artificial chromosome (YAC) clones covering 63 percent of the 12 rice chromosomes. Genome 44:32-37.

Sakata K, Antonio BA, Mukai Y, Nagasaki H, Sakai Y, Makino K, Sasaki T. 2000. INE: a rice genome database with an integrated map view. Nucleic Acids Res. 28:97-102.

San Miguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. Nat. Genet. 20:43-45.

Sasaki T, Burr B. 2000. International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr. Opin. Plant Biol. 3:138-141.

Wu J, Shimokawa T, Maehara T, Yazaki J, Harada C, Yamamoto S, Takazaki Y, Fujii F, Ono N, Koike K, Yamamoto K, Matsumoto T, Sasaki T. 1999. Current progress in rice EST mapping. Abstract of Plant & Animal Genome VII. p 163 (available at http://www.intl-pag.org/pag/7/abstracts/pag7264.html).

Yamamoto K, Sasaki T. 1997. Large-scale EST sequencing in rice. Plant Mol. Biol. 35:135-144.

Notes

Authors' address: Rice Genome Research Program (RGP), NIAR/STAFF, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan, E-mail: tsasaki@abr.affrc.go.jp.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Strategies and techniques for finishing genomic sequence

M. de la Bastide, D. Johnson, V. Balija, and W.R. McCombie

Finishing genomic sequence is the process of assembling and refining raw sequence data to produce a complete and accurate final sequence. The goals of finishing are to achieve contiguity, delineate insert-vector junctions, and resolve sequence discrepancies and ambiguities such that the error rate of the sequence is less than one in 10,000 bases. Such a level of accuracy is often difficult to establish due to a repetitive sequence or other secondary structures that require the use of more complicated finishing techniques. Because of such complexities, the finishing process is the bottleneck in the sequencing queue, and as such it is extremely beneficial to make the draft data publicly available before the sequence is completed. Much of the valuable information contained in the sequence can be determined from data at this earlier stage.

Finishing genomic sequence is the process of assembling and refining raw sequence data to produce a complete and accurate final sequence. Finishing is a combination of automated and manual editing with directed sequencing, followed by assembly verification. The goals of sequencing are to achieve contiguity, delineate insert-vector junctions, and resolve sequence discrepancies and ambiguities such that the error rate of the sequence is less than one in 10,000 bases. Such a level of accuracy is often difficult to establish due to repetitive sequences or other secondary structures. Because of such complexities, the finishing process is the bottleneck in the sequencing queue, and as such it is extremely beneficial to make the draft data publicly available before the sequence is completed. Much of the valuable information contained in the sequence can be determined from data at this earlier stage. However, finishing to a high accuracy is an extremely valuable endeavor. Current assembly algorithms (such as PHRAP, [Green 1996]) alone are not always sufficient to correctly assemble the sequence for gene prediction analysis, and the gene/pseudo-gene difference can be as simple as a single base change. Finished sequence provides the most complete information for scientific study.

Background

In the random shotgun phase of production, sequencing tasks are much more routine and repetitive, making them amenable to a division of labor that maximizes throughput. In comparison, finishing is a more directed and specific process, requiring a higher level of experience and a more focused effort. Therefore, to finish the sequence in a timely and cost-efficient manner, it is imperative to effectively utilize the production-sequencing group. In general, sequencing in the random phase to a level of approximately 8-fold (assembled, PHRED 20 [Ewing et al 1998, Ewing and Green 1998]) coverage greatly reduces the time required to manually close sequence gaps by oligo walking or other directed techniques. Furthermore, the use of both single- and double-stranded subcloning vectors in the shotgun phase is useful to help eliminate the occurrence of vector-specific cloning biases that can result in sequence gaps. Similarly, sequencing with a mixture of both dye primer and dye terminator chemistries can be a complementary strategy due to the difference in effectiveness of the two chemistries on specific sequence problems. For example, dye terminators often resolve the guanine-cytosine compressions commonly found in dye primer reactions, whereas dye primer reactions are often more processive than dye terminators in dinucleotide repeats. Figure 1 shows a schematic of the problems resolved in the finishing process.

Software

Effective and dependable data analysis software is crucial in the finishing process. CONSED (Gordon et al 1998) is a sequence editing and viewing tool written by David Gordon and used in conjunction with the basecaller PHRED (Ewing et al 1998, Ewing and Green 1998) and the sequence assembler PHRAP (Green 1996). CONSED

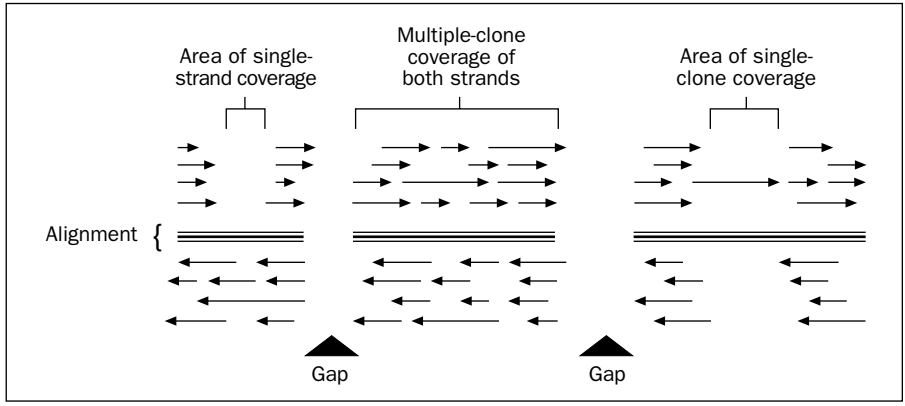


Fig. 1. Schematic overview of a clone entering the finishing phase of genome sequencing. Examples of selected problems to be resolved by the finishing process include single-stranded regions, single-clone regions, insert-vector junctions, and gap filling.

uses the quality values generated by PHRED to navigate through the sequence assembly, stopping on all regions that fall below a set quality value, regions covered by a single subclone, regions covered by a single reaction chemistry, regions of unaligned sequence, or sequences with high-quality bases in disagreement. Navigation by quality greatly increases the efficiency of the editing sequence by virtually eliminating manual intervention in areas where most of the sequence data are of good quality, but a small number of poor-quality sequences cause discrepancies. CONSED also allows the finisher to override quality values, make editing changes, add and remove data, and break and rejoin contigs to alter the assembly, thus giving the finisher the ultimate authority to correct the sequence. A second advantage of the CONSED package is the Autofinish (Gordon et al 1998) feature. Autofinish automatically scans the assembly and picks universal dye terminator, custom oligo dye terminator, and reverse dye terminator reactions to fix low-quality areas in the assembly, to provide coverage, and to attempt to close gaps. Automated reaction calling decreases the time required for a finisher to manually choose directed reactions to solve problem areas. Autofinish has a variety of parameters that may be modified to the specifications of the sequencing center. The number and type of reactions called by the Autofinish program can be tailored to the workflow, cost restrictions, and target accuracy of the project. For instance, if custom oligos are extraordinarily expensive to order, the Autofinish parameters can be modified to choose more universal primer reactions. Autofinish can also be run multiple times to call additional reactions after data from the first round are entered into the assembly. After the third time Autofinish is run, PCR primers are automatically picked to close remaining gaps. It is important to note, however, that, to maximize the efficacy of the reactions called by Autofinish, a finisher needs to manually break and reassemble areas within the database that have been incorrectly aligned by the assembler before running the program. More information on the PHRED, PHRAP, and CONSED programs (including the Autofinish feature) is available at www.genome.washington.edu/consed/consed.html. Figure 2 shows a screenshot of the CONSED software.

Finishing strategy and techniques

Basic finishing strategy

It is important to streamline the finishing process as much as possible to increase efficiency. Although each clone will require a slightly different set of techniques to achieve contiguity and the requisite accuracy level, most clones will require a common set of reactions to solve typical sequence problems. For instance, if all of the dye terminator reactions needed for each project are entered into a queue and performed simultaneously, the turnaround time for reactions is greatly reduced. Organization of labor is also essential for the establishment of the finishing pipeline. Centralizing all of the first-pass assembly procedures in one group, and all of the final quality assurance protocols in another, for example, can increase overall efficiency. Employing personnel of different experience levels can increase cost-effectiveness. Senior personnel are more suited for the assembly manipulation and verification processes, along

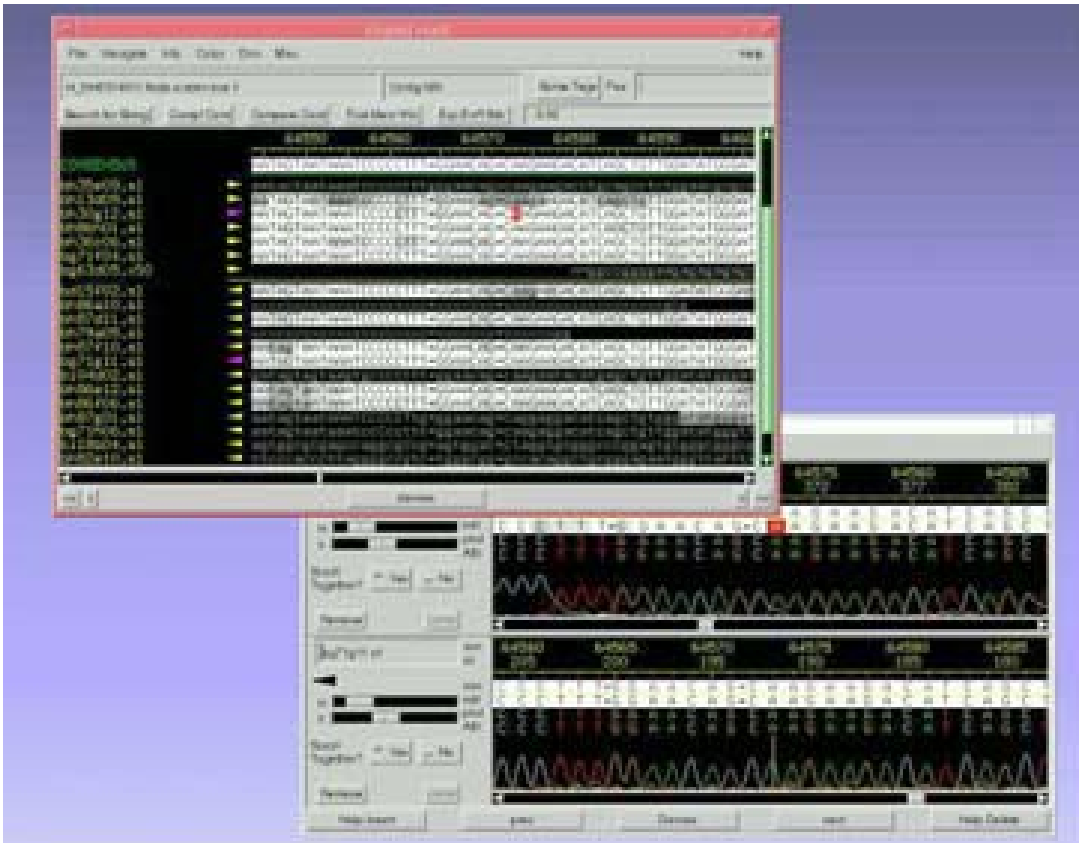


Fig. 2. Screenshot of the CONSED (Gordon et al 1998) sequence editing software. CONSED allows the finisher to view and manually edit the sequence alignments and the basecalling of the primary sequence data.

with the advanced techniques required for difficult regions of sequence, whereas less experienced technicians can handle the more standard sequence editing, templating, reactions, and loading of finishing samples. Putting a difficult clone into the hands of an appropriately experienced finisher will save a great deal of time, money, and frustration.

Steps in the finishing process

1. Finishing begins with clone confirmation and setting the finishing boundaries. It is necessary to perform searches to determine the map location of your clone, the neighboring clones (if any), and the extent of the overlaps with the neighboring clones. This information should be recorded to minimize redundancy in finishing. Searches should also be done to screen for vector or *E. coli* contamination, or contamination with sequence from other projects. Search engines available on the Web will find and display pairwise sequence alignments for areas of sequence that match other sequences in the database. The BLAST server at the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov) has an extensive database of sequences with which to compare your sequence. The Wisconsin package (Genetics Computer Group) offers the program Wordsearch, which identifies areas of similarity between your query sequence and either database sequences (i.e., vector sequences, another bacterial artificial chromosome, BAC, clone, etc.) or any user-defined sequence. The output is similar to that of a BLAST search. The RepeatMasker program masks simple repeats in sequence data, which can otherwise confuse and complicate the search results. For example, ALUs are short, high-copy repeats of about 300 bp found in the human sequence that must be masked in order to perform a meaningful search.
2. The next step is to complete the Autofinish reactions. CONSED's Autofinish reactions are automatically selected to extend the ends of contigs, build coverage, and resolve ambiguities. Three types of reads are chosen: custom primer reactions, dye terminators with universal primers, and reverse dye terminators. Terminators resolve ambiguities including compressions. Reverse reads and custom primers attempt to order or join contigs and generate coverage in single-clone regions.
3. Open the CONSED database to examine the sequence contigs. Check the quality of the ends of the contigs. Trim any poor-quality sequence and extend any sequences of good quality that have been prematurely trimmed by PHRED (barring clones that clearly run into the subcloning vector). This may uncover the BAC vector that was trimmed by the automatic vector cutoff program. Create consensus files and BLAST-search the ends of the contigs for the insert/vector junction.
4. Using the CONSED software, create a consensus (FASTA) file of the major contigs and use Miropeats to check for possible joins and to gauge the repeat structure of the clone. Miropeats (Parsons, unpublished) is a program that graphically displays regions of DNA that are repeated somewhere else on the same or

a different sequence or contig. Figure 3 shows the graphical output generated by the Miropeats program. You can search with repeat tags or use PHRAPVIEW (part of the CONSED package [Gordon et al 1998]) to get a closer sequence view of any possible joins suggested by Miropeats, including those that may have been blocked in the initial assembly by poor-quality sequence on the ends of contigs.

5. Use PHRAPVIEW to check forward-reverse pairs within the assembly. Any read pairs bridging a gap will help to order the contigs, and those within the contigs monitor the integrity of the assembly (i.e., repeats).
6. Using the primer picking function in CONSED, pick two custom primers to "walk" on subclones (or to use for polymerase chain reaction, PCR, if no appropriate clones are available) at the ends of each of the contigs. Custom primers that are to be used for PCR or sequencing must be unique to avoid random priming events. Primer walking will extend the sequence on the ends of the remaining contigs and facilitate joins. Remember to check for unique primers by searching the primer sequence against the rest of the assembly, to choose primers in the proper direction, and to choose the priming site approximately 100 bases from the end of the contig, allowing a buffer zone for a poor-quality beginning sequence.
7. Begin the first-pass edit. Generally, the first pass is the most time-consuming because of discrepancies that arise from the presence of one or a small number of poor-quality sequences in an otherwise high-quality area. These areas some-

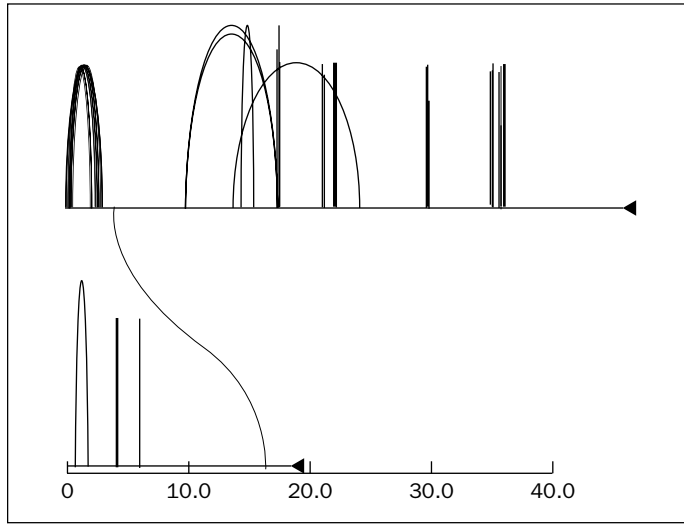


Fig. 3. Graphical output of the Miropeats program (Parsons, unpublished). Miropeats displays regions of highly similar sequence both within and among contigs by connecting the regions with lines indicating the length and orientation of the sequence match.

times contain sufficient data to make editing changes without requiring additional reactions. Always check the primary chromatogram data as well as the PHRED quality values of the available sequence before making an edit. The first pass is also the first attempt to resolve sequence ambiguities that do require additional data (i.e., compressions, low-quality areas, and single-clone areas). In problem areas that require more sequencing data, it is important to employ an aggressive finishing strategy to maximize efficiency. An aggressive approach is one in which multiple templates (2–5) are called to resolve a given problem so that, if one reaction fails, another will be available to take its place. Additionally, in an aggressive approach, custom primers would be ordered to cover single-clone areas, rather than trying to extend reads that may not be long enough to cover the area. Multiple custom oligos should be ordered to extend contigs or cover problem areas in case the first oligo does not work. Remember to tag any problem areas (i.e., compressions, repeats, clone variations, single base runs in which the number of bases varies, etc.) so that you can easily return to the area for study. Do not remove tags even if the problem is resolved! During the first-pass edit, it is a good idea to intermittently perform some of the sequencing reactions generated by editing to allow extension reactions to be entered into the database early in the finishing process. Early access to this data may expedite joins, reduce turnaround time of ordering additional primers for PCR or subclone walking if the first set fails, or elucidate BAC insert/vector junctions or read pairs that will help to order contigs.

8. After the first-pass editing is complete and all of the first-pass reactions have been assembled into the project, a second editing round begins. The second editing pass is more aggressive than the first round in several ways. Templates may be regrown to fix problem areas, custom oligos that may be used for PCR or subclone walking are ordered for any regions still in need of coverage, and PCR is done for any remaining gaps including vector junctions. This finishing round should resolve the majority of discrepancies and leave only a small number of problem areas for the third (and, we hope, final) edit. In the third-pass edit, more complicated techniques such as transposon insertion or small insert libraries may be employed to resolve any remaining gaps, repeats, or other secondary structures (see the next section, “Techniques for problem areas”).
9. The final check consists of checking the entire clone for contiguity, consensus quality (PHRED scores), coverage, vector junctions, and the proper resolution of repeats. Tagged regions should be checked to ensure that the problem has been sufficiently corrected. Single-stranded, single chemistry regions must have a minimum quality of PHRED30 at each base. The clone should be rechecked for single-clone regions and any areas lacking coverage. Gaps must be filled with a sufficient amount of good-quality data and coverage. Regions covered only by PCR must be amplified with a high-fidelity enzyme and must be tagged for annotation. Transposon sequences must be excised and the region must be tagged for annotation. Vector junctions must contain a restriction site and must be tagged as a BAC cloning vector. Any other region need-

ing annotation must be tagged according to the proposed finishing standards of the Bermuda genome meetings and the International Rice Genome Sequencing Project (Bentley 1996, Guyer 1998, Finishing standards of the IRGSP at http://demeter.bio.bnl.gov/Guidelines.html).

10. Restriction digests must be performed on every clone prior to final submission and the data from the digest must be compared to the bands predicted by the *in silico* restriction pattern of both the clone and the cloning vector. The Wisconsin Package (GCG) contains various programs for DNA sequence analysis, including the Mapsort feature, which shows the location of predicted restriction sites in your DNA sequence as well as the predicted sizes of the bands following digest with a chosen enzyme. Restriction enzymes will have sites in the cloning vector in which the insert is contained and will therefore add bands to the resulting fingerprint. The restriction data must agree with the predicted data within experimental error (~1 kb in a 10-kb fragment, for example) to ensure that the assembly is correct and that the repeats are correctly sorted. Disagreements between the experimental and predicted fingerprints may indicate a misassembly and must be noted and evaluated.
11. Complete the finisher's clone submission checklist (see Appendix) to ensure that the clone meets the proper standards (see "Finishing standards of the IRGSP" at http://demeter.bio.bnl.gov/Guidelines.html) and attach any other information that pertains to the finalization of the clone. Submit the clone for final review. The final review should be carried out by an experienced finisher, so that the base changes and assembly manipulation can be checked and confirmed.

Techniques for problem areas

Difficult sequence regions such as repeats and secondary structures often pose problems in finishing genomic sequence. A variety of techniques have proven helpful in the resolution of regions of sequence that are refractory to standard sequencing techniques.

Physical gaps

Physical gaps are regions in which no template is available that spans the gap. Physical gaps may result from little or no representation of an area in the subclone libraries (cloning bias), or the presence of repeats and secondary structures can often cause physical gaps. The standard solution for gaps between contigs is PCR. If standard PCR does not sufficiently fill the gap (this is often the case in repeat areas), the following strategies may be of use:

1. High-fidelity enzymes are now available for such amplifications. These enzymes have a lower rate of base misincorporation than the standard Taq enzyme and work over larger amplification size ranges. Good results have been obtained with Klentaq polymerase (Sigma, St. Louis, Missouri), the GC-rich PCR kit (Boehringer Mannheim, Indianapolis, Indiana), and Platinum Taq with the PCR Enhancer System (Gibco BRL Life Technologies, Gaithersburg, Maryland). Use of these enzymes with their modified buffers has allowed specific

amplification of regions that previously gave either no product or multiple bands. (Note: Because many enzymes have some polymerase activity at 4 °C, false annealing and extension may occur while the reaction mixture is prepared on ice. To reduce the occurrence of this problem, use a "hot start" in which the enzyme is the last component added to the reaction mixture in the thermal cycler at 95 °C.)
2. Add DMSO (5–8% maximum) to the PCR reaction. DMSO tends to relax the DNA template, thereby facilitating smooth extension. Higher annealing temperatures (55–65 °C) may also be useful.
3. Pick longer custom primers (30mers) and move the priming sites around, trying multiple combinations from multiple distances.
4. Try PCR and/or direct sequencing off of different template types, including BAC DNA, plasmid subclones, and M13 subclones.

Techniques in addition to PCR include using *in vitro* transposons (see "Techniques for problem areas") and subcloning and subsequently sequencing restriction fragments that are known to span the gap.

Sequence gaps

Sequence gaps differ from physical gaps in that there are templates available that readily cover the gap, but sequencing reactions are hindered and the enzyme does not process through the region. To sequence through such regions, it is often necessary to employ methods other than primer walking, which is both inefficient and sometimes ineffective. These gaps may be the result of compressions in the sequence, internal structures that deter sequence extension, or mono-polynucleotide runs. Possible solutions follow:

1. Compressions—sequence both strands (compressions in a specific area usually reside on only one strand) or use terminator chemistry (the large dye molecule on the 3′ extension end of the sequence tends to "flatten" compressions). BigDye terminator kits (Perkin Elmer, Foster City, Calif.) have proven quite effective at compression resolution. Figure 4 shows a schematic of a typical sequence compression.
2. Hard sequence stops, or regions where the processing of the enzyme is significantly and abruptly hampered, often occur in GC-rich sequences. The dGTP dye terminator kit from Perkin Elmer (Foster City, Calif.) has proven extremely useful in such regions. Most available terminator kits use the inosine analog to help combat GC compressions, but it is thought that the lower efficiency of inosine incorporation coupled with the reduced processing of the enzyme in regions of secondary structure or "hairpins" leads to repressed sequencing, or stops. By replacing inosine with guanine, the dGTP kit enhances processing and allows sequencing to advance through the structure. Compressions may be evident in such regions, but are generally less problematic to resolve. Table 1 shows the commonly used sequencing reagents and their uses.
3. Mono-polynucleotide strings—mono- and dinucleotide repeats also challenge the effectiveness of typical sequencing chemistries. Dye terminators tend to

slip in these regions, producing sequences with varying copies of the repeat. Energy Transfer dye primers (Amersham Pharmacia Biotech, Buckinghamshire, England) have proven useful in these areas, showing higher fidelity and less slippage, thus allowing longer single-base incorporations. Sequitherm XL (Epicentre Technologies, Madison, Wis.) may also be more processive through such A/T or G/C runs (see Table 1). Use subclones with these chemistries, as PCR will slip. Raising the cycling temperature in the sequencing reactions for GC-rich regions and the use of degenerate oligos are techniques that may be useful in some cases as well. Thermolabile sequenase used in noncycling sequencing may also inhibit slippage.

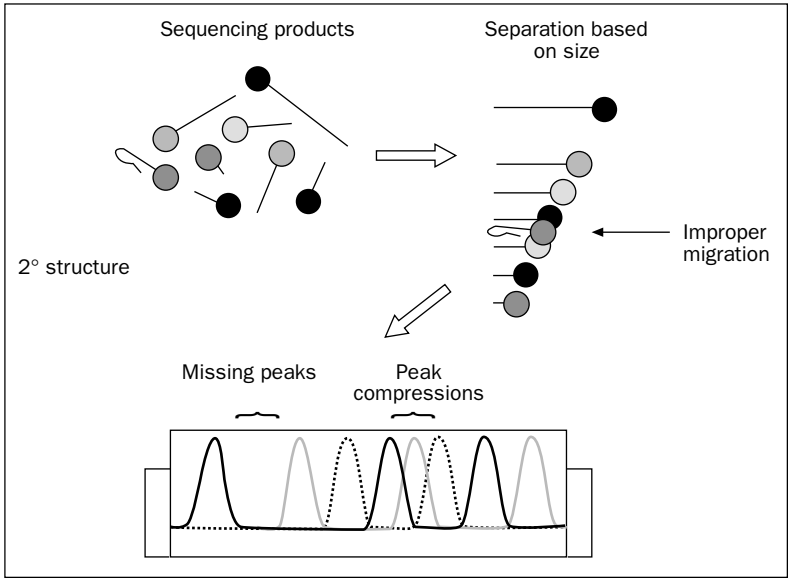


Fig. 4. Schematic of a typical sequence compression. The altered base spacing seen in sequence chromatograms results from the altered migration of extension products containing secondary bonding structures.

Table 1. Widely used sequencing chemistries, including their most useful applications.

Chemistry	Supplier	Advantages	Disadvantages
BigDye terminators	ABI	Even peak heights No compressions	Frequent sequence stops
dGTP dye terminators	ABI	Processive through G-C rich regions	Compressions Expensive
Sequitherm XL	Epicenter	Processive through homopolymeric runs	Expensive 4-tube reactions Must use labeled primer
ET dye primers	Molecular Dynamics	Very processive Even peak heights	Compressions 4-tube reactions Must use labeled primer

4. Structure problems—subcloning is a viable option for areas in which a PCR product is readily reproducible, but sequence quality is poor or stunted (cloning individual PCR molecules and using a universal primer usually produce a better sequence). (Note: Because of base misincorporation, high-fidelity enzymes must be used and at least three different concurring subclones on both strands are necessary for coverage.) Invitrogen's TA subcloning kit (Carlsbad, Calif.) has been particularly helpful in obtaining a clean sequence from PCR products that were refractory to direct sequencing with nested primers.

Small insert libraries (SILs, McMurray et al 1998) are an alternate technique for disrupting the secondary structures of repeats. In this method, clones spanning the problem region are sheared into 200–500-bp segments that are subsequently subcloned and sequenced. These small pieces of DNA are generally not subject to the secondary folding, which deters sequencing in larger inserts (see Fig. 5).

In vitro transposons may also be useful. The Template Generation System (Finnzymes, Watertown, Mass.) uses the *Mu* transposon to randomly insert novel universal priming sites into segments of DNA. Inserting multiple priming sites within the problem area tends to break up the secondary structure and allow sequencing to occur from within the problem region. Transposition has been a successful

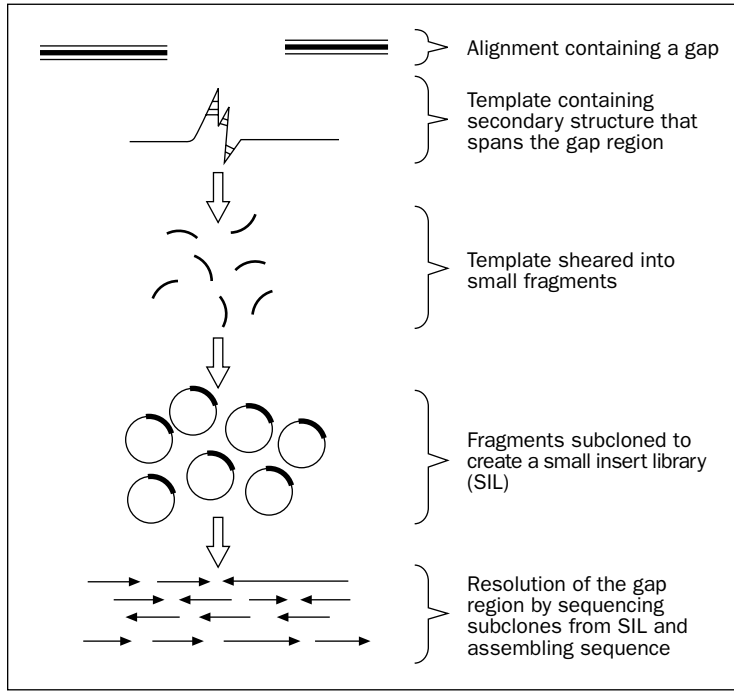


Fig. 5. Schematic of the small insert library (SIL) sequencing technique (McMurray et al 1998). Shearing a template covering a difficult sequence region into 200–500-bp segments may disrupt secondary structures that often deter sequencing.

sequencing method for us on cDNA clones for which oligo walking was both costly and time-consuming because of the turnover time of ordering oligos and primer failures. The linking of universal primer sequences to both sides of the transposon allows sequencing to be done in a "shotgun" fashion on multiple clones, thereby vastly expediting the coverage of the problem region (see Fig. 6).

Repeats

Repeats may cause physical and/or sequence gaps and are often difficult to detect. In the human sequence, ALUs are small, frequent repeats that are, on average, 200–300 bases in length. However, repetitive elements may be quite large in size and may differ by as little as one base in several kilobases of sequence. Careful inspection and interpretation of the sequence data during editing and assembly are imperative lest repeats be misassembled. Restriction digests are often clues to improper assembly and repeat problems. The various types of repeats and some strategies to resolve them follow:

1. Direct repeats—may have a unique sequence between the repetitive elements. Try primer walking, long reads, etc., to walk into the unique area. Physical

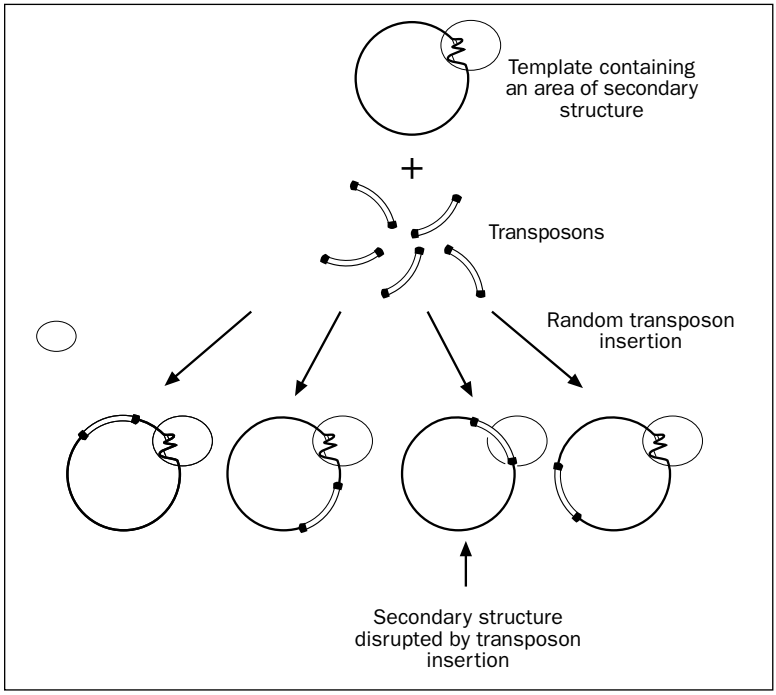


Fig. 6. Schematic of the *in vitro* transposon sequencing strategy. Insertion of a transposon flanked by two universal priming sites into clones containing regions of secondary structure can often help resolve difficult sequence regions by disrupting the secondary bonding.

restriction maps or optical maps are extremely useful in sorting direct (as well as inverted) repeats. Optical maps are anchored restriction digests that provide information not only about the size of the restriction fragments but also about their order along the contiguous clone. Comparing multiple optical maps to predictions can uncover misassemblies and provide clues as to where the fragments actually belong. Subcloning and sequencing restriction fragments may also resolve these repeats (see Fig. 7A).

2. Inverted repeats—often cause gaps because the sequence proceeds in opposite directions. PCR with custom oligos chosen in unique areas away from the gap may provide a template that can be sequenced through to make the join. Sequencing plasmid subclones rather than single-stranded subclones is sometimes more effective (the bonding of the double-stranded templates may inhibit secondary folding). Transposons have also aided finishers with inverted repeats, which are notoriously difficult to sequence through and often cause sequence gaps. Repeats have been effectively resolved by various methods: (1) transposing into a plasmid subclone from the initial shotgun library known to span the repeat, (2) subcloning a PCR product that spanned the repeat and subsequently transposing those clones, (3) small insert libraries (SILs), (4) subcloning and sequencing restriction fragments that span the gap, and (5) PCR, if some unique sequence within the hairpin structure is available. In the case of PCR, it is necessary to design a PCR primer in the unique sequence and use one primer on either side of the gap to generate products representing each side of the inverted repeat (see Fig. 7B).
3. Tandem repeats—these repeats cause enzyme slippage in the sequencing and PCR reactions and it is often difficult to determine the number of copies and the size of the repeat. Sequencing in both directions (with long reads) on subclone inserts of varying sizes may allow you to sequence through the string and into a unique sequence. Noncycling sequenase reactions may help prevent slippage (primers work better than terminators, but be aware that noncycling reactions require ~1 μg of template to generate enough sequence signal). If there are no unique bases within a long tandem repeat, it may be necessary to analyze the size of the repeat band on a restriction digest and attempt to estimate the number of tandem copies (see Fig. 7C).

In summary, these new techniques and chemistries have vastly increased both the efficiency and accuracy of finishing a difficult sequence. The most critical lesson in finishing is to seek the advice of an experienced finisher when a problem is encountered. There is no benefit in repeating techniques that fail. The experience of others often provides an efficient and effective resolution.

Quality assurance of the finished sequence

Once the assembly has been found to meet the requisite quality values and assembly verification, additional quality assurance exercises can be performed for outside verification of the sequence. Reassembly of the database (including the added directed reads) with PHRAP and comparing the computer-generated assembly with the origi-

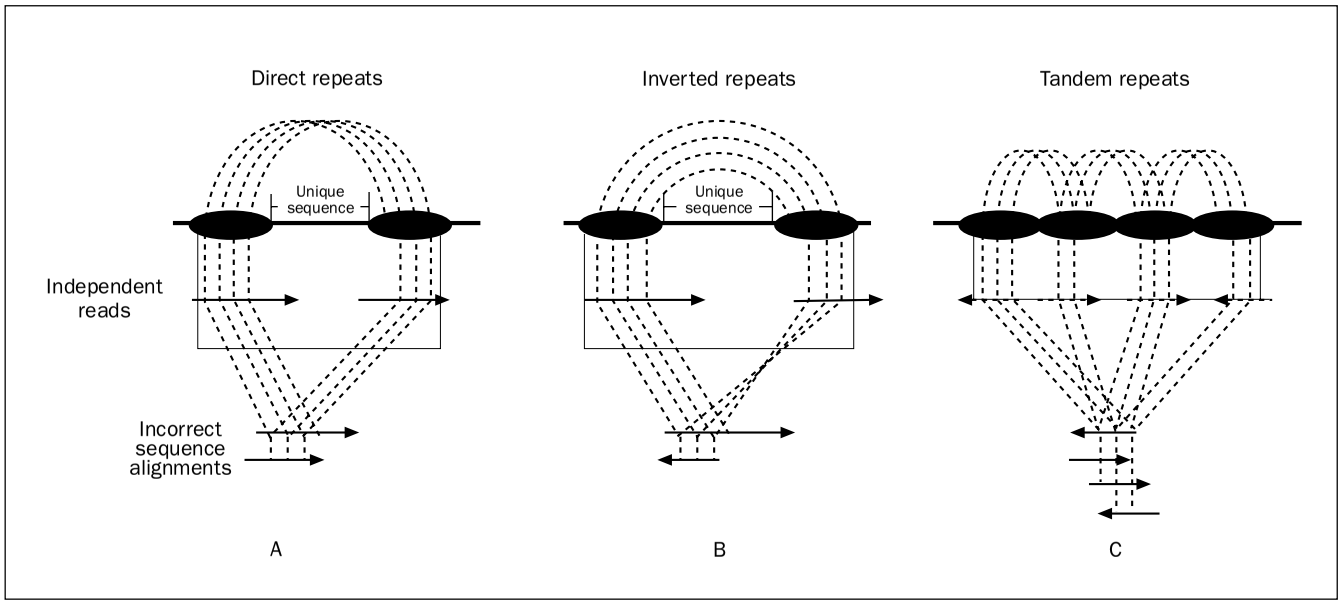


Fig. 7. (A) Direct repeats are regions of highly identical sequence that lie in the same orientation. Direct repeats are often mistakenly aligned on top of each other, thereby eliminating the unique sequence that lies between the repeat copies. (B) Inverted repeats are regions of highly identical sequence that lie in opposite orientations. Inverted repeats often cause physical gaps in the sequence. Secondary bonding between the repeat units results in hairpin loops, which may be both unstable to clone and notoriously difficult to sequence through. (C) Tandem repeats are regions of highly similar sequence with little or no unique sequence separating the repeat copies. Tandem repeats are often incorrectly aligned on top of each other and are difficult to sort. Most polymerases are unable to sequence through multiple copies without slipping.

nal assembly may point to areas that were incorrectly sorted by the finisher. One must always remember not to rely entirely on assembly algorithms, however, as they are not always adept at sorting high-identity repeat sequences. Combined PHRAP quality values can be generated for each base in the assembly based on the quality, strand, and chemistry of the underlying primary data. PHRAP scores of 40 or better indicate an extremely accurate sequence. Overlapping data from clones that share a common portion of sequence can be independently finished, and the resulting data can be compared for differences and discrepancies. Analysis of the assembly using alternate navigation software (such as Gap4 from the Staden package [Staden et al 1998]) and independent resequencing of questionable areas within the assembly can also be informative.

The sequencing pipeline, finishing time, and data release

In most if not all large-scale sequencing operations, an assembly line is constructed. In these labs, clones are processed one group after another from beginning until end. The first step is the making of subclone libraries. This typically takes 1 to 2 weeks per clone. Following subclone library construction is a phase of production sequencing. As an example, in our lab, the production phase typically takes 1 to 2 weeks as well. The next phase is the finishing phase, which involves assembly, resolution of problems, filling of gaps, and verification of the assembled sequence. On an uncomplicated clone, this might take about 3–5 weeks. Some complicated clones might take months or longer to complete. Some of the clones we sequenced from the centromere of *Arabidopsis* chromosome IV took more than 1 year to finish (Mayer et al 1999). Following finishing and validation of the clones, the sequence is analyzed.

Several factors concerning this process need to be emphasized. First, this is an extremely parallel process, that is, many clones, 10 or more, may be processed at about the same time. The clones are not done sequentially. Second, most of the data are available after the production-sequencing phase. The gaps and ambiguities remaining prior to finishing often account for only a small percentage of the total sequence. Although the data are definitely more difficult to use prior to finishing, they are mostly complete, and hence very valuable. Last, as a result of the process at most large sequencing centers, a very large amount of partially finished sequence is available at any given time. In fact, it would not be surprising if the amount of unfinished sequence exceeds the amount of finished sequence available for much of the life of a project, even if the sequence is being finished as rapidly as possible. This is the result of the parallel nature of the process, the bottleneck of finishing, and the typical gradual increase in sequencing capacity. This phenomenon has been known since large-scale sequencing began in the early 1990s.

Aware of this phenomenon, virtually all groups doing large-scale genome sequencing since 1995 have agreed to a policy of immediate release of assembled contiguous fragments of a sequence of more than 2 kilobases. This is the currently accepted international standard for data release referred to as the Bermuda Standards (Bentley 1996, Guyer 1998, HUGO Web site at http://www.gene.ucl.ac.uk/hugo/publications-reports.html; Wellcome Trust Web site at http://www.wellcome.ac.uk/en/1/

biopoldat.html). The 2-kb threshold minimizes the possibility of low-quality or contaminated sequences from being released. As a result, large amounts of sequence are made available to researchers months or years earlier than they would be without preliminary data release. The result is a vast savings of precious resources by saving many researchers from being forced to waste time cloning genes that are sitting on the computers of sequencing centers. Such a data release policy is also very important to the sequencing labs themselves. In the case of the *Arabidopsis* project, virtually all of the groups had clones that were mismapped in the genome. By sharing their preliminary data, researchers were more likely to identify these problems before another group mistakenly sequenced the same clone. A single mistake of this sort, if not prevented by preliminary data release, could easily cost $50,000 or more.

As a result of these factors, almost all, if not all, public large-size sequencing projects have followed a policy of releasing their preliminary data as soon as they are assembled. The rice project is perhaps unique in not following this policy of immediate release of preliminary data.

Resources

For further information, the following finishing resources are available for reference on the Internet:

- Problem Clone Browser—GSC Washington University, St. Louis
  www.genome.wustl.edu/gsc/cgi-bin/ripit.pl
- Cold Spring Harbor finishing manual
  www.nucleus.cshl.org/genseq/Protocol%20Index.html
- University of Washington finishing protocol
  www.genome.washington.edu/UWGC/finish.html
- Sanger Center finishing software
  www.sanger.ac.uk/software/Sequencing/DOCUMENTATION/CAF_SOFTWARE/CAF_PROGS.shtml

References

Bentley D. 1996. Genomic sequence information should be released immediately and freely in the public domain. Science 274:533-534.

Ewing B, Hillier L, Wendl M, Green P. 1998. Basecalling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. 8:175-185.

Ewing B, Green P. 1998. Basecalling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. 8:186-194.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195-202.

Green P. 1996. Documentation for Phrap. http://bozeman.mbt.washington.edu. Genome Center, University of Washington, Seattle.

Guyer M. 1998. Statement on the rapid release of genomic DNA sequence. Genome Res. 8:413.

Mayer K and 228 others. 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana.* Nature 402:769-777.

McMurray AA, Sulston JE, Quail MA. 1998. Short-insert libraries as a method of problem solving in genome sequencing. Genome Res. 8:562-566.
Staden R, Beal KF, Bonfield JK. 1998. The Staden Package. In: Misener S, Krawetz S, editors. Computer methods in molecular biology. Totowa, N.J. (USA): The Humana Press Inc.
Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, WI. Oxford Molecular Group, Inc.

Notes

Authors' addresses: M. de la Bastide, V. Balija, W.R. McCombie, Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA; D. Johnson, Washington University Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Blvd., St. Louis, MO 63108, USA.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Appendix

Finisher's Clone Submission Checklist

Finisher: Clone name:

Date assigned: Date completed:
Estimated clone size: Organism:
Actual clone size: Map location—genetic:

Neighbors/overlaps:

Extent of overlap:

Vector type:

Restriction site:

Is the entire clone finished?

If not, delineate the boundaries:

What finished clone(s) contains the unfinished overlap(s):

Have you checked single-strand/single-chemistry areas at phred30 quality?

Have you checked high-quality disagreements?

Have you eliminated single-clone areas?

Have you checked tags?

Do any areas need annotation? If so, list the locations and reasons:

Does the mapsort agree with the restriction digest? List enzymes used, and corresponding mapsort/digest band sizes, including vector bands! Attach mapsort and digest data.

Other remarks:

Sequence-tagged connector/ DNA fingerprint framework for rice genome sequencing

R.A. Wing, Yeisoo Yu, G. Presting, D. Frisch, T. Wood, Sung-Sick Woo, M. Arief Budiman, Long Mao, Hye Ran Kim, T. Rambo, E. Fang, B. Blackmon, J.L. Goicoechea, S. Higingbottom, M. Sasinowski, J. Tomkins, R.A. Dean, C. Soderlund, W.R. McCombie, R. Martienssen, M. de la Bastide, R. Wilson, and D. Johnson

The CCW Rice Genome Sequencing Consortium is funded to sequence and annotate the short arms of rice chromosomes 10 and 3—approximately 30 Mb of DNA. To efficiently sequence the rice genome, the Clemson University Genomics Institute has developed a framework consisting of two deep-coverage bacterial artificial chromosome (BAC) libraries and BAC fingerprint, genome-anchoring, and sequence-tagged connector databases. These resources have been provided to the International Rice Genome Sequencing Project and are being used extensively. In this chapter, we summarize the framework project and its use to sequence the short arm of chromosome 10.

Rice is one of the most important food crops in the world. Its compact genome, evolutionary relationship with other cereals, and sophisticated molecular genetic tools have made sequencing the rice genome a top priority for plant science. As part of the International Rice Genome Sequencing Project (IRGSP: http://rgp.dna.affrc.go.jp/Seqcollab.html), the CCW Rice Genome Sequencing Consortium (Clemson University Genomics Institute [CUGI], Cold Spring Harbor Laboratory [CSHL], and Washington University Genome Sequencing Center [WU-GSC]) was funded in 1999 to sequence the short arms of rice chromosomes 10 and 3. Together with The Institute for Genomic Research (TIGR) and the Plant Genome Initiative at Rutgers (PGIR—3 Mb of chromosome 10), we are scheduled to complete chromosome 10 by the end of 2001 and chromosome 3 by the end of 2002.

Here we present the development of a bacterial artificial chromosome (BAC)/fingerprint/sequence-tagged connector framework to sequence the rice genome and a summary of CCW's rice genome sequencing project.

The CUGI Rice Genome Framework Project

As a prelude to sequencing the rice genome, CUGI has developed an integrated framework consisting of (1) two deep-coverage large insert BAC libraries, (2) a BAC fingerprint database, (3) a genome-anchoring database, and (4) a sequence-tagged connector database (STC-DB) (Venter et al 1996). Together, these four components are used to select a minimum tiling path of BAC clones across the genome to efficiently sequence rice. Figure 1 illustrates how the framework is used. First, a "seed" BAC is selected that is physically and genetically anchored to the rice genome and shotgun-sequenced. The sequence of the seed BAC is then compared with the STC-DB and a single BAC clone is selected that overlaps minimally with the seed BAC on each end. The fingerprint and genome-anchoring DBs are consulted with each step to ensure that the BAC clones selected are not rearranged and are physically anchored to the contig being sequenced. By selecting several seed BACs across a chromosome, one can, in principle, use this framework to efficiently sequence the rice genome.

Below is a brief description of the four components of the framework project and a preliminary analysis of the data.

Two deep-coverage BAC libraries

Two BAC libraries were constructed from rice (*Oryza sativa* subsp. *japonica*) cultivar Nipponbare with each library representing a different restriction enzyme (Budiman

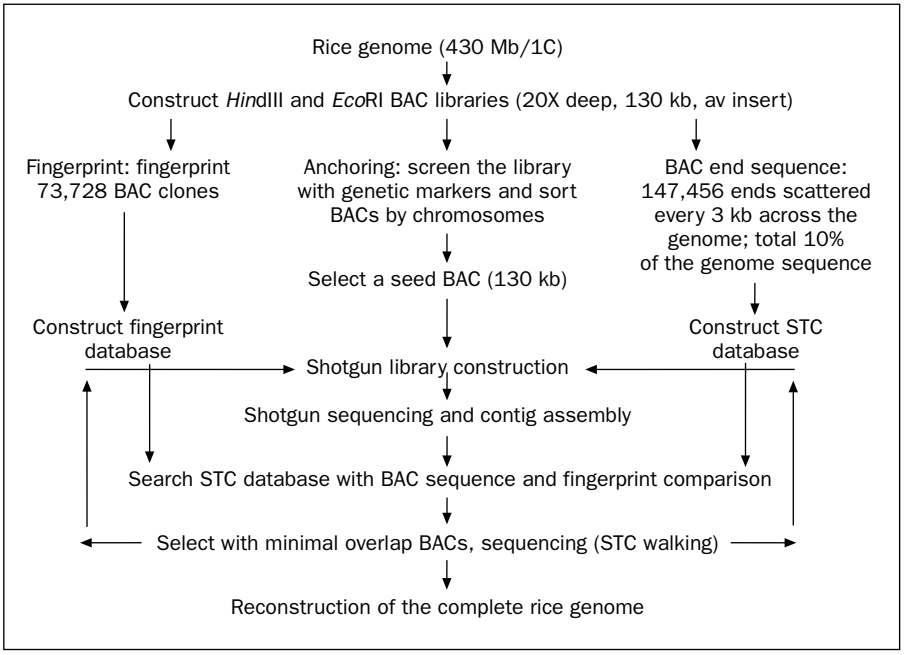


Fig. 1. STC approach to sequence the rice genome.

1999). The development of multiple-enzyme BAC libraries allows for improved genome representation. The rice BAC libraries consist of 36,864 and 55,296 clones for *Hin*dIII (vector: pBeloBAC11) and *Eco*RI (vector: pIndigoBAC536), respectively. The average insert size of the *Hin*dIII library was 129 kb and the *Eco*RI library 121 kb. Approximately 5% of each library contains no inserts and organelle DNA (chloroplast and mitochondria) contamination. Based on a haploid genome size of 431 Mb for rice (Arumuganathan and Earle 1991) and taking into account levels of empty vectors and organelle DNA contamination, the coverage of the *Hin*dIII and *Eco*RI libraries is approximately 10.6 and 15.0 haploid genome equivalents, respectively. Combined coverage for both libraries provides a total of at least 25 haploid genome equivalents.

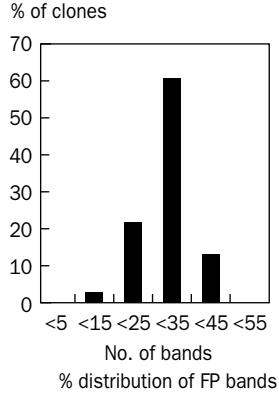
Both BAC libraries (libraries, filters, and clones) are available at cost from the CUGI BAC/EST Resource Center (www.genome.clemson.edu).

BAC fingerprint database

To develop a physical map of the rice genome, we attempted to generate unique restriction fragment fingerprints from 36,864 BAC clones from each library (for methods, see Marra et al 1997). DNA was isolated from 74,496 BAC clones, digested with *Hin*dIII, run on high-resolution agarose gels, stained, and imaged. The *Hin*dIII fingerprint images were then processed through IMAGE (Sulston et al 1989), which tracks the lanes, calls the band, and creates band files. The band files were then input into FPC (Soderlund et al 2000), which assembles BAC clone fingerprints into contigs based on the number of shared bands. Table 1 summarizes our preliminary results in

Table 1. CUGI rice FPC project summary.

Total	Successful	Clones in contig	Singleton	No. of contigs
74,496	64,233 (86.2%)	61,309 (82.3%)	2,924 (3.9%)	1,039



- Total unique bands: 93,532
- Unique bands contig^{-1}: 90
- Average number of BACs contig^{-1}: 59
- Average number of bands clone^{-1}: 28
- Average number of bands singleton clone^{-1}: 13.5
- Average contig size: 378 kb
- Estimated genome coverage: ~20x

the development of a fingerprint DB. Eighty-six percent of the clones were successfully fingerprinted and assembled into 1,039 contigs (FPC parameters: cutoff = 10^{-12}, tolerance = 7) and 2,924 singletons. An average contig contained 59 BAC clones of about 380 kb in size.

Genome-anchoring database

To anchor the 1,039 contigs to the rice genetic map, we used three approaches. The first was to develop probes from mapped restriction fragment length polymorphism (RFLP) markers and hybridize these markers to high-density hybridization filters representing both BAC libraries. Second, overgo primers (http://genome.wustl.edu/gsc/overgo/overgo.html) were developed from STC (as described below) anchored contigs (step 1) to extend the contigs. Third, the STCs were used to tentatively anchor contigs based on sequence homology with sequenced RFLP markers—*in silico* anchoring (Yuan et al 2000). Using these approaches, we have tentatively anchored 472 contigs to the rice genetic map containing 215–234 Mb of DNA or more than 50% of the rice genome.

The complete integrated fingerprint and anchoring databases are publicly available using WebFPC (Soderlund et al, unpublished) at www.genome.clemson.edu.

Sequence-tagged connector database (STC-DB)

To develop STC-DB, we attempted to sequence both ends of every rice DNA insert in the BAC libraries described above. DNA was isolated from 63,729 BAC clones and end-sequenced using a high-throughput DNA sequencing protocol (Yu and Wing, unpublished). Table 2 summarizes the results. An average of 87% of all BAC ends were successfully sequenced, resulting in a total of 41 Mb of high-quality sequence (phred 20 or greater) being released to the dbGSS (database of Genome Survey Sequence) section of Genbank. On average, our results generated about 1 STC every 3.9 kb across the rice genome.

To obtain an early glimpse of the rice genome and to provide rice breeders with a potentially new set of genetic markers, we mined the *Hin*dIII STC data set for transposable element and simple sequence repeat content.

Figure 2 summarizes our analysis of the transposable element content of the rice *Hin*dIII STC-DB. Approximately 13% of the STCs have sequences similar to class I,

Table 2. Publicly available rice sequence-tagged connectors (http://www.genome.clemson.edu). Nipponbare *Hin*dIII and *Eco*RI bacterial artificial chromosome (BAC) libraries, 40.67 Mb of high-quality[a] rice genomic sequence, 9.5% of the total rice genome, 3.9 kb STC^{-1}.

BAC	Total rxn	Successful (%)	Av HQ[a] (bp)
OSJNBa[b]	73,362	63,432 (86.5)	387
OSJNBb[b]	54,097	47,006 (86.9)	343
Total	127,459	110,438 (86.6)	368

[a]No. of bases having phred value ≥20. [b]OSJNBa is *Hin*dIII and OSJNBb is *Eco*RI BAC clones.

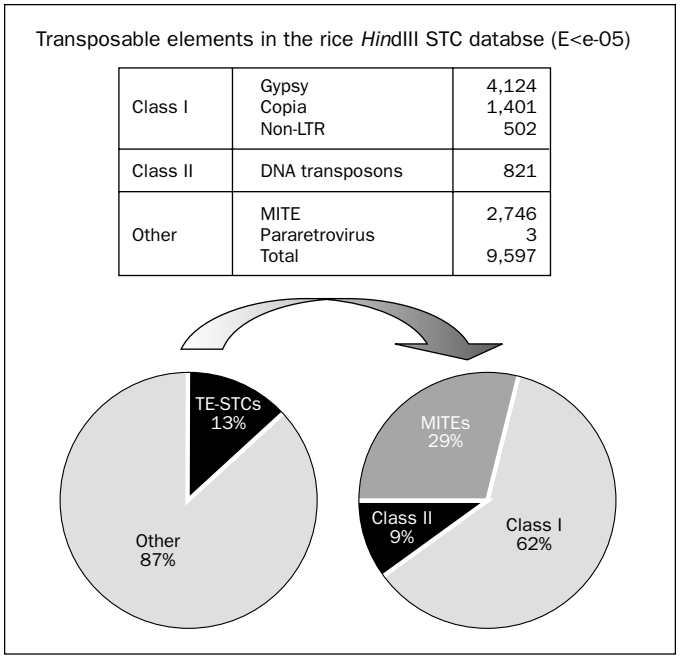
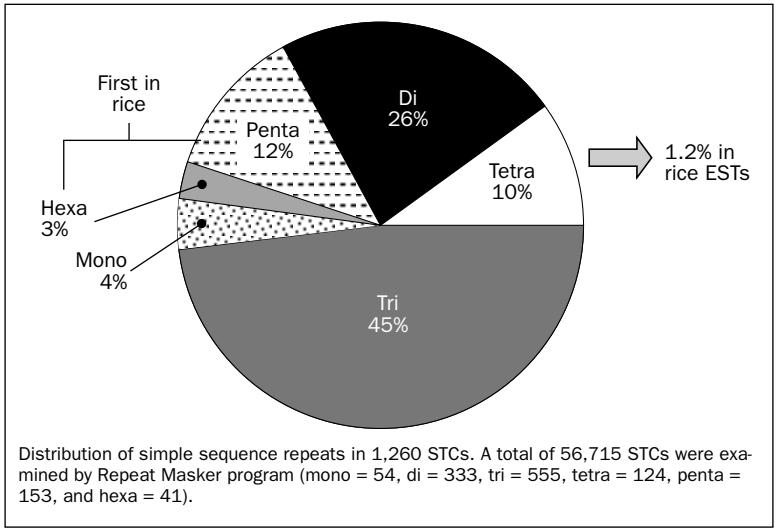


Fig. 2. Transposable elements in the rice sequence-tagged connector database. TE-STCs = transposable elements-sequence-tagged connectors, MITEs = miniature inverted repeat transposable elements.

II, and miniature inverted repeat transposable elements (MITEs). The predominant class, 62%, is similar to the gypsy, copia, and non-LTR class I transposable elements. The next largest class is MITEs (29%). For a more detailed analysis, see Mao et al (2000).

Simple sequence repeats (SSRs) have proven to be extremely useful as highly polymorphic genetic markers. We examined the SSR content of the *Hin*dIII STC-DB and identified 1,260 SSRs (Yu and Wing, unpublished). Tri-nucleotide SSRs accounted for 45% and di- and penta-nucleotide SSRs for 26% and 12%, respectively (Fig. 3). To test whether STC-derived SSRs could be used as genetic markers, primer pairs were designed for ten potential polymorphic SSRs linked to known locations on chromosome 10. All ten primer pairs amplified their expected fragments and seven were polymorphic between Nipponbare (japonica) and Kasalath (indica). Two primer pairs were mapped using 98 backcross inbred lines (BILs) and were found to map to the exact location predicted (Yu and Wing, unpublished). Thus, we conclude from this preliminary study that the STC-derived SSRs will probably serve as a valuable and potent source of new SSRs for rice genetic mapping. Further, since more than half of the fingerprinted contigs are genetically anchored to the rice genome, rice geneticists now have approximately 630 new SSR candidates that could be used for positional cloning, QTL mapping, and breeding.



Distribution of simple sequence repeats in 1,260 STCs. A total of 56,715 STCs were examined by Repeat Masker program (mono = 54, di = 333, tri = 555, tetra = 124, penta = 153, and hexa = 41).

Fig. 3. Frequency and distribution of simple sequence repeats in the *Hin*dIII sequence-tagged connector database. ESTs = expressed sequence tags.

Table 3. CCW chromosome 3 and 10 sequencing timetable.

Activity	Year 1	Year 2	Year 3
	Sep 1999-Aug 2000	Sep 2000-Aug 2001	Sep 2001-Aug 2002
Clone validation	40 BACs	62 BACs	58 BACs
Production sequencing	40 BACs 5.2 Mb draft	62 BACs 8.06 Mb draft	58 BACs 7.54 Mb draft
Finishing/annotation (I)	32 BACs	48 BACs	80 BACs
Finished data release	4.16 Mb	6.24 Mb	10.40 Mb

The CCW Rice Genome Sequencing Consortium

Rice genome sequencing at CCW is divided into three sites. CUGI, led by R. Wing, is responsible for overall management of the project, physical mapping, clone selection/validation, production sequencing, finishing, and annotation. CSHL, led by W.R. McCombie, is responsible for production sequencing, finishing and finishing management, and annotation. Scientists at WU-GSC, led by R. Wilson, are responsible for finishing difficult clones and developing novel strategies and chemistries to finish difficult clones. Table 3 outlines our goals for finishing the short arms of chromosomes 10 and 3 over the 3-year project.

The short arm of chromosome 10—difficult to finish

Using the CUGI Rice Genome Framework and all other available resources (primarily from the Japanese Rice Genome Program and Cornell University: genetic mark-

ers, maps, mapping populations, and a P-1 derived artificial chromosome, PAC, library), the CCW has focused on sequencing the short arm of chromosome 10 (10s). Figure 4 shows the current physical map of the short arm of chromosome 10. We have mapped 31 contigs on 10s covering an estimated 75–85% of this arm. Twelve gaps remain, covering an estimated 2.6–4.0 Mb of DNA.

An early surprise during the physical mapping of 10s was that the centromere was miss-mapped. In collaboration with J. Jiang (Cheng et al 2001), using fluorescence *in situ* hybridization, the chromosome 10 centromere was mapped between 15.2 cM and 15.9 cM. This places it directly in the middle of the ~30 cM region assigned to CCW.

Table 4 shows the current status of the CCW sequencing project. As shown, we have submitted 6.1 Mb of assembled sequence to the HTGS section of Genbank and have finished 1.39 Mb of rice DNA from nine BACs and submitted this to Genbank. All data are also available from the CUGI (www.genome.clemson.edu) and CSHL (http://nucleus.cshl.org/genseq) Web sites and the U.S. Rice Genome Sequencing and IRGSP Web site (http://www.usricegenome.org). As can be seen, CCW did not meet its year-one target of finishing and annotating 32 BACs because of several factors, such as the fact that the 30 cM assigned to CCW is highly heterochromatic and contains numerous repeat regions that are difficult to finish. Further, most sequencing groups have determined that rice is inherently difficult to finish and sequencing results in many GC (guanine cytosine) stops. To demonstrate the repeat nature of some of the BACs we have encountered, Figure 5 shows a Print Repeats/Miropeats (J. Parsons, http://www.ebi.ac.uk/~jparsons/) output of the repeat relationships of four BACs on 10s. Rice BACs OSJNBa0035B06 and OSJNBa0053D03 are located near the centromeric region of 10s, whereas BACs OSJNBa0065H03 and OSJNBa0020E23 are near 1.2 and 3.0 cM, respectively. Here we used a threshold of 500, which means that any repeat 500 bases or longer will be displayed. Figure 5 clearly shows that the repeat nature of the BACs near the centromere is much more complex than near the telomere and will require a significant effort to finish.

CCW has gained an intimate knowledge of and experience with the repeat nature of chromosome 10s and is confident that it will accomplish the finishing goals for the remainder of the project. In addition, at the Interim IRGSP Meeting held in Clemson in September 2000, the IRGSP finishing problems were discussed and a new set of finishing guidelines was established to aid in resolving difficult regions of the rice genome. These new guidelines will take effect in February 2001.

Finishing the rice genome

The original goal to finish the rice genome was 2008 (Sasaki and Burr 2000). In April 2000, Monsanto announced that it had sequenced a 6X draft of about 3,000 BACs across the rice genome and would share it with the IRGSP. The IRGSP is in the process of evaluating these new data; however, it is now apparent that just over half of the rice genome is covered by the Monsanto 6X draft. We are hopeful that, by combining the CUGI Rice Genome Framework with the Monsanto draft and the Rice Genome Program high-density genetic map, the IRGSP will be able to define a minimum tile of BAC clones across the entire rice genome. Such a tile would provide

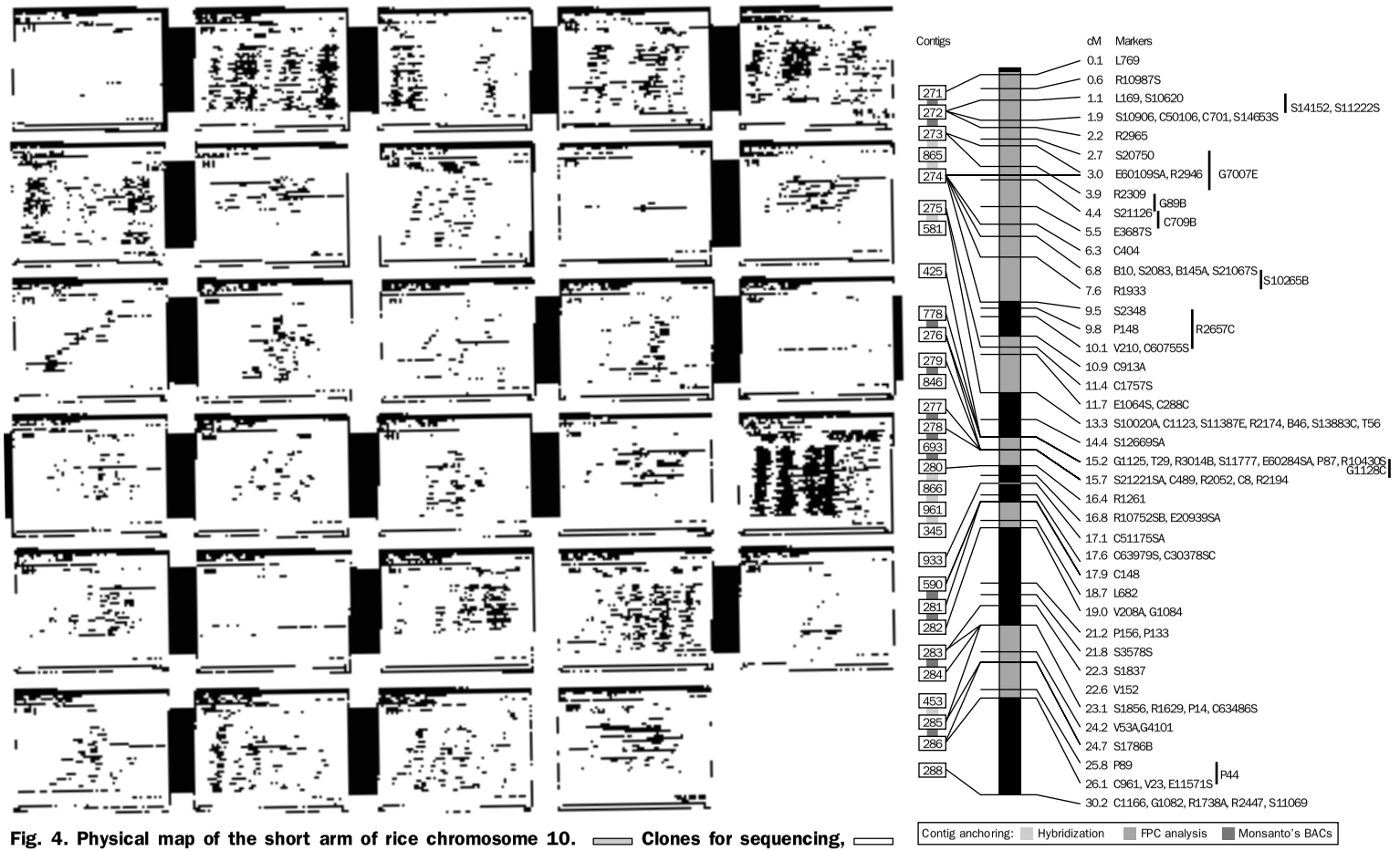


Fig. 4. Physical map of the short arm of rice chromosome 10. ▭ **Clones for sequencing,** ▭ **Monsanto's BACs.**

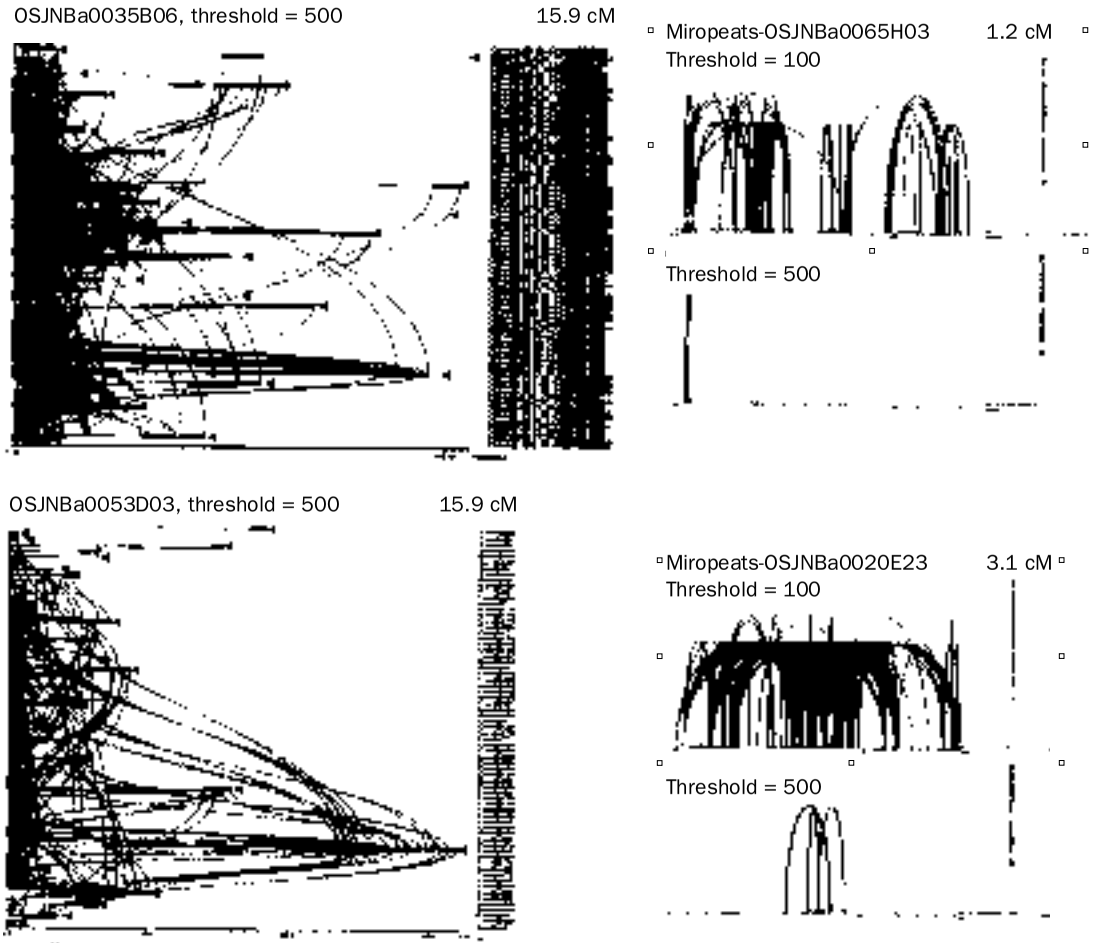


Fig. 5. Print Repeats/Miropeats output of the relationship between repeats among bacterial artificial chromosome (BAC) sequence assemblies for four BACs sequenced by CCW. OSJNBa0035B06 and OSJNBa0053D03 are located at 15.9 cM near the centromere. OSJNBa0065H03 and OSJNBa0020E23 are located near 1.2 and 3.0 cM, respectively. The BACs near the centromere have a complex repeat structure and will be difficult to finish.

Table 4. CCW sequencing status for year 1.

Goal	CUGI	CSHL	WUGSC	CCW	Sequence (Mb)
Finishing	12	20	0	32	4.17
Finished	3	5	1	9	1.39
In finishing	12	16	4	32	4.72
In production	0	0	0	0	0.00
In library	0	0	0	0	0.00
Total	15	21	5	41	6.11
Sequence (Mb)	2.17	3.21	0.74	6.11	

immediate access to the rice genome for positional cloning and marker-assisted selection. Moreover, it would allow the IRGSP to provide a set of defined BAC clones to sequence to each member of the IRGSP based on sequencing capacity and funding. Such a scheme would accelerate the completion of the sequencing of the rice genome, possibly as early as 2004.

References

Arumuganathan K, Earle E. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9:208-219.

Budiman M. 1999. Chapter III: Construction and characterization of a BAC library of *Oryza sativa* L. subsp. *japonica* cv. Nipponbare for genomic studies. PhD dissertation. Texas A&M University, College Station, Texas, USA. p 50-80.

Cheng Z, Presting G, Buell C, Wing R, Jiang J. 2001. High resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. Genetics 157:1749-1757.

Mao L, Wood T, Yu Y, Budiman M, Woo S, Sasinowski M, Goff S, Dean R, Wing R. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors (STCs). Genome Res. 10:982-990.

Marra M, Kucaba T, Dietrich N, Green E, Brownstein B, Wilson R, McDonald K, Hillier L, McPherson J, Waterston R. 1997. High throughput fingerprint analysis of large-insert clones. Genome Res. 7:1072-1084.

Sasaki T, Burr B. 2000. International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr. Opin. Plant Biol. 3:138-141.

Soderlund C, Humphray S, Dunham A, French L. 2000. Contigs built with fingerprints, markers and FPC V4.7. Genome Res. 10:1772-1787.

Sulston J, Mallett F, Durbin R, Horsnell T. 1989. Image analysis of restriction enzyme fingerprint autoradiograms. CABIOS 13:101-106.

Venter C, Smith H, Hood L. 1996. A new strategy for genome sequencing. Science 381:364-366.

Yuan Q, Liang F, Hsiao J, Zismann V, Benito M, Quackenbush J, Wing R, Buell R. 2000. Anchoring of rice BAC clones to the rice genetic map *in silico*. Nucleic Acids Res. 28:3636-3641.

Notes

Authors' addresses: R.A. Wing, Yeisoo Yu, G. Presting, D. Frisch, T. Wood, Sung-Sick Woo, M. Arief Budiman, Long Mao, Hye Ran Kim, T. Rambo, E. Fang, B. Blackmon, J.L. Goicoechea, S. Higingbottom, M. Sasinowski, J. Tomkins, R.A. Dean, C. Soderlund, Clemson University Genomics Institute, 100 Jordan Hall, Clemson, SC 29634; W.R. McCombie, R. Martienssen, M. de la Bastide, Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring Harbor, NY, 11724; R. Wilson, D. Johnson, Washington University Genome Sequencing Center, 4444 Forest Park Boulevard, St. Louis, MO 63108.

Acknowledgments: The authors would like to acknowledge financial support from the Rockefeller Foundation for BAC library development (RF98001#630 to RAW), the NSF Plant Genome Program for the CUGI BAC/EST Resource Center (#DBI-987276 to RAW, RAD, DF, MS), the Novartis Foundation for the fingerprint and STC databases, and the USDA-CSREES/NSF–Rice Genome Sequencing Project for the physical mapping, sequencing, and annotation of the short arms of chromosomes 10 and 3 (#99-35317-8505). We also thank the staff of CUGI, CSHL, and WU-GSC, past and present, for their hard work and dedication to the project. Finally, we especially thank T. Sasaki and members of the Japanese Rice Genome Program for providing access to their valuable molecular genetic resources.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Naturally occurring allelic variations as a new resource for functional genomics in rice

M. Yano

Recent progress in rice genome analysis has made it possible to analyze naturally occurring allelic variation underlying complex traits. Using heading date as a model for complex traits, we detected and characterized quantitative trait loci (QTLs) and identified genes at QTLs at the molecular level. QTLs for heading date were mapped by using several types of progeny derived from a cross between varieties Nipponbare and Kasalath. Nine QTLs were mapped precisely as single Mendelian factors by the use of advanced backcross progeny. Nearly isogenic lines of QTLs were also developed by marker-assisted selection and were used to identify the function of each detected QTL. Combining two QTLs into the genetic background of Nipponbare allowed us to investigate epistatic gene interactions among QTLs. We analyzed a large segregating population by genetic and physical mapping to narrow down candidate genomic regions for target QTLs. These analyses revealed 10- to 50-kb regions as candidate regions for the target genes. We identified genes of the most probable candidates for the photoperiod-sensitivity loci *Hd1* and *Hd6*. Naturally occurring allelic variation could be a new resource for the functional analysis of rice genes.

Structural analysis of the rice genome has progressed much in the past decade. More than 40,000 expressed sequence tags (ESTs) have been sequenced and deposited in the public DNA databases (Sasaki et al 1994, Yamamoto and Sasaki 1997). Moreover, a high-density restriction fragment length polymorphism (RFLP) linkage map has been constructed and about 60% genome coverage of yeast artificial chromosome (YAC) clone contigs has been achieved (Harushima et al 1998, Kurata et al 1997). This progress has allowed us to embark on the sequencing of the whole rice genome (Sasaki and Burr 2000). Resources—genetic markers, sequence data, and genomic clones—derived from such efforts will be used for the functional analysis of rice genes in the next decade. Although artificially induced variations, such as mutants, have been used mainly for genetic and physiological studies in rice and other plant species, the development of DNA markers has allowed access to naturally occurring allelic variations underlying complex traits (Tanksley 1993, Paterson 1995, Yano and

Sasaki 1997). Such analysis is often referred to as quantitative trait locus (QTL) analysis. Many QTLs have been mapped for many complex traits in rice (McCouch and Doerge 1995, Yano and Sasaki 1997).

During the analyses of several quantitative traits by the DNA marker-assisted strategy, two questions about QTL analysis have been raised: (1) Does a QTL represent a single Mendelian locus or a cluster of multiple loci? (2) Is it possible to precisely map a QTL and identify QTLs at the molecular level using map-based or other strategies? To answer these questions, we have performed a series of analyses on heading date as a model for complex traits. This chapter describes a comprehensive analysis of QTLs for heading date, including the identification of putative QTLs, characterization and fine mapping of QTLs using near-isogenic lines (NILs), and identification of genes at QTLs for heading date by the map-based strategy. In addition, we have developed several primary permanent mapping populations and secondary genetic resources, such as chromosomal segmental substitution lines, to facilitate the genetic analysis of naturally occurring allelic variations. What we have learned clearly indicates that naturally occurring allelic variations will be a new resource for functional genomics in rice.

Launching pad for new genetics in rice

The most important aspect of the genetic dissection of complex traits is the effective production of genetic markers in the target chromosome regions. Figure 1 summarizes potential resources for the production of chromosomal region-specific markers

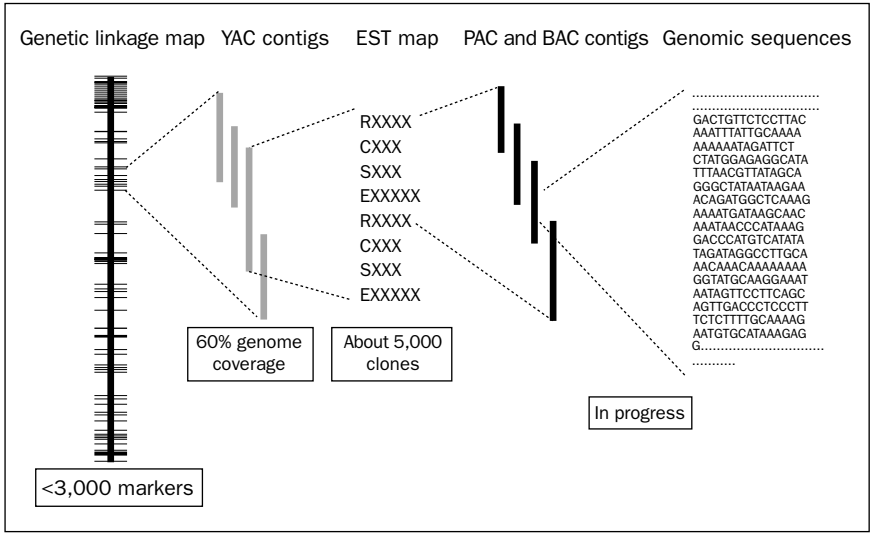


Fig. 1. Potential resources for the production of chromosomal region-specific markers in rice. EST = expressed sequence tags, YAC = yeast artificial chromosome, PAC = P1-derived artificial chromosome, BAC = bacterial artificial chromosome.

in rice. To this end, two high-density RFLP linkage maps were constructed to provide a framework of markers for the detection of individual factors controlling complex traits (Causse et al 1994, Harushima et al 1998, reviewed by Nagamura et al 1997). Kurata et al (1997) have also achieved about 60% genome coverage of yeast artificial chromosome (YAC) clone contigs. P1-derived artificial chromosome (PAC) and bacterial artificial chromosome (BAC) libraries have also been constructed in rice genome sequencing activities (Sasaki and Burr 2000). These YAC, PAC, and BAC clones are also very useful materials for developing region-specific genetic markers. Moreover, about 5,000 cDNA clones have already been mapped on YAC contigs that were placed on the genetic linkage map (Wu et al 2000). These cDNA clones can be used as a potential candidate for genetic markers in target chromosomal regions. Furthermore, the release of rice genome sequence data has begun (Sasaki and Burr 2000). Although completion of the sequencing of the rice genome will need additional efforts, the existing sequence data have already provided and will continue to provide an effective basis for developing new region-specific genetic markers. These resources will become a launching pad for the analysis of naturally occurring allelic variations in rice.

Genetic dissection of complex traits

Detection of QTLs controlling heading date

We have performed comprehensive analyses of heading date as a model trait for complex traits using the resources described above. Thirteen QTLs controlling heading date have been identified by using several progeny derived from Nipponbare (japonica) and Kasalath (indica) (Fig. 2). Five QTLs, *Hd1–Hd5*, have been mapped based on analysis of the F_2 population (Yano et al 1997) and an additional two, *Hd7* and *Hd11*, have been detected by using BC_1F_5 lines (Lin et al 1998). In addition, new loci involving heading date—*Hd6*, *Hd8–Hd10*, *Hd12*, and *Hd13*—have been detected by using advanced backcross progeny, such as BC_3F_2 or BC_4F_2 but not F_2 or BC_1F_5 (Yamamoto et al 2000, M. Yano, unpublished data). These results clearly indicate that not all factors involved in the target traits could be detected by using primary mapping populations, such as F_2 and recombinant inbred lines. One reason is the statistical limitation of QTL analysis. It is often difficult to detect QTLs with small phenotypic effects in the analysis of a primary population because of the noise of QTLs with a large effect or environmental variation (Tanksley 1993, Yano and Sasaki 1997). Another reason may be due to the existence of epistatic interaction among QTLs (Yamamoto et al 2000, Lin et al 2000). This point is described later.

Fine mapping of QTLs using advanced backcross progeny

Because of the statistical nature of QTL analysis, the chromosomal location of a QTL is not accurate and allows no information about the function of the genes. Populations used in previous QTL analyses of rice were mainly F_2, BCF_1, recombinant inbred lines, or doubled-haploid lines (McCouch and Doerge 1995, Yano and Sasaki 1997). These populations, which segregate multiple genetic factors on the whole genome simulta-

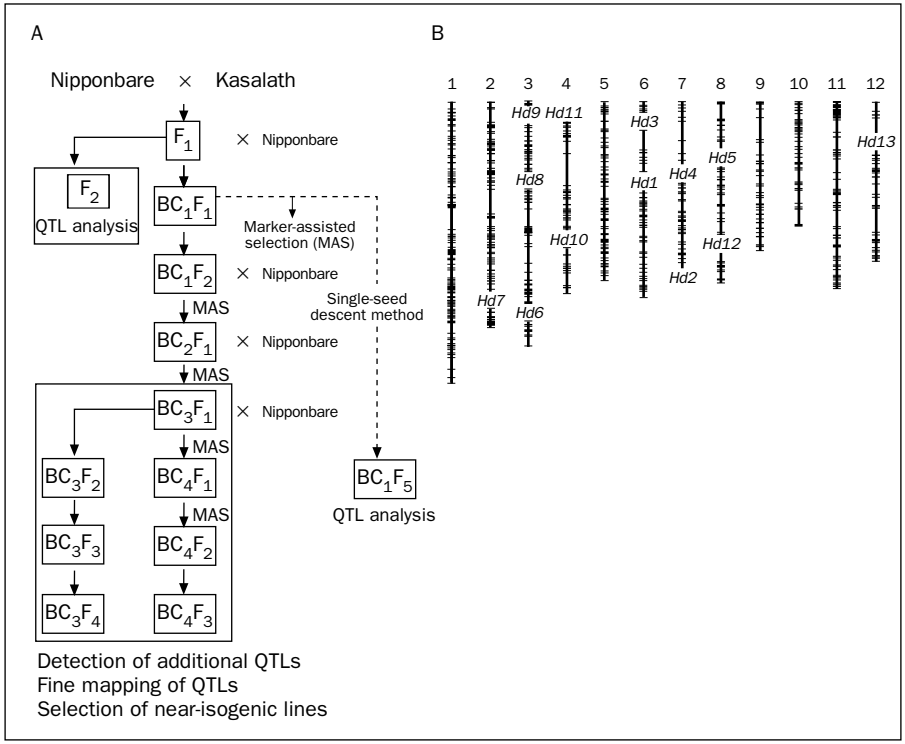


Fig. 2. Plant materials used to detect QTLs and chromosomal locations of QTLs controlling heading date. (A) Mapping populations derived from a cross between Nipponbare and Kasalath. QTL mapping was performed by using F_2, BC_1F_5, and BC_3F_2 lines. Fine mapping and selection of near-isogenic lines were conducted using advanced backcross progeny. (B) A high-density RFLP linkage map showing chromosomal locations of putative QTLs (*Hd1–Hd13*) for heading date.

neously, cannot be used for the precise mapping of one of multiple QTLs. Moreover, determination of the true genetic action of the QTL is more difficult because the genetic parameters of a given QTL are often affected by the segregation of other QTLs.

In plant genetics, NILs developed by backcrossing have been widely used to perform accurate genetic analysis because the segregating populations obtained from crossing NILs and their recurrent parents simplify genetic variation by excluding extra-genetic factors. To obtain an answer to the main question about QTLs—Does a QTL represent a single locus or multiple loci?—we performed fine mapping of a QTL as a single Mendelian factor using advanced backcross progeny. So far, we have shown that 9 of 13 QTLs for heading date could be mapped precisely on the RFLP linkage map as single factors (Yamamoto et al 1998, 2000, H.X. Lin and M. Yano, unpublished data). Only one chromosomal region could be dissected into two factors with different functions in the control of heading (L. Monna and M. Yano, unpublished data).

Characterization of genes at QTLs using NILs

To characterize QTLs, we used marker-assisted selection (MAS) to develop NILs differing only in the presence of a single specific QTL (QTL-NILs) for heading date in rice (Fig. 3). Each line contained the chromosomal region of the target QTL from the donor variety Kasalath in the genetic background of variety Nipponbare. Days to heading was investigated under various controlled-daylength conditions. The QTLs

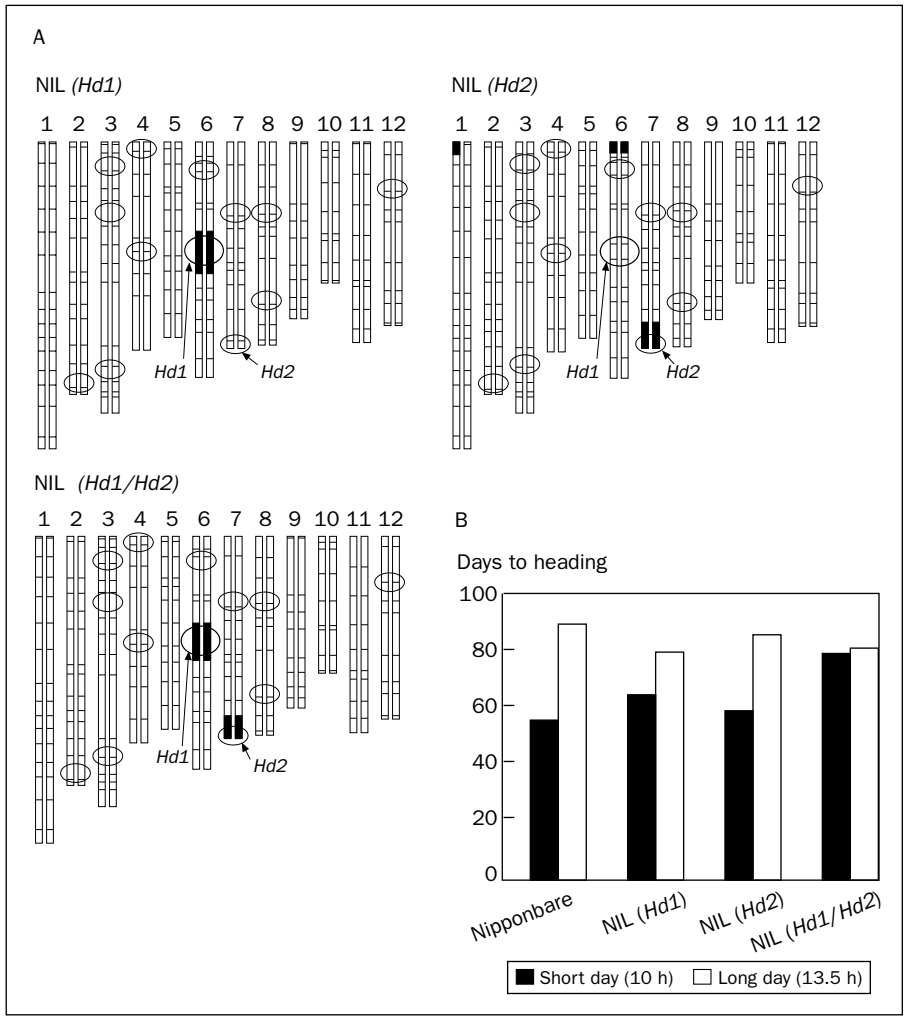


Fig. 3. (A) Graphical representation of genotypes of near-isogenic lines for *Hd1*, *Hd2*, and *Hd1/Hd2*, and (B) days to heading under short- and long-day conditions. The 12 pairs of bars represent the chromosomes, numbered at the top. The horizontal lines on the bars show positions of marker loci used in marker-assisted selection. The 13 circles on the chromosomes represent regions for heading-date QTLs detected. Open bars and solid bars show segments of the chromosomes derived from Nipponbare and Kasalath, respectively.

were classified into two groups based on response of QTL-NILs to daylength. Five QTLs—*Hd1*, *Hd2*, *Hd3*, *Hd5*, and *Hd6*—were found to confer photoperiod sensitivity (Yamamoto et al 2000, Lin et al 2000). Three NILs, for *Hd7*, *Hd8*, and *Hd9*, however, did not vary in response to daylength. This result indicates that these three loci do not confer photoperiod sensitivity. Functions of other QTLs are under investigation.

Analysis of epistatic interaction among QTLs

Many studies have been performed to detect epistatic interactions among QTLs by using primary populations, such as F_2 and recombinant inbred lines, but successful examples of detection seem to be relatively few (Tanksley 1993, Yano and Sasaki 1997). The more the number of contributing QTLs increases, the more difficult it is to detect significant differences to distinguish individual QTLs without using a huge population size. Yano et al (1997) predicted an interaction between *Hd1* and *Hd2*, the two largest QTLs. But the existence of *Hd6* and its interaction could not be detected in their analysis of an F_2 population (Yamamoto et al 2000). This result suggests that many epistatic interactions do exist among minor QTLs that are not detected in the primary population. Thus, it is necessary to develop new experimental materials, such as chromosomal substitution lines or NILs, for a better understanding of epistasis among QTLs.

The effectiveness of this strategy was proved by the analysis of photoperiod-sensitivity QTLs in two studies (Lin et al 2000, Yamamoto et al 2000) and in other work (Doebley et al 1995, Eshed and Zamir 1996). It was possible to detect epistatic interactions among QTLs in three ways: constructing QTL-NILs for each detected QTL by MAS based on the results of primary QTL analysis, combining QTLs by crossing all QTL-NILs, and comparing each phenotype with each combining QTL genotype. In addition, secondary F_2 populations derived from a cross between QTL-NILs were used to detect and confirm epistasis between QTLs. The results showed that three photoperiod-sensitive QTLs, *Hd1*, *Hd2*, and *Hd3*, interacted with each other (Lin et al 2000). Based on these results, we suggest that the Kasalath allele of *Hd3* does not affect photoperiod sensitivity by itself but that it is involved in enhancing the expression of the Nipponbare alleles of *Hd1* and *Hd2*. We also showed that digenic epistatic interaction prevented us from detecting major QTLs in a primary mapping population such as F_2 (Yamamoto et al 2000).

In the QTL analysis of 186 F_2 plants by Yano et al (1997), the phenotypic difference caused by segregation of the chromosomal region where *Hd6* is located was surveyed under the simultaneous segregation of two major photoperiod-sensitive QTLs, *Hd1* and *Hd2*. This situation could not secure a large enough population to detect the gene effect of *Hd6*, which shows epistatic interaction. As a result, the variance due to the difference in genotypes might not be distinguishable from the variance due to the segregation of other QTLs and to environmental error. On the other hand, the existence of an additional QTL, *Hd6*, and epistatic interaction between *Hd2* and *Hd6* were clearly detected in the analysis of the advanced progeny.

In summary, in QTL analysis of a population in which a QTL with a large effect will segregate, a putative gene effect of an epistatic QTL can be recognized as a small effect only, even if the actual gene effect is large. The complex inheritance of quantitative traits is probably explained mainly by multigenic control, but epistatic interaction is also an important factor in such complexity.

Molecular dissection of a complex trait

Map-based cloning or candidate identification of genes at QTLs

Figure 4 shows a procedure for the molecular identification of genes at QTLs from detection to cloning. After the detection of putative QTLs with major effects in a primary mapping population, the construction of special genetic stocks, such as NILs or substitution lines, can be effectively and efficiently achieved by using MAS. For some traits that can be evaluated with high reliability, such as heading date and culm length, high-resolution and fine-scale mapping of putative QTLs as Mendelian factors will be feasible by combining the use of NILs and ordinal linkage mapping based on progeny testing. For traits whose expression and performance are affected largely by environmental factors, it would be more difficult to employ simple progeny testing, such as F_2 plant/F_3 lines. In such a case, advanced progeny testing might be

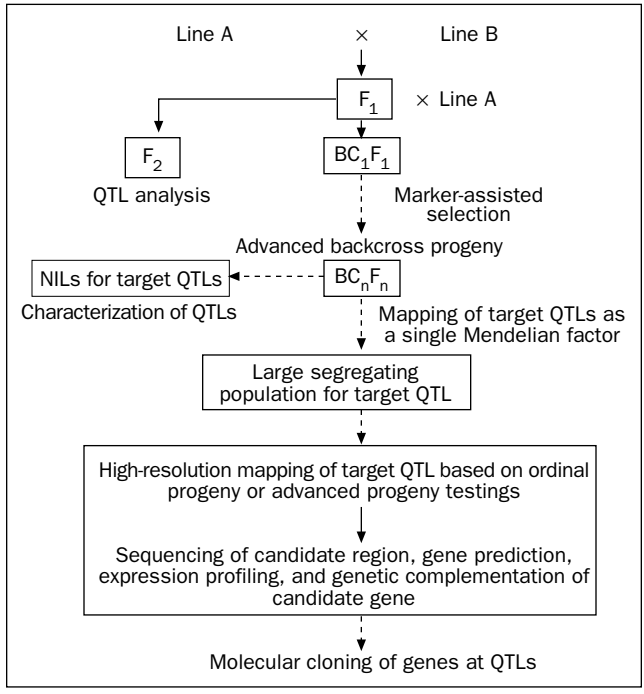


Fig. 4. Schematic representation of a strategy for map-based cloning of genes at QTLs. NILs = near-isogenic lines.

effective at determining the genotypes of target QTLs. We should raise generations beyond the second to obtain fixed lines of a recombinant chromosome for more reliable phenotyping.

Once a given QTL can be mapped as a single Mendelian factor using advanced backcross progeny, chromosomal walking and landing methods can be used to identify genes at QTLs (Tanksley 1993, Tanksley et al 1995, Yano and Sasaki 1997). It will be necessary to increase the population size enough to minimize the candidate genomic region. In our analysis of heading-date QTLs, high-resolution linkage mapping allowed us to define a genomic region of <50 kb as a candidate using 1,000–2,000 plants. We have developed a new polymerase chain reaction-based marker, the cleaved amplified polymorphic sequence (CAPS) marker, and have used it to facilitate these labor-intensive experiments.

The production of region-specific markers is another crucial factor in the fine-scale mapping of QTLs. As mentioned above, a large number of resources—YAC, PAC, and BAC clones and about 5,000 ESTs—are available for the production of new markers. Combining the analysis of a large segregating population and region-specific markers allowed us to define a candidate genomic region of <50 kb. After defining a candidate genomic region, we can use several molecular approaches, such as sequencing of the candidate region, gene prediction, expression profiling, and genetic complementation of the candidate gene.

A major QTL for photoperiod response, *Hd1*, was recently isolated and found to encode a protein with a zinc finger structure (Yano et al 2000). This protein shows high similarity with *CO*, a gene for photoperiod response in *Arabidopsis*. The most probable candidates of *Hd3a* and *Hd6* were also identified; genetic complementation is in progress. We have also conducted high-resolution mapping for *Hd2*, *Hd3b*, *Hd5*, and *Hd9*. The development of genomic clone contigs and identification of candidate genes are in progress for these QTLs.

Genetic control of photoperiod sensitivity in rice

So far, there is no model for the genetic control pathway of rice heading. A series of analyses of heading date based on naturally occurring allelic variations gave us several factors and their relationships in the genetic control pathway of photoperiod response in rice.

The function of *Hd1* is likely to affect transcription activation because of the presence of a zinc finger domain, which is required to bind DNA. As *Hd1* transcription itself was not greatly affected by daylength change, we speculate that *Hd1* affects the transcription level of genes whose expression level is controlled by daylength change (Yano et al 2000). Nonfunctional alleles of *Hd1* and other genes at QTLs for photoperiod response were combined. Analysis of epistatic interactions revealed that *Hd1* is epistatic to other genes enhancing photoperiod response, such as *Hd2* and *Hd3* (Lin et al 2000). These results suggest that *Hd1* plays a central role in the expression of photoperiod response under both short- and long-day conditions. However, this study suggests that the expression of *Hd1* might not be affected by daylength changes. These observations raise a major question: Which factors are involved in the dramatic change

in the response to daylength in rice? Other factors whose expression is affected by daylength change might be involved downstream of *Hd1* in the genetic control pathway.

The identification of other QTLs by map-based cloning is progressing. This simultaneous approach to identifying QTLs controlling photoperiod sensitivity will contribute to elucidating the genetic control pathway for photoperiod response in rice.

Plant materials for the analysis of quantitative traits

To facilitate the genetic analysis of complex traits in rice, permanent mapping populations based on japonica × indica or japonica × japonica crosses are being constructed. Recombinant inbred lines (RILs), backcross inbred lines (BILs), and doubled-haploid lines have been developed and framework linkage maps for permanent use are being constructed (Table 1, Tsunematsu et al 1996, Lin et al 1998, M. Yano and S.Y. Lin, unpublished data). These permanent mapping populations will facilitate the detection of naturally occurring allelic variations for target traits. However, these materials are not enough to allow us to proceed with further analysis, such as fine mapping and characterization of target QTLs. Secondary mapping populations, such as chromosomal substitution lines or NILs, will be required to facilitate more comprehensive analysis of target QTLs. To this end, we have been developing chromosomal substitution lines derived from crosses between Nipponbare and Kasalath and between Koshihikari (japonica) and Kasalath (Table 1).

The construction of a series of intraspecific substitution lines of rice with overlapping chromosomal segments is progressing by marker-assisted selection (Table 1). In

Table 1. Permanent mapping populations for the analysis of naturally occurring allelic variations.

Cross combination	Population structure[a]	Generation	No. of lines	Genotype data[b]	Institution involved[c]
Nipponbare/Kasalath	BIL	BC_1F_{10}	98	245 RFLPs	NIAR, STAFF, Hokuriku NAES
Nipponbare/Kasalath	SL	BC_3F_7/BC_4F_6	115	125 RFLPs	NIAR, STAFF, Hokuriku NAES
Asominori/IR24	RIL	F_{12}	71	375 RFLPs	Kyushu University, NIAR, STAFF
Koshihikari/Akihikari	DH	A_4	214	108 RFLPs 48 RAPDs	Ishikawa Agricultural College, NIAR, STAFF
Sasanishiki/Habataki	BIL	BC_2F_6	85	238 RFLPs	Hokuriku NAES, NIAR, STAFF
Koshihikari/Kasalath	BIL	BC_1F_7	183	135 RFLPs	NIAR, STAFF
Koshihikari/Kasalath	SL	BC_3F_2/BC_4F_1	Under development		NIAR, STAFF

[a]RIL = recombinant inbred line, BIL = backcross inbred line, SL = chromosome segment substitution line, DH = doubled-haploid line. [b]RFLPs = restriction fragment length polymorphisms, RAPDs = randomly amplified polymorphic DNAs. [c]NIAR = National Institute of Agrobiological Resources, STAFF = Institute of Society for Techno-innovation for Agriculture, Forestry and Fisheries, Hokuriku NAES = Hokuriku National Agricultural Experiment Station.

addition, some accessions of wild relatives have been used as donor parental lines to develop chromosome segmental introgression lines at Kyushu University (Doi et al 1997, Sobrizal et al 1999). A wide range of cross combinations will be very important for detecting naturally occurring allelic variations because wild relatives and ecotypes are adapted to specific environmental conditions and variations in genes give advantages for adaptation. We might find a wider range of allelic variation in these wild relatives than in cultivated species.

Future prospects

Although many mutants have contributed to our understanding of gene functions so far, the application of mutational approaches in rice, compared with the model plant *Arabidopsis thaliana*, is limited for several reasons, including large plant size, long life cycle, and large seed size. As described in this chapter, the use of naturally occurring allelic variations will be an alternative resource for functional genomics in rice as well as in *Arabidopsis* (Alonso-Blanco and Koornneef 2000).

The use of naturally occurring allelic variations compared with mutational approaches has several advantages. First, naturally occurring allelic variations may determine not only the presence or absence of allele function but also leaky or weakly functional alleles. This would be more informative for the analysis of an important gene; for example, when the gene function is completely lost, the plant does not survive. Multicopy genes (redundant genes), whose function is always complemented by other members, cannot be easily analyzed by simple gene disruption analysis because genes with no or very small phenotypic effect cannot be analyzed genetically. On the other hand, various kinds of allele combinations of redundant genes will occur in the progeny of different ecotypes or wild relatives. Dramatic phenotypic changes will be detected in this type of progeny, even if the frequency is very low. An example is gametic abortion causing seed sterility.

Second, phenotype assays often require large-scale field or greenhouse experiments and a long time period. Applications of these types of assays are limited to mutational approaches based on chemical, physical, and transposon mutagenesis. It is currently not possible to perform such large-scale assays on transgenic mutagenesis, such as T-DNA tagging or Ac/Ds mutagenesis. In contrast, the use of naturally occurring allelic variations needs the survey of only 100–200 lines or individuals from the primary mapping population. This approach will be a very powerful strategy for analyzing yield performance, abiotic and biotic stress tolerance, eating and cooking qualities, and so on. Although the analysis of naturally occurring allelic variations has several disadvantages compared with mutational analysis, such as small allelic differences in phenotypes, the approach will complement mutational approaches in the functional analysis of rice genes.

We are now employing this strategy to analyze several complex traits, such as seed shattering habit, plant height, seed size, and pollen sterility; tolerance for abiotic stresses such as cold temperature and ultraviolet irradiation; and tolerance for nutrient imbalances. Only one cross combination has given us a wide range of resources for

functional genomics. However, it is obvious that the most important issue for the effective use of naturally occurring allelic variations is to provide a wide mapping population derived from wild relatives and different ecotypes. We will need to develop not only primary mapping populations, such as RILs or doubled-haploid lines, but also secondary mapping populations, such as chromosomal segmental substitution lines. In addition, precise and reliable phenotype assays will be necessary in the use of naturally occurring allelic variations in functional genomics. Combining old techniques such as crossing and selection with new tools such as DNA markers and sequences will contribute greatly to the functional analysis of rice genes.

References

Alonso-Blanco C, Koornneef M. 2000. Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. Trends Plant Sci. 5:22-29.

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251-1274.

Doebley J, Stec A, Gustus C. 1995. *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 141:333-346.

Doi K, Iwata N, Yoshimura A. 1997. The construction of chromosome substitution lines of African rice (*Oryza glaberrima* Steud.) in the background of Japonica rice (*O. sativa* L.). Rice Genet. Newsl. 14:39-41.

Eshed Y, Zamir D. 1996. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147-1162.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

Kurata N, Umehara Y, Tanoue H, Sasaki T. 1997. Physical mapping of the rice genome with YAC clones. Plant Mol. Biol. 35:101-113.

Lin SY, Sasaki T, Yano M. 1998. Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa* L., using backcross inbred lines. Theor. Appl. Genet. 96:997-1003.

Lin HX, Yamamoto T, Sasaki T, Yano M. 2000. Characterization and detection of epistatic interactions of three QTLs, *Hd1*, *Hd2* and *Hd3*, controlling heading date in rice using nearly isogenic lines. Theor. Appl. Genet. 101:1021-1028.

McCouch SR, Doerge RW. 1995. QTL mapping in rice. Trends Genet. 11:482-487.

Nagamura Y, Antonio BA, Sasaki T. 1997. Rice molecular genetic map using RFLPs and its application. Plant Mol. Biol. 35:79-87.

Paterson AH. 1995. Molecular dissection of quantitative traits: progress and prospects. Genome Res. 5:321-333.

Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkala I, Minobe Y. 1994. Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J. 6:615-624.

Sasaki T, Burr T. 2000. International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr. Opin. Plant Biol. 3:138-141.

Sobrizal, Ikeda K, Sanchez PL, Doi K, Angeles ER, Khush GS, Yoshimura A. 1999. Development of *Oryza glumaepatula* introgression lines in rice, *O. sativa* L. Rice Genet. Newsl. 16:107-108.

Tanksley SD. 1993. Mapping polygenes. Annu. Rev. Genet. 27:205-233.

Tanksley SD, Ganal MW, Martin GB. 1995. Chromosome landing: a paradigm for map-based cloning in plants with large genomes. Trends Genet. 11:63-68.

Tsunematsu H, Yoshimura A, Harushima Y, Nagamura Y, Kurata N, Yano M, Sasaki T, Iwata N. 1996. RFLP framework map using recombinant inbred lines in rice. Breed. Sci. 46:279-284.

Wu J, Yamamoto S, Maehara T, Harada C, Shimokawa T, Ono N, Mukai Y, Takazaki Y, Yazaki J, Koike K, Shomura A, Fujii F, Yano M, Ando T, Kono I, Sasaki T, Matsumoto T. 2000. A comprehensive map of rice expressed sequence tags (ESTs): its characteristics and application to genome analysis. Abstracts for the 6th International Congress of Plant Molecular Biology. S1, Plant Genomics, S01–92.

Yamamoto K, Sasaki T. 1997. Large-scale EST sequencing in rice. Plant Mol. Biol. 35:135-144.

Yamamoto T, Kuboki Y, Lin SY, Sasaki T, Yano M. 1998. Fine mapping of quantitative trait loci *Hd-1*, *Hd-2* and *Hd-3*, controlling heading date of rice, as single Mendelian factors. Theor. Appl. Genet. 97:37-44.

Yamamoto T, Lin HX, Sasaki T, Yano M. 2000. Identification of heading date quantitative trait locus *Hd6*, and characterization of its epistatic interaction with *Hd2* in rice using advanced backcross progeny. Genetics 154:885-891.

Yano M, Sasaki T. 1997. Genetic and molecular dissection of quantitative traits in rice. Plant Mol. Biol. 35:145-153.

Yano M, Harushima Y, Nagamura Y, Kurata N, Minobe Y, Sasaki T. 1997. Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. Theor. Appl. Genet. 95:1025-1032.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T. 2000. *Hd1*, a major photoperiod-sensitivity QTL in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell. 12:2473-2484.

Notes

Author's address: National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba 305-8602, Japan.

Acknowledgments: This work was supported by funds from the Ministry of Agriculture, Forestry and Fisheries of Japan (MAFF), the Japan Racing Association (JRA), and the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN).

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Deletion mutants for functional genomics: progress in phenotyping, sequence assignment, and database development

H. Leung, C. Wu, M. Baraoidan, A. Bordeos, M. Ramos, S. Madamba, P. Cabauatan, C. Vera Cruz, A. Portugal, G. Reyes, R. Bruskiewich, G. McLaren, R. Lafitte, G. Gregorio, J. Bennett, D. Brar, G. Khush, P. Schnable, G. Wang, and J. Leach

A collection of IR64 mutant populations has been established using diepoxybutane, fast neutron, and gamma ray mutagenesis. Phenotypic screening is being conducted on about 12,000 M_3 or M_4 lines for morphological variations and altered response to biotic and abiotic stresses. About 6% of the mutants exhibit morphological variations in the vegetative and reproductive stages. Disease-response mutants are recovered at approximately 0.3%. Based on the frequency of visible and conditional mutants observed under biotic stresses, the number of mutated sites per genome is estimated to be high (>10) in these mutants.

Mutants are screened for gain and loss of resistance against bacterial blight, blast, sheath blight, and tungro viruses. We have identified mutants with a gain in resistance to tungro viruses, blast, and bacterial blight. Lesion mimic mutants with enhanced resistance to both blast and bacterial blight are found. To dissect the defense pathways, double disease-response mutants have been produced to examine the epistatic interactions and expression profiles of host defense genes. The mutant collection is being evaluated for tolerance of submergence and salinity. Mutants are also being screened for altered response under water stress at the flowering stage.

Nucleotide-binding site and leucine-rich repeat sequences are used to detect genomic changes in loss-in-resistance mutants; however, the efficiency of this approach is limited by the number of available candidate genes. To efficiently assign DNA sequences to mutant phenotypes, we are developing a gene array-based screening strategy using subtractive cDNA/expressed sequence tag libraries enriched for stress-response genes. We are also producing additional mutant populations to generate a range of deletion sizes that are suitable for array-based and high-throughput PCR screening. As a first step, we are using stress-response and morphological mutants with phenotypic and genetic descriptions to develop a mutant database to be linked to sequence databases. We expect the mutant database to grow as the research community continues to use this mutant resource.

Mutational analysis has long been the foundation of genetic analysis and discovery of gene function. With sequence information becoming increasingly abundant and accessible through whole-genome sequencing projects, mutational analysis takes on an even more important role in the assignment of putative functions to these genes in a genome. In rice, the systematic production of rice mutants began in several laboratories largely in response to the rapid progress in structural delineation of the rice genome. As described in several papers from this symposium, the production and characterization of mutants have been the centerpiece of functional genomics programs around the world.

Most of the mutant populations established or being produced by researchers worldwide are insertional mutants using T-DNA and heterologous or endogenous transposons (Greco et al, An et al, and Hirochika et al, this volume). The main advantage of insertion lines is their ability to disrupt a gene and at the same time tag the sequences responsible for the altered phenotypes. Furthermore, specially designed insertion vectors can be used to tag enhancer elements and to activate gene expression, which cannot be done by other means (Martienssen 1998, Weigel et al 2000). At present, because of the genotype dependency of transformation and tissue manipulation, all insertional mutants have been made in a few japonica rice genotypes.

Since more than 80% of the rice grown in the world is indica type with a distinct suite of adaptive characteristics, it is important to produce indica rice mutants that encompass a large amount of biological variation relevant to improving agronomic traits. A large deletion stock induced by chemical or irradiation mutagenesis would serve this purpose. Because of the ease and low cost of production, it is feasible to produce mutant collections in different genetic backgrounds according to the need for identifying biological variation. We expect that the disadvantage of not tagging mutations will be alleviated by high-throughput genotyping and gene array technologies. The availability of the complete genome sequences will further simplify positional cloning. Here we describe our progress in producing and characterizing a physical and chemical mutagen-induced population of IR64, a popular indica rice variety widely grown in the tropics. IR64 carries many valuable agronomic traits related to yield, plant architecture, grain quality, and tolerance of biotic and abiotic stresses. Creating mutations in such an elite genetic background can facilitate the detection of phenotypic changes in important agronomic traits. We expect that such a large collection of nontransgenic mutants will complement the insertional lines as a global resource for rice functional genomics.

Strategy for producing an indica rice mutant bank by chemical and irradiation mutagenesis

Throughout the long history of rice breeding and genetic studies, geneticists and plant breeders have collected many spontaneous and induced mutants over time. The notable collections are maintained at Kyushu University, Hokkaido University, and the National Institute of Genetics in Mishima, Japan. IRRI has a collection of about 400 morphological mutants accumulated over many years of breeding and genetic work.

Although these mutants are useful for identifying single-gene defects, their diverse genetic backgrounds make them less suitable for the systematic comparison of biological variation. Thus, a comprehensive mutant population in an identical genetic background is needed for phenotyping and functional assignment of rice sequences.

To begin the mutant project, we evaluated several mutagens that would give sizable deletions that could be physically detected by reverse genetics (Aguirrezabalaga et al 1995, Yandell et al 1994). We selected diepoxybutane (DEB), fast neutron, and gamma ray to build the mutation collection based on such properties (Bruggemann et al 1996, Reardon et al 1987, Shirley et al 1992). For instance, DEB is known to produce kilobase range deletions in plants and *Drosophila*, whereas fast neutron is expected to produce larger deletions and translocations. The requirement to produce large deletions, however, is no longer necessary with the development of a high-throughput technique to detect single-base changes (McCallum et al 2000a; see discussion below). Our goal is to develop a collection of mutants with a range of deletion sizes and point mutations that are suitable for both forward and reverse genetics.

Prior to the production of IR64 mutants, several mutant collections were made to identify specific genes. G. Khush and D. Brar (unpublished) produced an ethyl methanesulfonate-induced mutant population of IR36 in an attempt to select for mutations related to apomixis. P. Ronald, University of California-Davis, made another collection in IRBB21 using DEB and fast neutron. The IRBB21 mutants carry *Xa21* in an IR24 background, providing the materials for investigating genes involved in the *Xa21* gene-mediated resistance pathway (see Wang et al, this volume).

We have used the IRBB21 mutant collection to establish the parameters in screening for disease-response mutants and to determine the molecular events associated with the mutations. Approximately 4,000 IRBB21 M_2 lines were screened for a change from resistance to susceptibility against the bacterial pathogen strain PXO99 (avirulent to *Xa21*). We recovered 31 mutants that have become fully susceptible (10 mutants) or partially susceptible (21 mutants) to nine races of the bacterial blight pathogen. Southern and polymerase chain reaction (PCR) analyses of the molecular changes at the *Xa21* locus revealed two classes of fully susceptible mutants: those with a loss of both the kinase domain and leucine-rich repeats and those with rearrangements in the kinase domain. Based on the presence or absence of a marker flanking *Xa21*, we estimated that some deletions are >100 kb whereas others are under 10 kb, consistent with those reported in mutagenesis of other eukaryotes.

For the production of IR64 mutants, our target is to produce 40,000 independent lines to give a high probability of inducing a mutation in most genes (except homozygous lethals). Following the calculation of Krysan et al (1996) for T-DNA tagging in *Arabidopsis*, the probability of inducing a mutation in a gene in a given population size can be estimated. Assuming that the average size of rice genes is approximately 3.2 kb (as in *Arabidopsis*), the rice genome of 430,000 kb could be imagined as 134,375 targets (430,000/3.2). The mutation rate of 1/1,000 observed in the rice mutants suggests that each mutant probably harbors 10–20 mutated sites. Thus, estimating conservatively, 10,000 mutant lines would carry $10 \times 10{,}000$ mutated sites. We can apply the formula of $P = 1 - (1 - f)^N$ to determine P, the probability of any gene being

marked by a deletion/chromosomal rearrangement, where f = 1/number of targets and N = number of mutations surveyed. With 10,000 mutants screened, the probability of detecting a mutation in a specific gene is about 52%.

Figure 1 summarizes the status of the IR64 mutant collections produced by three mutagens—DEB, fast neutron, and gamma ray (cobalt 60). It should be emphasized that we deliberately chose to induce as many mutations as the genome can tolerate. We reason that the background mutations will mostly be silent. Once an interesting phenotype is identified, genetic control can be determined by cosegregation analysis. In the first phase of mutant production, we maintained the pedigree by extracting 6–10 plants from an M_2 family. Although it was useful to have sister lines to infer inheritance, this advantage is outweighed by the burden of maintaining a large number of lines with redundant mutations. Thus, in the second stage, we advanced all M_1 lines by single-seed descent.

Phenotypic characterization

Morphological variations

We used several obvious mutant phenotypes with known genetic control to compare the spectrum of mutations produced by different mutagens (Table 1). Approximately 6% of the mutant population shows morphological variants. Most of the morphological mutants have obvious phenotypes that affect the vegetative and reproductive stages. It appears that the frequencies of inducing variability by three mutagens are similar (0.2% to 0.05%). We have not made special attempts to examine quantitative varia-

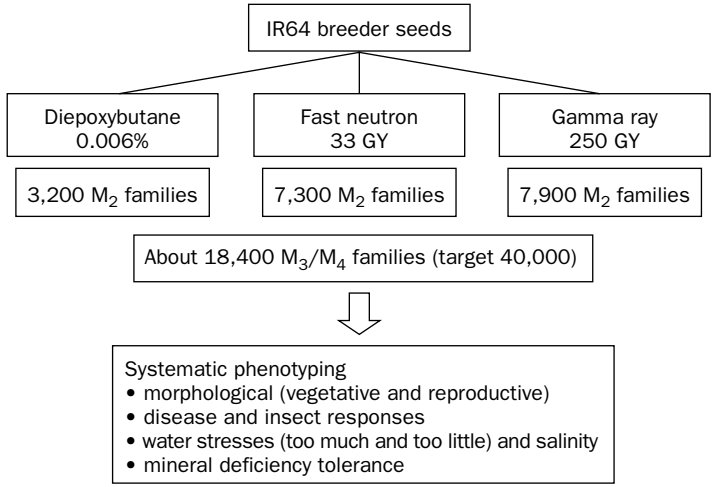


Fig. 1. Development of IR64 mutants using three mutagens: diepoxybutane (DEB), fast neutron (FN), and gamma ray (GR). Treatment doses are indicated. Both FN and GR treatments were performed at the International Atomic Energy Commission, Vienna, Austria. Gy = Gray, 1 joule kg^{-1} of target specimen.

Table 1. Frequencies of mutations observed in IR64 populations treated with three mutagens.

Mutant phenotype	Treatment					
	Gamma ray		Diepoxybutane		Fast neutron	
	Observed (no.)	%	Observed (no.)	%	Observed (no.)	%
Zebra	1/482	0.21	1/500	0.07	2/531	0.38
Spotted leaf (*spl1*)	2/1,000	0.20	2/3,000	0.07	1/2,000	0.05
Elongated upper internode	1/100	0.10	1/1,500	0.07	1/531	0.18
White stripe	1/482	0.21	2/1,500	0.13	1/336	0.30
Extra glume	2/425	0.47	0/1,500	0	1/531	0.18
Gold hull	1/1,000	0.10	1/1,500	0.07	2/531	0.38

Table 2. Categories of IR64 disease-response mutants against blast and bacterial blight.

Mutation	Morphology[a]	Number of mutants resistant to[b]				
		Magnaporthe grisea races		*Xanthomonas oryzae* pv. *orzyae* races		Both diseases
		Single	Multiple	Single	Multiple	
Loss in resistance	Normal	0	2	2	1	–
Gain in resistance	Normal	2	3	2	–	3
	Abnormal	2	–	2	–	2
Lesion mimics	Normal	–	–	–	–	–
	Abnormal	0	–	16	–	3

[a]Abnormal morphology includes dwarfing, stunting, and reduced tillering. [b]Numbers refer to mutants genetically confirmed by M_3 segregation data. Four *M. grisea* isolates representing diverse genetic groups were used to screen for susceptibility to blast. *Xoo* isolate PXO61 was used to screen for a change from resistance to susceptibility to bacterial blight. PX099 was used to screen for a change from susceptibility to resistance. – = mutant class has not been found.

tion, such as flowering time, number of tillers, and degree of sterility. More careful observation in quantitative variation is expected to yield additional mutants.

Biotic stresses

To identify genes involved in broad-spectrum resistance, we have searched for mutants with a gain or loss in resistance to multiple races of a pathogen and to multiple pathogens. We have screened about 10,000 lines for altered response to the blast fungus *Magnaporthe grisea* and the bacterial blight pathogen *Xanthomonas oryzae* pv. *oryzae (Xoo)*. About 0.3% of the mutants showed altered response to the pathogens. The observed mutation frequencies are similar to those found in other plant species mutagenized by chemicals or irradiation (Okubara et al 1994). Single-gene inheritance has been confirmed in approximately 30 mutants with altered response to one or two pathogens (Table 2). Of special interest are those mutants with a loss or gain in resistance to multiple races of one or more pathogens. For example, a gamma-ray-induced mutant GR-740-12-8 shows a loss in resistance to five of six races of *Xoo* to which IR64 is resistant and enhanced susceptibility to four races. Segregation

analysis suggests that the mutation in GR-740-12-8 is dominant. Another gamma-ray-induced mutant, GR-978-18-18, shows strong resistance to 10 races of *Xoo* from the Philippines and to two isolates of the blast pathogen. This mutant is slightly shorter than the wild-type IR64. Another mutant, GR-282-12-15, showed a gain in resistance to a single isolate of *Xoo* and two blast isolates. Segregation patterns in the M_3 generation suggest that both mutations in GR-978-18-18 and in GR-282-12-15 are recessive. Overall, a majority of these gain-in-resistance mutants are morphologically abnormal, ranging from semidwarf to severe stunting. In *Arabidopsis*, some gain-in-resistance mutants were also morphologically abnormal (Clough et al 2000). As of now, we have not identified mutants showing susceptibility to multiple pathogens, indicating a need to expand the screening to detect these presumably rare mutations.

The most common class of disease-response mutants is those with lesion mimics. Yin et al (2000) demonstrated that some but not all lesion mimic mutants (in another genetic background of indica rice IR68) exhibit enhanced resistance to blast and bacterial blight. To further delineate the pathways involved in disease and defense responses, we selected 21 lesion mimic mutants from the IR64 mutant collection. We found that these mutants showed a varying degree of enhanced resistance to different bacterial blight strains and that resistance was not race-specific. Two mutants, D6-1851 and D6-242, with reduced resistance to *Xoo* and *M. grisea* were monitored for expression of a pathogenesis-related (PR) protein gene *PBZ1* (Fig. 2). In the wild-type and single mutants, the expression of *PBZ1* was strongest in the second youngest leaf (L2) but decreased in the older leaves (L3 and L4). In the double mutant, which showed enhanced resistance, high expression of *PBZ1* was maintained in the older leaves.

Double mutants have been created by crossing independently inherited lesion mimic mutants to evaluate the epistatic relationships among these mutations. Defense response (DR) genes *PR-1*, peroxidase 22.3 (*POX22.3*), and peroxidase POC1 (*POC1*) were used to probe mRNA from four leaf stages of the single and double mutants derived from D6-1851 × D6-242. *POX22.3* expression was high in young leaves and diminished as lesions developed. *PR-1* and *POC1* expression was enhanced relative to that observed in IR64 in the second and third leaves. In general, DR gene expression was positively correlated with enhanced disease resistance to bacterial blight. These results suggest that the expression of these DR genes is developmentally controlled and regulated by the lesion mimic genes and that mis-regulation of certain defense-related genes was correlated with the resistant phenotypes. Several subtractive cDNA libraries induced by pathogens have been constructed and a panel of candidate genes is being used to correlate gene expression with broad-spectrum disease resistance (J. Leach, published data).

Sheath blight caused by *Rhizoctonia solani* is a difficult disease to evaluate because of the variability associated with the disease response and the strong environmental influence on pathogen-host interactions. However, there is an urgent need to search for endogenous genes that can enhance sheath blight resistance as no genetic resistance has yet been found through conventional screening of rice germplasm. Our strategy is to first identify mutants with a gain or loss in resistance to blast and bacte-

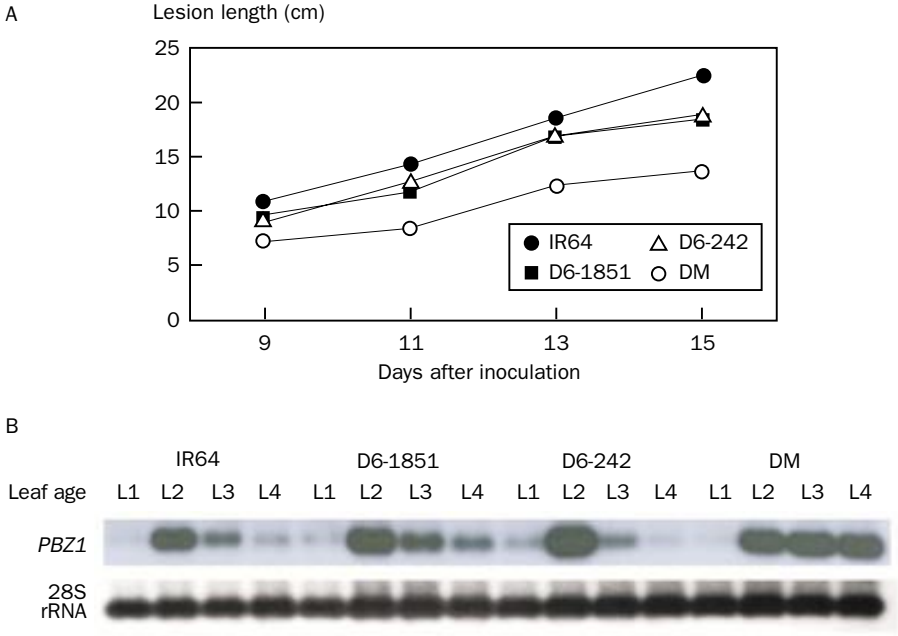


Fig. 2. (A) Reduced susceptibility to bacterial blight in single and double mutants (DM) carrying lesion mimic mutations. (B) Expression of *PBZ1* in leaf of different ages (L1 is youngest and L4 is oldest). The double mutant maintains a high level of expression in older leaves, which corresponds to enhanced resistance to bacterial blight.

rial blight, and then evaluate these mutants with sheath blight under field conditions. So far, several lesion mimic mutants have been evaluated for reaction to sheath blight but no enhanced resistance has been found. In addition, we have begun screening for tolerance of tungro viruses using viruliferous green leafhopper carrying both the spherical (RTSV) and bacilliform (RTBV) viruses. From a screening of about 2,000 lines, six putative mutants with different degrees of tolerance of RTSV were identified, and these are being confirmed for both the phenotypes and genotypes.

Abiotic stresses

We have initiated mutant screenings under drought, submergence, salinity, and zinc deficiency conditions. In preliminary screening experiments, quantitative variation was observed in zinc deficiency tolerance but no variability for salinity stress (100 mM NaCl) was found. From a screening of approximately 3,000 lines for submergence tolerance under field conditions, G. Gregorio and D. Adorada (personal communication) identified six putative mutants that give 75% recovery after 14-d submergence treatment at the seedling stage. These mutants are being reevaluated to confirm inheritance of submergence tolerance and yield performance.

For drought-response screening, our objective is to identify mutant lines that affect the transduction of stress signals that influence grain filling under stress. In 1999, 1,200 mutant lines were evaluated for yield under continuous flooding and with floodwater drained prior to flowering. The stress treatment delayed flowering by 2 d and reduced yield by about 28%. After using spatial analysis to compensate for variation within the field, we identified 65 lines with superior yield in either the stress or control treatment. These lines were evaluated in replicated experiments in 2000; they differed significantly in yield in both the control and stress plots, but the stress treatment (field drained 18 d after panicle initiation) had little effect on yield because of unseasonable rain. An additional 1,200 deletion lines were evaluated in 2000. They were uniform in appearance and flowering date. Under stress, some variation in senescence rate was observed. Based on visual scores and on grain yields under stress relative to the IR64 check, 26 lines were identified for further assessment of variation in component traits related to tolerance of late-season stress (R. Lafitte, unpublished data, Fig. 3).

Opportunities for other traits

Results from screening for phenotypic variation in morphology and responses to biotic and abiotic stresses suggest that the IR64 population is rich in genetic variability.

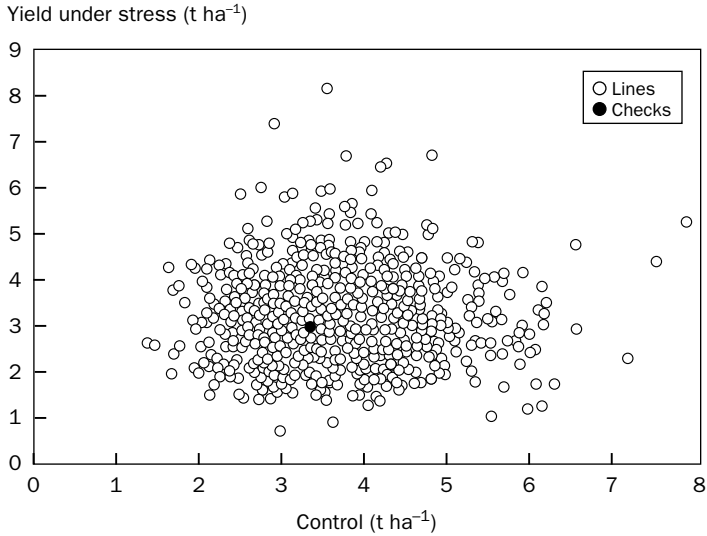


Fig. 3. A scatter plot of yield response of IR64 mutants to water stress at the IRRI farm. Mutant lines were grown under water stress and control conditions. The irrigated control was flooded a few weeks later and remained with 5–10 cm of standing water until about 1 wk before harvest. The water-stress block was flooded a few weeks after transplanting but was drained 1 wk before flowering; thus, plants were under stress during grain filling. Accelerated senescence was observed in the stress. Grain yield (t ha^{-1}) was extrapolated from plot-size yield. Wide deviation from the 1:1 slope suggests a contrasting performance under the two environments.

However, to detect genetic variability in complex traits, a reproducible screening technique is much needed. For example, stringent and uniform submergence screening allows for recovery of lines with quantitative enhancement in submergence tolerance. On the other hand, considerable variability exists in grain yields among the mutant lines under drought stress and control environments, illustrating the difficulties in identifying quantitative response using parameters that could be affected by multiple factors. It appears that measurement of multiple characters with adequate replications would be essential in detecting mutations that affect quantitative traits.

Another factor to consider in conducting a productive mutant screening is the likelihood of detecting a gain or loss in functions in the parental genotype. Using the parental phenotype as a guide, we need to design screening methods with sufficient sensitivity to detect deviations from the parental value. For example, IR64 has an intermediate level of tolerance of zinc deficiency. Whether we can detect variation in oversensitivity or enhanced tolerance may depend on the level of available zinc in the test soil to differentiate the phenotypic extremes.

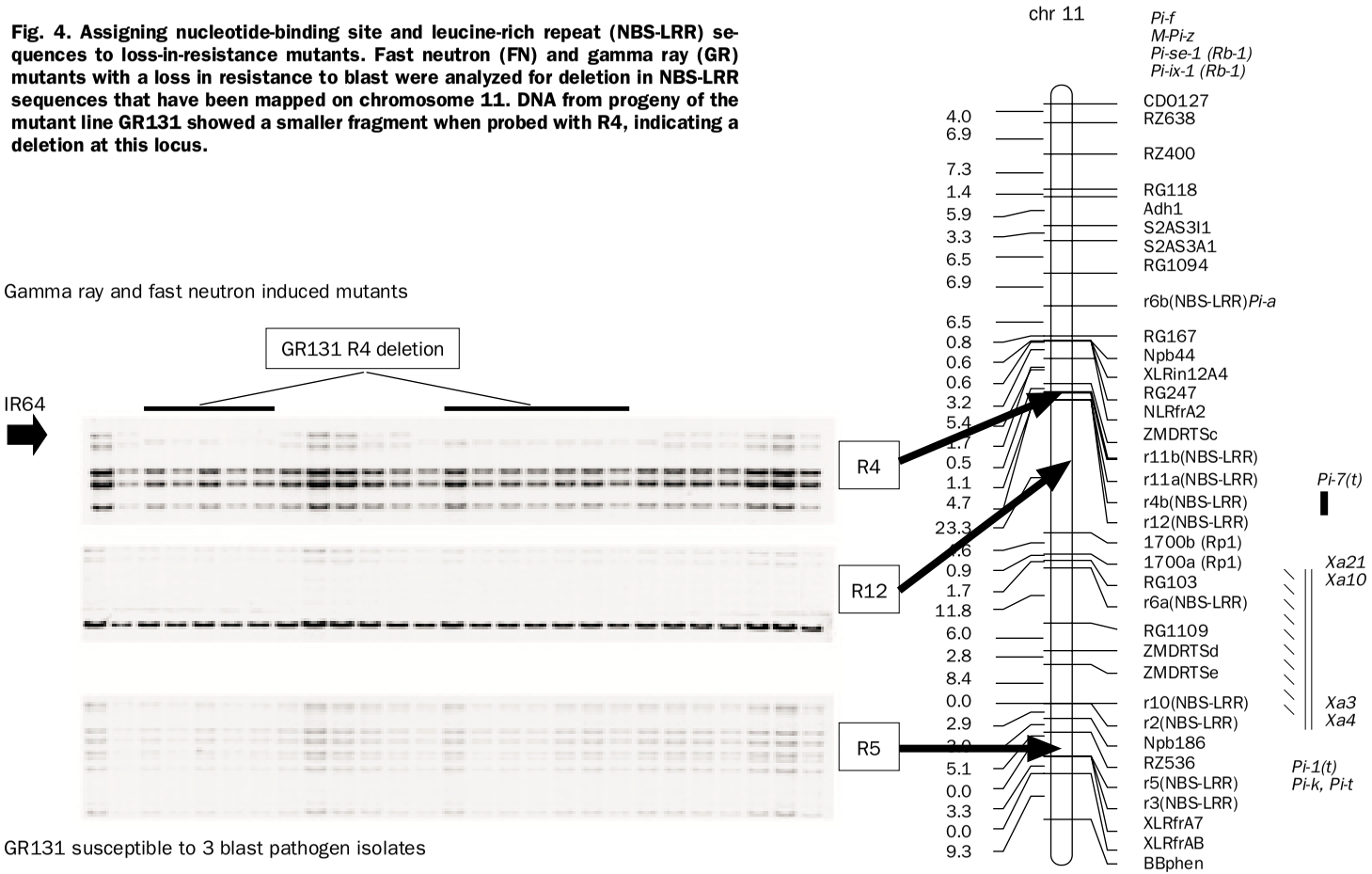
Given the abundance of useful traits in IR64, opportunities exist to search for conditional mutants under laboratory and field conditions. Possible candidates are variants under photosynthesis-limiting conditions, tolerance of soil toxicity, and quantitative resistance to brown planthopper. Recently, Wurtzel et al (2001) described a simple approach to isolate rice mutants exhibiting a block in carotenoid biosynthesis. Although this class of mutants may not be useful in developing pro-vitamin A rice, the approach of isolating mutants blocking micronutrient biosynthesis and availability may prove useful for identifying genes to improve nutrients in rice grain.

Reverse genetics systems

Our original goal in producing a deletion stock is to create a sizable gap in a gene such that it can be physically detected by PCR or hybridization methods. To examine whether deletions can be detected in our mutant population, we selected disease-susceptible mutants and examined a set of putative disease-resistance-gene sequences (nucleotide-binding site and leucine-rich repeat, NBS-LRR, sequences) for detecting deletions. In one experiment, DNA from several blast-susceptible mutants was probed with three NBS-LRR sequences—R4, R12, and R5—previously mapped on chromosome 11 (Leister et al 1998). A small deletion was detected in a line (GR131) with R4 but not with the other probes (Fig. 4, M. Ramos-Pamplona and H. Leung, unpublished). Cosegregation of this deletion and loss of resistance is in progress to determine whether R4 is responsible for the change from resistance to susceptibility. This experiment indicates that deletion can be detected if a collection of candidate genes and preselected mutants is available. However, the challenge is to be able to apply DNA sequences to screen DNA from a pool of random mutants and isolate the mutants with specific genetic lesions.

At least two screening methods have been developed as reverse genetics tools in nontransgenic mutants. Liu et al (1999) described the use of PCR screening of a DNA pool from a deletion stock of *Caenorhabditis elegans*. This approach uses gene-spe-

Fig. 4. Assigning nucleotide-binding site and leucine-rich repeat (NBS-LRR) sequences to loss-in-resistance mutants. Fast neutron (FN) and gamma ray (GR) mutants with a loss in resistance to blast were analyzed for deletion in NBS-LRR sequences that have been mapped on chromosome 11. DNA from progeny of the mutant line GR131 showed a smaller fragment when probed with R4, indicating a deletion at this locus.



cific PCR primers to screen a DNA pool of deletion lines. If the target gene has a deletion bracketed by the primers, a length polymorphism would be detected in the PCR products. More recently, McCallum et al (2000a,b) described a technique called targeting induced local lesions in genomes (TILLING) to detect base-pair mismatches. In this case, primers are designed to bracket the target gene. A single base-pair mismatch between the mutant and wild-type DNA sequences results in a heteroduplex that can be detected by denaturing high-performance liquid chromatography (DHPLC) or by specific enzyme cleavage precisely at the 3′-end of the mismatch (Oleykowski et al 1998, Colbert et al 2001). The TILLING procedure is attractive as it is well suited for high-throughput screening. Furthermore, the high density of mutations induced by EMS makes it possible to cover the genome with relatively few individuals (approximately 10,000 lines). We have generated a collection of EMS-induced mutants and DNA pools are being produced and subjected to TILLING. Successful application of a reverse genetics system to the IR64 mutants will greatly enhance the value of the collection of mutants and increase our ability to assign sequences to each of the mutant lines.

IR64 mutant database

We have systematically collected phenotypic information on the mutant stocks with the goal of linking such information to genomic databases. As rice genome data become available, we envision that all mutant lines can be assigned to one or more physical locations on the sequence map. Together with other collections of mutants, it is feasible to assign an allelic series of insertions, deletions, and point mutations to every gene in the genome.

As a first step, we have used the International Crop Information System (ICIS) database to document phenotypic and molecular data on the IR64 mutant stocks. An ontology of component traits of mutant phenotypes is being defined to allow structured searches of the database. The database of mutants has about 200 phenotypically described mutants. A prototype Web interface (www.cgiar.org/irri/genomics/index.htm) has been developed to allow users to search the database according to mutant phenotypes. The next phase of database development will involve linking phenotypic data to molecular characterization of the mutants to facilitate the identification of candidate genes.

Conclusions

The IR64 mutant collection provides a needed resource for gene discovery, particularly for indica varieties that contain many adaptive traits for rice grown in the tropics. Our data suggest that the IR64 mutants exhibit a wide range of variability useful for detecting conditional mutants. We will expand the mutant collection with a range of deletion sizes and point mutations such that an efficient reverse genetics system can be developed using high-throughput genotyping methods. Through database linkages, we will place these mutants onto the chromosomal regions by association with known mutants or traits and by reverse genetics.

Although mutants with discrete genetic lesions are ideal genetic materials for assigning function to a particular gene, there are only two possible alleles at a locus in each parental genotype. The success of a mutant screening therefore depends on discriminating the phenotypic effect of changing a single allele. Some mutated alleles may not have observable phenotypes because of minor effects or gene redundancy. The production of double or triple mutants is probably needed to understand the effects and interactions of multiple genes. Even after a sequence is assigned to a mutant by reverse genetics, phenotyping, possibly guided by the known disrupted sequence, remains the critical step for assigning function. With the availability of genetic tools and sequence information, ample opportunities exist to make use of the allelic diversity created in mutants together with the natural variability present in breeding materials and raw germplasm to better understand the pathways that control complex traits.

References

Aguirrezabalaga I, Nivard MJM, Comendador MA, Vogel EW. 1995. The cross-linking agent hexamethylphosphoramide predominantly induces intra-locus and multi-locus deletions in postmeiotic germ cells of *Drosophila*. Genetics 139:649-658.

Bruggemann E, Handwerger K, Essex C, Storz G. 1996. Analysis of fast neutron-generated mutants at the *Arabidopsis thaliana* HY4 locus. Plant J. 10:755-760.

Clough SJ, Fengler KA, Yu I-c, Lippok B, Smith Jr. RK, Bent AF. 2000. The *Arabidopsis dnd1* "defense, no death" gene encodes a mutated cyclic nucleotide-gated ion channel. Proc. Natl. Acad. Sci. USA 97:9323-9328.

Colbert TG, Till BJ, Tompa R, Reynolds SH, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S. 2001. High-throughput screening for induced point mutations. Plant Physiol. (In press.)

Krysan PJ, Young JC, Tax F, Sussman MR. 1996. Identification of transferred DNA insertions within *Arabidopsis* genes involved in signal transduction and ion transport. Proc. Natl. Acad. Sci. USA 93:8145-8150.

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci. USA 95:370-375.

Liu LX, Spoerke JM, Milligan EL, Chen J, Reardon B, Westlund B, Sun L, Abel K, Armstrong B, Hardiman G, King J, McCague L, Basson M, Clover R, Johnson CD. 1999. High-throughput isolation of *Caenorhabditis elegans* deletion mutants. Genome Res. 9:859-867.

Martienssen RA. 1998. Functional genomics: probing plant gene function and expression with transposons. Proc. Natl. Acad. Sci. USA 95:2021-2026.

McCallum CM, Comai L, Green EA, Henikoff S. 2000a. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol. 123:439-442.

McCallum CM, Comai L, Green EA, Henikoff S. 2000b. Targeted screening for induced mutations. Nature Biotechnol. 18:455-457.

Okubara PA, Anderson PA, Ochoa OE, Michelmore RW. 1994. Mutants of downy mildew resistance in *Lactuca sativa* (lettuce). Genetics 137:867-874.

Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT. 1998. Mutation detection using a novel plant endonuclease. Nucleic Acids Res. 26:4597-4602.

Reardon JT, Liljestrand-Golden CA, Dusenbery RL, Smith PD. 1987. Molecular analysis of diepoxybutane-induced mutations at the *rosy* locus of *Drosophila melanogaster*. Genetics 115:323-331.

Shirley BS, Hanley S, Goodman HM. 1992. Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent *testa* mutations. Plant Cell 4:333-347.

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dixon RA, Harrison MJ, Lamb C, Yanofsky M, Chory J. 2000. Activation tagging in *Arabidopsis*. Plant Physiol. 122:1003-1013.

Wurtzel ET, Luo R, Yatou O. 2001. A simple approach to identify the first rice mutants blocked in carotenoid biosynthesis. J. Exp. Bot. 52:161-166.

Yandell MD, Edgar LG, Wood WB. 1994. Trimethylpsoralen induces small deletion mutations in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 91:1381-1385.

Yin Z, Chen J, Zeng L, Goh M, Leung H, Khush GS, Wang GL. 2000. Characterization of rice lesion mimic mutants and identification of a mutant with broad-spectrum resistance to rice blast and bacterial blight. Mol. Plant-Microbe Interact. 13:869-876.

Notes

Authors' addresses: H. Leung, C. Wu, M. Baraoidan, A. Bordeos, M. Ramos, S. Madamba, P. Cabauatan, C. Vera Cruz, A. Portugal, G. Reyes, R. Bruskiewich, G. McLaren, R. Lafitte, G. Gregorio, J. Bennett, D. Brar, G. Khush, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines; P. Schnable, Department of Agronomy, Iowa State University, Ames, IA 50011, USA; G. Wang, Department of Plant Pathology, Ohio State University, Columbus, OH 43210, USA; J. Leach, Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502, USA.

Acknowledgments: This work has been supported by IRRI core research funds and in part by grants from the Rockefeller Foundation, United States Department of Agriculture, and United States Agency for International Development.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Generation of T-DNA insertional tagging lines in rice

Gynheung An, Jong-Seong Jeon, Sichul Lee, Ki-Hong Jung, Sung-Hoon Jun, Dong-Hoon Jeong, Jinwon Lee, Seonghoe Jang, Shin-Young Lee, Kiyoung Yang, Byoungho Lee, Sinok Moon, Kyungsook An, Min-Jung Han, Jung-Hwa Yu, Namok Lee, Su-Young An, Sun-Hee Park, Eun-Sik Song, In-Soon Park, and Hyun-Sook Lee

T-DNA insertions have produced approximately 30,000 transgenic lines of rice. Polymerase chain reaction and genomic DNA gel-blot analyses have shown that approximately 65% of the population contains more than one copy of the inserted T-DNA. Hygromycin resistance tests determined the number of T-DNA inserts to be an average of 1.4 genetic loci per plant. From this, it can be estimated that at least 40,000 taggings have been generated.The binary T-DNA vector used in the insertion contained the promoterless *ß-glucuronidase* (*gus*) reporter gene with an intron as well as multiple splicing donors and acceptors immediately next to the right border. This gene trap vector is designed to detect a gene fusion between *gus* and the endogenous gene that is tagged by the T-DNA. The leaves, roots, mature flowers, and developing seeds of the transgenic rice plants were subjected to histochemical GUS assays. The results showed that 1.6–2.1% were GUS-positive in the tested organs and that their *gus* expression patterns were organ- or tissue-specific or ubiquitous in all parts of the plant. This large population of T-DNA tagged lines will be useful for identifying insertional mutants in various genes and for discovering new genes in rice.

Strategies to discover the functions of plant genes have recently been developing rapidly. These strategies have been largely based on genetic approaches such as mutant identification and map-based gene isolation (reviewed in Martin 1998). The use of a transposon insertion to inactivate genes has been employed for functional studies in several plant species. Similarly, mutagenesis through the use of transfer DNA (T-DNA) has also been developed for tagging genes in *Arabidopsis* (Babiychuk et al 1997, Feldmann 1991, Krysan et al 1999). T-DNA insertion is believed to be a random event and the inserted genes are stable through multiple generations (reviewed in Azpiroz-Leehan and Feldmann 1997).

The development of several strategies for screening T-DNA or transposon insertions in a known gene and recovering sequences flanking the insertions have made

insertional mutagenesis an attractive method for functional analysis (Cooley et al 1996, Couteau et al 1999, Frey et al 1998, Krysan et al 1999, Liu and Whittier 1995). Through the sequencing of polymerase chain reaction (PCR)-amplified fragments adjacent to the inserted element, a flanking sequence database has been constructed in *Arabidopsis* (Parinov et al 1999, Tissier et al 1999).

Reporter genes have been used as insertional elements in order to aid in the identification of insertions within functional genes (Campisi et al 1999, Kertbundit et al 1991, 1998, Topping et al 1991, Sundaresan et al 1995). An enhancer trap contains a weak minimal promoter fused to a reporter gene and a gene trap contains multiple splicing sites fused to a reporter gene. The *gus* gene has been the most frequently used reporter gene because its gene products can be accurately detected and because of the tolerance of amino-terminal translational fusions in its enzyme activity (Jefferson et al 1987).

Rice is a model plant of cereal species because of its relatively small genome size, efficient tools for plant transformation, construction of physical maps, large-scale analysis of expressed sequence tags (ESTs), international genome sequencing projects, and economic importance (Hiei et al 1994, Sasaki 1998). Therefore, the development of insertional mutant lines will be extremely valuable for the functional genomics of rice. To this end, we are producing a large population of rice lines that are tagged with T-DNA.

Results

Vector construction for insertional mutagenesis in rice

Two binary vectors were constructed for T-DNA insertional mutagenesis of rice (Fig. 1A). The first plasmid, pGA1633, contains the promoterless *gus* gene immediately next to the right border and the cauliflower mosaic virus (CaMV) 35S promoter–*hygromycin phosphotransferase (hph)* chimeric gene as a selectable marker. The second plasmid, pGA2144, was constructed to increase gene trap efficiency. In this plasmid, an intron carrying three putative splicing donors and acceptors was placed in front of *gus*. In pGA2144, we replaced the CaMV 35S with the strong promoter of the rice α-tubulin gene *OsTubA4* along with its first intron for expression of the selectable marker *hph* gene.

Production of T-DNA-tagged transgenic rice plants

Scutellum-derived embryonic calli were cocultivated with *Agrobacterium tumefaciens* LBA4404 carrying the binary tagging vector. Approximately 20–40% of the cocultivated calli produced hygromycin-resistant cells. The frequency of plant regeneration from the calli ranged from 50% to 85% (data not shown). *Agrobacterium*-mediated rice transformation procedures have been developed using the system based on the super-virulent strain and super-binary vectors carrying the virulence region of pTiBo542 (reviewed in Hiei et al 1997). Our results showed that the transformation efficiency of our system was as high as that of the super-binary vector system, indicating that *Agrobacterium* strain LBA4404 and a common binary vector can be used for efficient transformation of rice. With this system, we have produced 1,600 transgenic

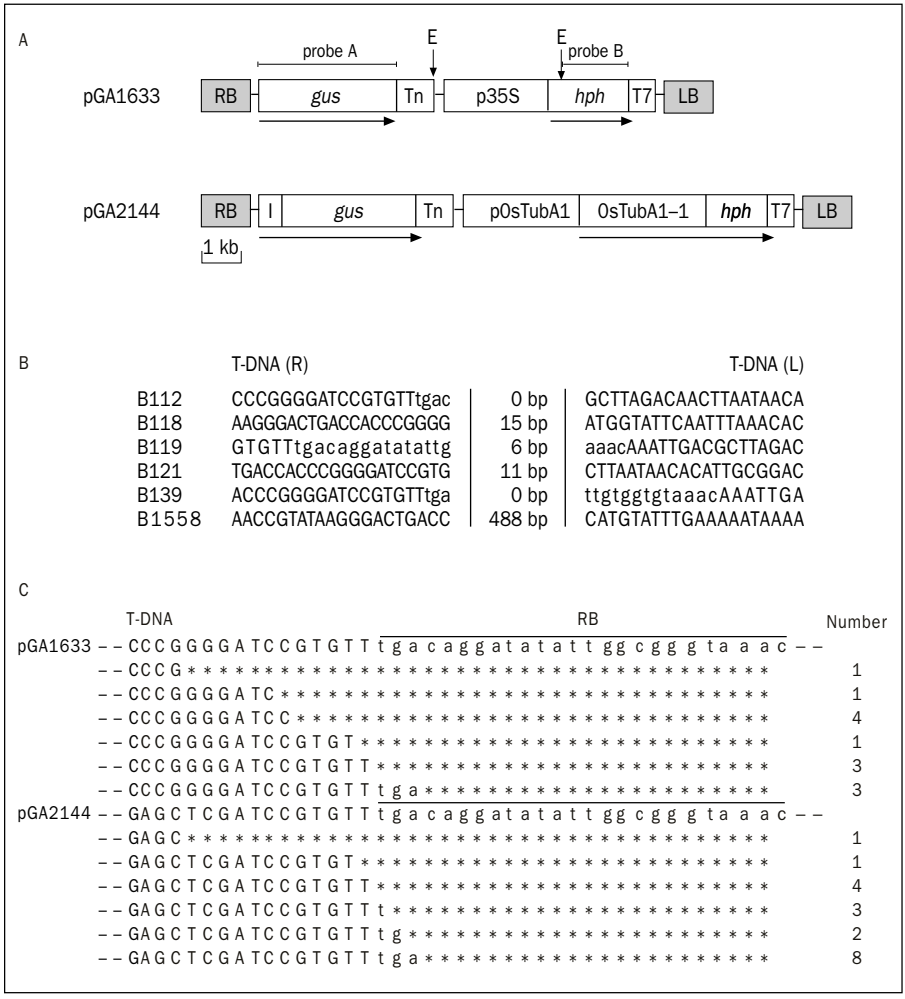


Fig. 1. Maps of the T-DNA tagging vectors and analysis of T-DNA integration patterns. (A) Schematic diagrams of pGA1633 and pGA2144. The RB and LB in shaded boxes represent the right border and left border of T-DNA, respectively. E = *Eco*RI site, *gus* = *β-glucuronidase*, Tn = nopaline synthase (*nos*) terminator, p35S = cauliflower mosaic virus 35S promoter, *hph = hygromycin phosphotransferase*, T7 = transcription termination region of gene 7 of the *pTiA6*, I = the *OsTubA4* intron 3 carrying three putative splicing acceptor and donor sites, *pOsTubA4* = the promoter of the rice α-tubulin gene, and *OsTubA4* = OsTubA4-1, the first intron of *OsTubA4*. The *gus* probe (probe A) and *hph* probe (probe B) used for DNA blot analyses are indicated. (B) Sequences at the junctions between two directly integrated T-DNA copies. The capitals indicate T-DNA sequences immediately next to the right (R) and left (L) border sequences (lower case). The length of filler sequences is depicted. (C) Sequences of T-DNA at the junction to plant DNA. The T-DNA sequences of pGA1633 and pGA2144 are shown in capitals and the right border (RB) sequences are shown in lower case. Asterisks signify any nucleotides presumably originated from the rice genome.

plants transformed with pGA1633 and 30,000 transgenic plants transformed with pGA2144.

Molecular characterization of the T-DNA integration pattern in transgenic rice plants

The number of integrated T-DNA in each plant was estimated from randomly selected primary transformants. Figure 2 is a result of the genomic DNA gel-blot analysis of 34 transgenic lines that were hybridized with the *gus* or *hph* coding region. It showed that 11 lines carried a single copy of the *gus* gene and 13 carried a single copy of the *hph* gene. The remaining lines carried two or more copies of *gus* or *hph*. This result indicates that approximately 35% of the transgenic lines carry a single T-DNA insert. In several lines, the numbers of the *gus* and *hph* genes were different from each other, probably due to T-DNA rearrangement during the transformation process (Ohba et al 1995, see below).

The number of T-DNA insertion loci was estimated by scoring hygromycin-resistant progeny (T2). Twenty-four of 34 lines appeared to carry T-DNA at one locus, whereas the remaining 10 lines contained an unlinked T-DNA insertion (Fig. 3). This indicates that transgenic plants contain an average of 1.4 genetic loci of T-DNA inserts. These data are similar to the result observed in *Arabidopsis* that T-DNA tagged plants contain an average of 1.4 inserts (Feldmann 1991). The number of insertion loci estimated by hygromycin resistance was smaller than the number of T-DNA copies evaluated by the DNA gel-blot analysis. This result was probably due to tandem integration of two or more T-DNA copies into a single chromosome as observed previously in dicot plants (Krizkova and Hrouda 1998). A PCR approach was undertaken to investigate the T-DNA arrangement of the lines that carry multiple T-DNAs at a single chromosome. The result showed that T-DNA copies were arranged in direct or inverted repeats (data not shown). We carried out sequence analysis of the regions between the T-DNA borders from six lines that carry multiple T-DNA copies at a

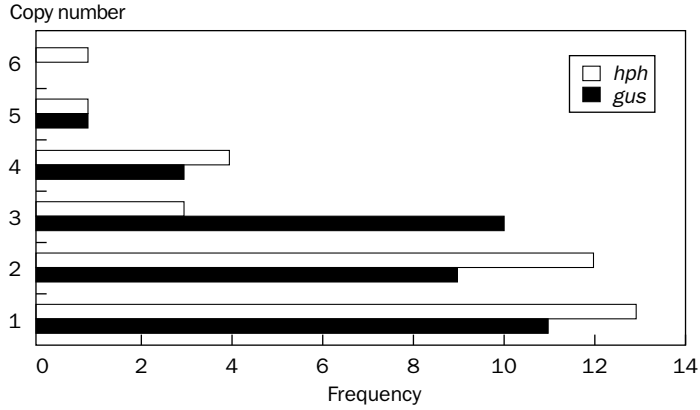


Fig. 2. Estimation of T-DNA copy numbers by DNA gel-blot analyses. Thirty-four transgenic lines were analyzed with *gus* or *hph* probes.

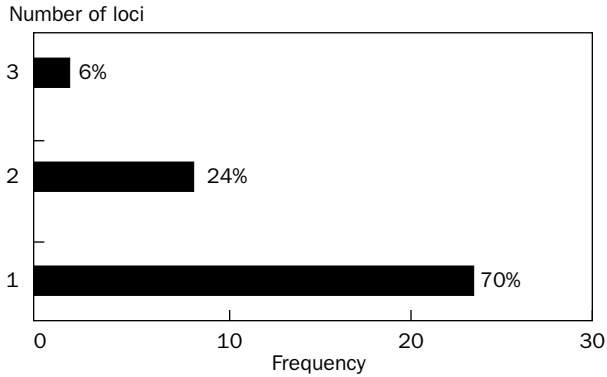


Fig. 3. Estimation of T-DNA insertion loci. The number of genetic loci was estimated by scoring hygromycin-resistant progeny.

single locus. The result revealed that two lines did not contain any DNA sequences between the T-DNAs (Fig. 1B). The remaining four lines carried 6 to 488 bp of filler DNA. Interestingly, the 488 bp of the longest filler DNA in the B1558 line was found to be a portion of the *gus* gene. A DNA gel-blot analysis confirmed that the B1558 line had one more copy of *gus* than *hph*. Such a partial T-DNA was previously reported from dicots such as tobacco (Krizkova and Hrouda 1998). It appears that the formation of repeated T-DNA copies might result from co-integration of several intermediates into one target site.

It has previously been reported that a majority of the T-DNA insertions occur within the right border at a specific locus (reviewed in Tinland 1996). To examine whether the same was true for our tagging lines, the junction regions between rice genomic DNA and the T-DNA right border were sequenced (Fig. 1C). The sequencing results revealed that the boundaries in most of the rice lines did not correspond to the T-DNA nicking position found in *Arabidopsis* and tobacco transgenic plants. In dicot species, most T-DNAs were nicked after the first or second base of the right border. In our tagging lines, five showed nicking positions similar to those of *Arabidopsis* and tobacco. However, the most frequent junction point (11 out of 32 lines) was after the third base of the right border. In seven lines, the junction was at the boundary between T-DNA and the right border. The remaining nine lines showed deletion of one to 12 bases of T-DNA. It was previously reported that two of three right boundaries in transgenic rice plants and four of ten in transgenic maize plants carried three bases originated from the right border (Hiei et al 1994, Ishida et al 1996).

Expression of the *gus* gene in transgenic rice plants

To evaluate the efficiency of the gene trap system, the *gus* expression pattern was examined from various organs of primary transgenic plants transformed with pGA2144. We have analyzed GUS activity in the leaves and roots from 5,353 lines, in mature flowers from 7,026 lines, and in developing seeds from 1,948 lines. The results re-

vealed that the efficiency of GUS staining was 2.0% (106 of 5,353) for leaves, 2.1% (113 of 5,353) for roots, 1.9% (133 of 7,026) for flowers, and 1.6% (31 of 1,948) for immature seeds (Fig. 4). Among the 106 GUS-positive lines in leaves, 15 (14.2%) were leaf-specific. Likewise, 25 (22.1%) lines were root-specific among the 113 GUS-positive lines in roots. We also obtained data indicating that the efficiency of *gus* expression in pGA1633 lines was 1.1% (8 of 750) for leaves and 0.9% (7 of 750) for roots. These values are lower than those of pGA2144, indicating that the modified *OsTubA4* intron increased *gus* tagging efficiency.

The staining patterns of the 106 lines that showed GUS activity in leaves were observed in detail (Table 1). The vein-preferential GUS staining pattern was the most frequently observed (43.4%) and 14 (13.2%) lines were stained preferentially in mesophyll cells between veins. In most samples, GUS staining was observed strongly in boundary regions exposed by cutting. A high concentration of cellulose, lignin, silica cells, and wax in rice leaves could have obstructed penetration of the GUS substrates. Table 2 summarizes the GUS staining patterns in roots. A majority of the lines showed

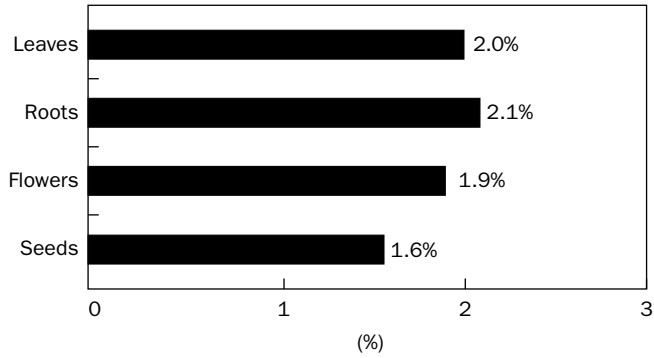


Fig. 4. The frequency of GUS expression in various organs of transgenic plants. GUS activities were analyzed from leaves and roots of 5,353 seedlings, mature flowers of 7,026 transgenic plants, and developing seeds of 1,948 plants.

Table 1. GUS assay in leaves of transgenic rice plants[a].

GUS staining patterns	GUS-positive	
	Lines	%
Mesophyll	14	13.2
Spots	14	13.2
Veins	46	43.4
Veins, mesophyll	28	26.4
Veins, spots	2	1.9
Mesophyll, spots	2	1.9
Total	106	100

[a]5,353 lines were tested.

Table 2. GUS assay in roots of transgenic rice plants.

GUS staining patterns	GUS-positive	
	Lines (no.)	%
GUS activity in various regions[a]		
Cap	1	0.9
Differentiation	36	31.9
Cap, division	4	3.5
Cap, differentiation	3	2.7
Division, differentiation	2	1.8
Elongation, differentiation	7	6.2
Differentiation, spots	3	2.7
Cap, division, elongation	2	1.8
Cap, division, differentiation	2	1.8
Division, elongation, differentiation	16	14.2
Cap, division, elongation, differentiation	37	32.7
Total	113	100
GUS activity in various tissues		
Epidermis	3	4.1
Vein	14	18.9
Epidermis, vein	1	1.4
Exodermis, endodermis, vein	4	5.4
Exodermis, cortex, endodermis	11	14.9
Exodermis, cortex, endodermis, vein	34	45.9
Epidermis, exodermis, cortex, endodermis, vein	7	9.4
Total	74	100

[a]5,353 lines were tested.

GUS staining in the area of cell differentiation and more than half of the lines exhibited GUS activity in the area of cell elongation or cell division. We also characterized the GUS staining patterns in transgenic flowers (Table 3). Among the 133 lines that showed GUS activity in flowers, 50 lines (37.6%) displayed intense GUS staining primarily in the palea and lemma. One line exhibited GUS activity only in glumes, eight lines showed GUS activity only in lodicules, and four lines only in a carpel. Of the 11 lines exhibiting stamen-specific GUS activity, seven showed pollen-specific GUS staining. Five to ten days after flowering, the developing seeds were also subjected to GUS staining. A large portion of these lines showed a tissue-preferential expression pattern. For example, line G930726 exhibited an aleurone-layer preferential GUS staining pattern, indicating that the trapped gene might be involved in the formation of the aleurone layer or in a specific function in the tissue.

Discussion

A variety of GUS staining patterns was observed from the tagged lines. Some were tissue- or organ-specific and others were expressed ubiquitously. This observation supports T-DNA insertion as a random event. The flanking sequences of the GUS-positive lines are being isolated to obtain the genes that provided GUS expression.

Table 3. GUS assay in flowers of transgenic rice plants[a].

GUS staining patterns	GUS-positive	
	Lines (no.)	%
Palea/lemma	50	37.6
Lodicules	8	6.0
Stamens	11	8.3
Carpel	4	3.0
Palea/lemma, lodicules	9	6.8
Palea/lemma, stamens	8	6.0
Palea/lemma, carpel	5	3.8
Lodicules, stamens	1	0.8
Lodicules, carpel	3	2.3
Stamen, carpel	9	6.8
Palea/lemma, lodicules, stamens	4	3.0
Palea/lemma, lodicules, carpel	2	1.5
Palea/lemma, stamens, carpel	3	2.3
Lodicules, stamens, carpel	1	0.8
Palea/lemma, lodicules, stamens, carpel	8	6.0
Rachilla	6	4.5
Glumes	1	0.8
Total	133	100

[a]7,026 lines were tested.

The next generation of the tagged lines is studied to determine whether these lines display any mutant phenotypes in the organs where the *gus* gene was activated, and whether the phenotypes cosegregate with the T-DNA.

Insertional lines that exhibit a particular GUS staining pattern should facilitate identifying genes that are regulated spatially and temporally for plant development. The *Arabidopsis LRP1* (*lateral root primordium1*) gene, which may play a role in lateral root development, was identified by expression of promoterless *gus* in tagging plants (Smith and Fedoroff 1995). The *Arabidopsis PROLIFERA* gene, which is related to the *MCM2-3-5* family of yeast genes, was also cloned by gene trap transposon mutagenesis (Springer et al 1995).

The progeny phenotypes of the pGA1633 transgenic lines were examined. Some of the tagged lines showed mutant phenotypes, including early flowering, tallness, dwarfism, spotted leaves, chlorophyll deficiency, depressed palea, filamentous flowers, extra glumes, long sterile glumes, zebra (transverse green and chlorotic bands in leaves), etc. (data not shown). Whether any of these mutant phenotypes are due to the T-DNA insertion is under investigation.

It is expected that the genome sequence of rice will be completed in the near future. This will produce a large number of genes such that their identification must be regarded as hypothetical. One of the most efficient ways to obtain information on the function of a gene is to create a loss-of-function mutation and study the phenotype of the resulting mutant. If a large population of mutagenized plants is available, it is possible to detect an insertion within the gene of interest by PCR using oligonucleotide primers from the insertional element and the gene of interest (Couteau et al

1999, Krysan et al 1999, Sato et al 1999). The identification of the desired mutant could be accomplished efficiently using a super-pooling strategy as suggested by Krysan et al (1999). They estimated that the maximum useful pool size is 2,350 lines in *Arabidopsis* based on the sensitivity for detecting a specific T-DNA insert and the total amount of template DNA. We are performing experiments to determine the upper size limit on DNA pools of the rice tagged lines.

Our T-DNA tagging lines will be useful in analyzing the function of several valuable genes by various approaches. With the increasing availability of rice genome sequences from public databases, it would be valuable to construct a database for the flanking sequences. The development of effective methods for amplification of sequences adjoining the insertion would enable us to construct the database effectively.

References

Azpiroz-Leehan R, Feldmann KA. 1997. T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. Trends Genet. 13:152-156.

Babiychuk E, Fuangthong M, Van Montagu M, Inze D, Kushnir S. 1997. Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach. Proc. Natl. Acad. Sci. USA 94:12722-12727.

Campisi L, Yang Y, Yi Y, Heilig E, Herman B, Cassista AJ, Allen DW, Xiang H, Jack T. 1999. Generation enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. Plant J. 17:699-707.

Cooley MB, Goldsbrough AP, Still DW, Yoder JI. 1996. Site-selected insertional mutagenesis of tomato with maize *Ac* and *Ds* elements. Mol. Gen. Genet. 252:184-194.

Couteau F, Belzile F, Horlow C, Grandjean O, Vezon D, Doutriaux MP. 1999. Random chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc1* mutant of *Arabidopsis*. Plant Cell 11:1623-1634.

Feldmann KA. 1991. T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. Plant J. 1:71-82.

Frey M, Stettner C, Gierl A. 1998. A general method for gene isolation in tagging approaches: amplification of insertion mutagenized sites (AIMS). Plant J. 13:717-721.

Hiei Y, Komari T, Kubo T. 1997. Transformation of rice mediated by *Agrobacterium tumefaciens*. Plant Mol. Biol. 35:205-218.

Hiei Y, Ohta S, Komari T, Kumashiro T. 1994. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. 6:271-282.

Ishida Y, Saito H, Ohta S, Hiei Y, Komari T, Kumashiro T. 1996. High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. Nat. Biotechnol. 14:745-750.

Jefferson RA, Kavanagh TA, Bevan MW. 1987. GUS fusions: *β-glucuronidase* as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6:3901-3907.

Kertbundit S, De Greve H, Deboeck F, Van Montagu M, Hernalsteens JP. 1991. *In vivo* random *β-glucuronidase* gene fusions in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 88:5212-5216.

Kertbundit S, Linacero R, Rouze P, Galis I, Macas J, Deboeck F, Renckens S, Hernalsteens JP, De Greve H. 1998. Analysis of T-DNA-mediated translational *β-glucuronidase* gene fusions. Plant Mol. Biol. 36:205-217.

Krizkova L, Hrouda M. 1998. Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. Plant J. 16:673-680.
Krysan PJ, Young JC, Sussman MR. 1999. T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11:2283-2290.
Liu YG, Whittier RF. 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25:674-681.
Martin GB. 1998. Gene discovery for crop improvement. Curr. Opin. Biotechnol. 9:220-226.
Ohba T, Yoshioka Y, Machida C, Machida Y. 1995. DNA rearrangement associated with the integration of T-DNA in tobacco: an example for multiple duplications of DNA around the integration target. Plant J. 7:157-164.
Parinov S, Sevugan M, Ye D, Yang WC, Kumaran M, Sundaresan V. 1999. Analysis of flanking sequences from *Dissociation* insertion lines: a database for reverse genetics in *Arabidopsis*. Plant Cell 11:2263-2270.
Sasaki T. 1998. The rice genome project in Japan. Proc. Natl. Acad. Sci. USA 95:2027-2028.
Sato Y, Sentoku N, Miura Y, Hirochika H, Kitano H, Matsuoka M. 1999. Loss-of-function mutations in the rice homeobox gene *OSH15* affect the architecture of internodes resulting in dwarf plants. EMBO J. 18:992-1002.
Smith DL, Fedoroff NV. 1995. *LRP1*, a gene expressed in lateral and adventitious root primordia of *Arabidopsis*. Plant Cell 7:735-745.
Springer PS, McCombie WR, Sundaresan V, Martienssen RA. 1995. Gene trap tagging of *PROLIFERA*, an essential *MCM2-3-5*-like gene in *Arabidopsis*. Science 268:877-880.
Sundaresan V, Springer P, Volpe T, Haward S, Jones JDG, Dean C, Ma H, Martienssen R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev. 9:1797-1810.
Tinland B. 1996. The integration of T-DNA into plant genomes. Trends Plant Sci. 1:178-184.
Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JDG. 1999. Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: a tool for functional genomics. Plant Cell 11:1841-1852.
Topping JF, Wei W, Lindsey K. 1991. Functional tagging of regulatory elements in the plant genome. Development 112:1009-1019.

Notes

Authors'address: National Research Laboratory of Plant Functional Genomics, Division of Molecular and Life Science, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Republic of Korea, Fax +82-562-279-2199, E-mail genean@postech.ac.kr.
Acknowledgments: This work was supported in part by grants from Novartis Crop Protection, Inc., and from the Korea Institute of Science and Technology Evaluation and Planning (KISTEP).
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Transposons and functional genomics in rice

R. Greco, P.B.F. Ouwerkerk, C. Sallaud, A. Kohli, C. Favalli, T. Beguiristain, L. Colombo, E. Pè, P. Puigdomènech, E. Guiderdoni, P. Christou, J.H.C. Hoge, and A. Pereira

Genome and expressed sequence tag sequencing in rice provides a vast resource of gene sequences whose functions need to be determined by reverse genetics methods for expression and mutational analysis. To develop insertional mutagenesis strategies in rice, we transformed japonica and indica cultivars with maize transposon constructs, for knockout and gene detection insertions. A green fluorescent protein (GFP) excision assay developed enabled the visualization of transposon excision in a variety of tissues. Surprisingly, early *Ac* excision was observed directly after transformation from a construct containing the strong double CaMV enhancer element adjacent to the *Ac* promoter. We identified genotypes with *Ac* amplification events and with a forward transposition rate of 15–50% that are useful for generating lines containing multiple transposons. The sequence of DNA flanking transposed *Ac* provided a resource of *Ac*-tagged sites, which represented about 50% in the target region, indicating insertional specificity appropriate for the identification of mutants of sequenced genes. Clustered *Ac* transposition was revealed by six insertions in 70 kb of chromosome 6. Gene detection *Ac-Ds* enhancer trap and activation tag transformants revealed active transposition in about half the lines. These resources for functional genomics are developed by an EU-funded consortium and will be made available to rice researchers worldwide.

Genomic technologies and DNA sequence information have become important in understanding the genetics of plants. This is particularly true for model organisms such as rice, in which the availability of vast amounts of genome information encourages biologists to ask questions in genetics, biochemistry, physiology, and molecular biology in a genomic perspective. *Arabidopsis* genomics and functional genomics have led the way and will provide the blueprint for the structure and function of many plant genes. Some of the remarkable insights emerging from whole-genome analysis (Lin et al 1999, Mayer et al 1999) are the number of gene clusters, large-scale duplications of chromosome segments, as well as a surprisingly high frequency (40%) of newly discovered genes of unknown function.

Functional genomics

The field of functional genomics has emerged to address the function of genes discovered by genome sequencing. In contrast to the previously prevalent gene-by-gene approaches, new high-throughput methods are being developed for expression analysis as well as for the recovery and identification of mutants. The experimental approach is consequently changing from hypothesis-driven to nonbiased data collection and an archiving methodology that makes these data available for analysis by bioinformatics tools. The functional genomics methodology is also changing the experimental strategy from a forward genetics (or mutant to gene) approach to a reverse genetics (or sequenced gene to mutant and function) approach. It is expected that the functional genomics of model plants will contribute to the understanding of basic plant biology as well as the exploitation of genomic information for crop improvement. This is because a large number of gene functions for generic traits will be functional across species, either directly or after identifying the functional homologues.

With its relatively small genome (450 Mb), rice has been selected to be the model monocot plant to be first sequenced. Although the other cereal genomes will not be sequenced completely soon, the synteny between the monocots (Gale and Devos 1998) will help decipher the structure and function of the more complex genomes.

The major developments in whole-genome analysis have been in the field of transcript expression analysis using a variety of high-throughput methods. In plants, different technologies have been employed (Baldwin et al 1999) based on high-density nylon filters, microarrays, various gel systems, and even a sequencing-based method, SAGE, that has been employed for rice (Matsumura et al 1999). With gene chips or microarrays, the representative genes of an organism can be placed on a solid support such as glass and used in hybridization experiments with different RNA samples to reveal gene expression patterns and help identify pathways by association. However, the expression patterns of genes supply correlative information and do not necessarily prove a causal relationship between gene sequence and function. As mRNA level alone does not reflect the actual expression of the gene product, additional means are provided by proteomics (Dove 1999), which addresses the protein expression of a cell type, and metabolomics, which reveals the metabolite profile (Trethewey et al 1999) of the cell/tissue type revealing the active biochemical processes. A combination of different expression analysis tools and mutant or overexpression analysis of genes will be able to provide a unique multidimensional picture of genetic circuits and pathways.

Knockout mutagenesis

The phenotype associated with a gene function is often the best clue to its role in the plant. Phenotype of mutants can be broadly defined at the morphological, biochemical, or physiological level and it provides information on the interactions among different processes. Classically, chemical and physical mutagens have provided loss-of-function mutants and have helped define genes involved in specific pathways or pro-

cesses. To study the interactions among genes, mutants in a particular pathway can be combined and the genetic hierarchy studied. These mutants can be precisely mapped and lead to gene isolation by map-based cloning procedures.

Insertional mutagenesis, with transposable elements or *Agrobacterium*-mediated T-DNA insertions, can generate mutants and also lead directly to gene identification. The insertion often causes a knockout mutation by blocking the expression of the gene and might display a mutant phenotype. The mutant gene tagged by the insertion can be isolated by recovering DNA flanking the insert and subsequently lead to the isolation of the wild-type gene. For forward genetics, T-DNA insertional mutagenesis is only practical on a large scale for *Arabidopsis* (avoiding somaclonal variation) and has resulted in a large number of tagged genes, although its use for reverse genetics in model plants such as rice has a future.

The ability of the well-characterized maize transposon systems *Ac-Ds* and *En-I(Spm)* to transpose in heterologous hosts (Baker et al 1986, Pereira and Saedler 1989) offers new possibilities for transposon tagging, which is an efficient tool for identifying genes. For effective tagging strategies, two component systems comprising a mobile transposon component (*Ds* or *I/dSpm*) and the corresponding stable transposase (*Ac* or *En/Spm*) source have been developed (reviewed in Pereira 1998). The mobile transposon components are often inserted in assayable/selectable marker genes to monitor their excision phenotype (Baker et al 1987) and can also contain other convenient marker/selectable genes to screen for their presence. To control transposition, the transposase is often put under the control of heterologous promoters (Swinburne et al 1992) or segregated out in progeny to yield stable transposon inserts. As transposons move preferentially to closely linked sites (Jones et al 1990) and tag genes efficiently near their original position, efforts are made to map transposons in heterologous systems in order to generate jumping pads all over the genome for efficient targeted mutagenesis.

Ac-Ds transposons were first introduced by electroporation into rice (Izawa et al 1991, Murai et al 1991, Shimamoto et al 1993) and shown to transpose. It was noted, however, that the two-component *Ac-Ds* system often gets inactivated after the first generation (Izawa et al 1997), justifying a systematic investigation of *Ac-Ds* biology in rice. Recently, the behavior of *Ac* has been followed through three successive generations and it has revealed characteristics suitable for functional genomics strategies (Enoki et al 1999). Another knockout mutagenesis tool in rice, an endogenous retrotransposon, *Tos17*, has been developed for reverse genetics screens (Hirochika 1997).

The use of knockout mutations is limited, however, as the majority of genes display no obvious phenotype (Burns et al 1994), probably because of functional redundancy, in which one or more other homologous loci can substitute for the same function. Even *Arabidopsis*, with its simple genome, contains large duplicated segments and redundant genes (Lin et al 1999, Mayer et al 1999). Therefore, the sequential disruption of redundant genes in an individual genotype might ultimately reveal a mutant phenotype and uncover the gene function.

Gene detection

Gene detection strategies have been developed to address the function of genes that do not directly reveal a knockout phenotype. One way is expression detection that can make use of inserts containing reporter gene constructs, such as enhancer traps or gene traps, whose expression depends on transcriptional regulatory sequences of the adjacent host gene. Another way is by creating misexpression mutants, such as activation tags that might reveal a gain-of-function phenotype.

Enhancer detection was developed to detect enhancers in the genome, which are capable of orientation-independent transcriptional activation from a distance, and thus identify genes based on their expression pattern (Skarnes 1990) even though they might not display an obvious mutant phenotype. Enhancer detection constructs contain a reporter gene such as β-glucuronidase (GUS) with a weak or minimal promoter, for example, with a TATA box (transcription initiation signal), situated near the border of the insert and are expressed when integrated adjacent to an enhancer in the genome. Gene trap-type inserts are designed to create fusion transcripts with the target gene. One type used effectively in plants is an exon trap that enables reporter gene fusions to be created at various locations within a gene. By introduction of splice acceptor sites upstream of the reporter gene, transcriptional fusions are created even for insertions in introns, thus increasing the frequency of inserts expressing the reporter gene.

A novel method for efficient selection of stable transpositions, using positive and negative selection markers, has been developed for an *Ac-Ds*-based gene trap and enhancer trap system (Sundaresan et al 1995). Stable transposed *Ds* elements were recovered and revealed GUS expression activity in about 50% of enhancer trap (*DsE*) and 25% of gene trap (*DsG*) inserts (Sundaresan et al 1995). A similar strategy for selection of stable transposed gene trap *DsG* elements has begun in rice (Chin et al 1999) using greenhouse-selectable marker systems.

Misexpression or gain-of-function mutants can be generated by insertion sequences that carry a strong enhancer element near the border and activate the expression of the adjacent genes. This method of "activation tagging" has been validated by an extensive screen of more than 25,000 T-DNA tags (Weigel et al 2000), revealing a 1/1,000 frequency of dominant mutants that were shown to be caused by the presence of an enhancer between 0.4 and 4 kb from the overexpressed gene. A transposon construct variant to isolate dominant gain-of-function alleles employed the CaMV 35S promoter transcribing outward from the *Ds* transposon end and helped identify a semidominant overexpression mutant (Wilson et al 1996). These examples demonstrate the use of activation tagging to generate dominant mutations by over-/misexpressing genes and generate phenotypes for processes or genes not uncovered by knockout mutants.

Generation of rice transposon insertion populations

To develop efficient transposon mutagenesis strategies in rice, we tested several parameters in order to select components for transposon constructs to be transformed.

Constructs were made with the aim of generating populations of transposon inserts for knockout mutagenesis as well as gene detection using enhancer traps and activation tags. Both the maize *Ac-Ds* and *En-I* transposon systems were used and tested for their characteristic advantages. For transformation, primarily *Agrobacterium*-mediated transformation (Hiei et al 1994) was used for japonica cultivars and particle bombardment (Christou et al 1991) for indica cultivars. The analyses of the transformants with different transposon constructs suggest strategies that could be employed for efficient generation of mutants in rice.

Development of a GFP transposon excision assay

Several promoter-GFP (green fluorescent protein) gene constructs were bombarded (particle gun) into suspension-cultured rice cells and the cells subsequently assayed for GFP expression by fluorescence microscopy. The gene constructs used differed in type of GFP (codon usage, protein solubility/folding, protein chromophore, and presence of an intron in the coding region) as well as in type of promoter. From the detected transient GFP expression levels (Table 1), it was concluded that the GFP types sGFP(S65T) and pGFPint(S65T), in combination with the promoter types 35S, d35S, ACT, GOS2, or UBI, would be best suited to serve as a transposon excision marker in

Table 1. Transient expression of promoter-GFP (green fluorescent protein) gene constructs in bombarded (particle gun) suspension-cultured rice cells. Indicated is the type of light (UV, ultraviolet, or blue) required for optimal (max) or suboptimal (min) excitation. The relative expression (emission of green, blue, or yellow light) of the constructs is indicated with minus or plus signs. Promoter abbreviations are 35S = cauliflower mosaic virus 35S; d35S = 35S with double enhancer; UBI = maize ubiquitin; ACT = rice actin; GOS2 = rice constitutive; CMV = (animal) cyto megalo virus. The origin of the GFPs is smG(B)FP (Davis and Vierstra 1998), sGFP(S65T) (Chiu et al 1996), mGFP5 (pCAMBIA vectors), EG(Y)FP (Clontech), and MON30063/MON30049 (Phang et al 1996).

GFP	Promoter	Excitation		Emission green (blue/yellow)
		UV	Blue	
smGFP	35S	Max	Min	–
smGFP	UBI	Max	Min	+/–
smBFP	35S	Max	0	– (blue)
smRS-GFP(S65T)	35S	Min	Max	+/–
mGFP5	35S	Equal		–
mGFP5	UBI	Equal		–
mGFP6(S65T)	35S	0	Max	++
sGFP(S65T)	35S	0	Max	++++
sGFP(S65T)	d35S	0	Max	++++
sGFP(S65T)	ACT	0	Max	++++
sGFP(S65T)	UBI	0	Max	+++
sGFP(S65T)	GOS2	0	Max	++++
pGFP-int S65T(MON30063)	d35S	0	Max	+++++
pGFP-int S65T(MON30049)	d35S	0	Max	++++
EGFP-C3	35S	0	Max	++
EGFP-C3	CMV	0	Max	+
EYFP	d35	0	Max	+++ (yellow)

rice. Transgenic plants were also generated and tested with the sGFP(S65T) under control of the d35S, ACT, UBI, and GOS2 promoters in comparison with the MON30063-GFP construct and confirmed the bombardment experiments.

A construct termed d35S-sGFP:Ac (Table 2), harboring an autonomous *Ac* element between the d35S promoter and the sGFP(S65T) gene, was introduced into rice via *Agrobacterium*. The obtained transgenic calli displayed easily detectable fluorescent GFP sectors (Greco et al, submitted), indicative of *Ac* excision. The GFP excision assay allowed the identification of excision at various stages during the transformation process and revealed a high excision rate. As GFP expression is easily visible in the seed, this phenotypic excision assay is potentially suitable for selecting independent germinal excision events before planting.

Ac transposition behavior in rice

The transformants with the autonomous *Ac* transposon were analyzed by PCR and DNA blot hybridization for molecular evidence of excision and reinsertion. Table 2 summarizes the results of the transformation experiments and molecular analyses of individual regenerants.

The most remarkable discovery was that, in transformants with the d35S-sGFP:Ac construct with an enhancer adjacent to the *Ac* promoter (Fig. 1), *Ac* transposed directly after transformation in the plant cell in every transformant. The evidence was obtained by DNA sequence analysis of excision and reinsertion alleles from about 40 regenerated plants (Greco et al, submitted). This has never been shown before and probably occurs because the strong CaMV 35S enhancer adjacent to the *Ac* promoter

Table 2. Molecular analysis of transposon construct transformants in rice.

Transposon	Construct	Variety	Independent lines analyzed[a]	Lines w/ entire T-DNA[a]	Independent active lines[a]	% active lines[b]
Autonomous Ac	35S-smGFP:Ac	Taipei 309	1 (9)	1 (9)	1 (9)	100
	d35S-sGFP: Ac	Taipei 309	11 (45)	11 (45)	11 (33)	100
		Nipponbare	8 (52)	7 (45)	6 (32)	86
Total			20 (106)	19 (99)	18 (74)	–
Direct DNA transfer	UBI-smGFP:Ac	Indica cvs.	348 –	– –	278 –	80
Ac total			368 (454)	– –	296 (352)	–
2-component Ac-Ds	2-comp. Ac-Ds	Taipei 309	3 (19)	1 (13)	1 (13)	100
Enhancer trap	Ac-Ds ET	Taipei 309	11 (45)	10 (37)	5 (11)	50
	Ac-Ds ET	Nipponbare	10 (31)	8 (27)	4 (14)	50
	Ac-Ds ET with GFP	Nipponbare	219 (278)	104 (119)	63 (72)	60
ET total			240 (354)	122 (183)	72 (97)	–
Activation tag	Ac-Ds AT	Nipponbare	9 (15)	5 (7)	3 (3)	60
	Ac-Ds AT with GFP	Nipponbare	42 (75)	14 (17)	4 (4)	29
AT total			51 (90)	19 (24)	7 (7)	–
Ac transposase	35s-AcTPase+ SU1	Taipei 309	15 (15)	12 (12)	–	–
		Nipponbare	12 (15)	11 (13)	–	–

[a]Total number of plants in parentheses. [b]The percentage is calculated on the number of independent lines that were not rearranged.

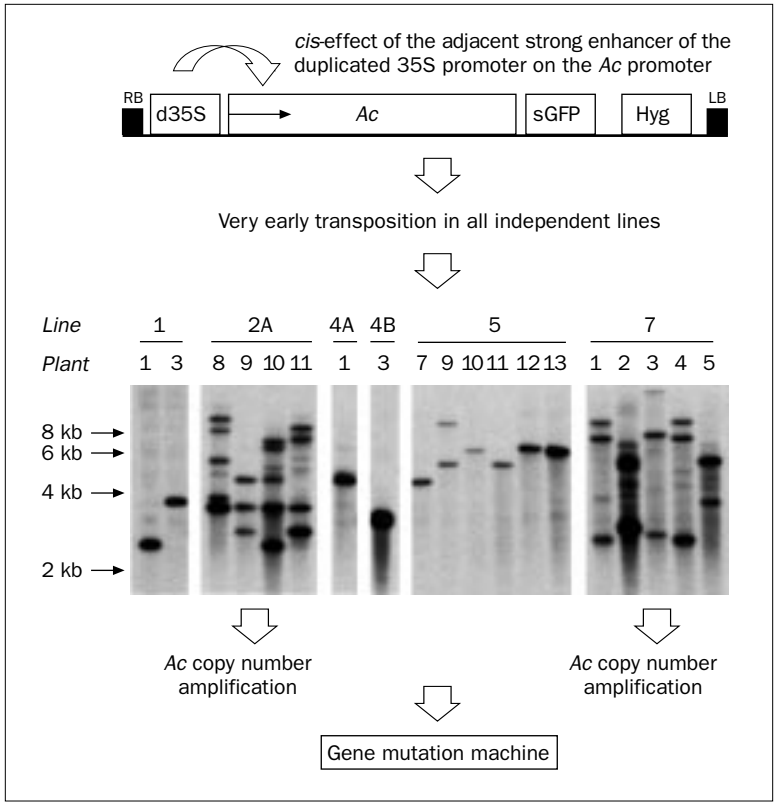


Fig. 1. Early transposition and amplification of autonomous *Ac*. The d35S-sGFP::Ac construct with a duplicate CaMV 35S enhancer-promoter adjacent to the *Ac* promoter (shown by arrow). On excision, the sGFP expression can be visualized by green fluorescent protein (GFP) fluorescence. Shown below is DNA blot hybridization of individual regenerants from six different transformed lines (as labeled), hybridized with an *Ac*-specific probe and fragment sizes given on left. All lines showed very early transposition and the individual regenerants (e.g., 8, 9, 10, 11) from line 2A (single T-DNA insert) display active transposition and amplification. Multiple *Ac* lines displaying independent transposition are used to make a mutation machine for reverse genetics. RB = right border, LB = left border.

influences transposase expression. The *cis* effect of the CaMV enhancer on the *Ac* promoter and transposase expression was observed earlier in *Arabidopsis* (Balcells and Coupland 1994), leading to earlier excision of *Ds*.

Ac reinsertion was evident in about 70% of the lines and one-third of these also showed an amplification in *Ac* copy number as demonstrated by the number of *Ac*-hybridizing fragments in the one- and two-copy *Ac* lines (lines 2A and 7 in Fig. 1). The process of transposition during replication has been described in maize (Greenblatt 1984), in which *Ac* transposes from replicated DNA to unreplicated DNA, resulting in a moderate increase in copy number. In the amplifications we observed, the pres-

ence of a single excision footprint in all the regenerants from a single T-DNA insertion (Greco et al, submitted) suggests that transposition must have begun from an unreplicated site. The mechanism of *Ac* amplification we propose is outlined in Figure 2 and is potentially useful for generating genotypes with multiple transposons (Fig. 1) that can reach genome saturation quickly.

To compare the transpositional activity between transformants, a parameter termed "independent transposition frequency" (ITF) given by the frequency of unique transposition events among the total inserts in a set of related plants (regenerants or seed progeny) was estimated. The somatic ITF calculated over three populations of regenerants varies between 50% and 65%. Transposed *Ac* insertions that were present in only a single plant were considered as unique late events that were useful for producing different insertions in the progeny.

In contrast to the active transposition in regenerating calli, leaf samples of mature plants revealed a stable transposon insertion pattern. The individual progeny again revealed new insertions with an ITF between 15% and 50% in different lines, suggesting that transposition took place prior to gamete formation. This indicates a biphasic transposition pattern: an active phase during regeneration from the callus as well as the pregametic cells and a stable phase in the mature plant.

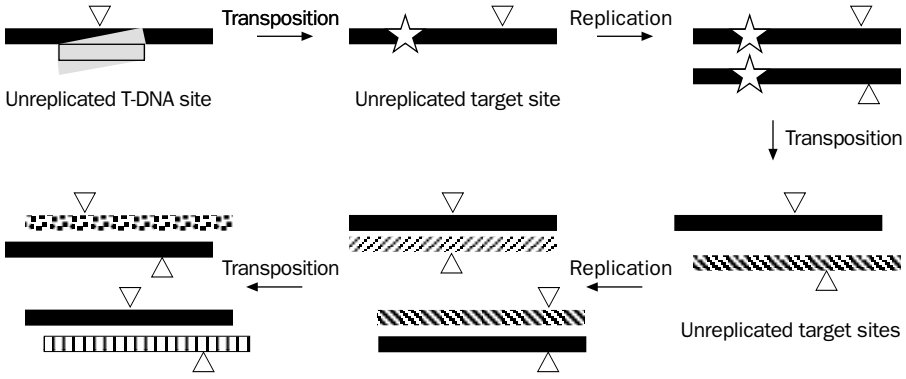


Fig. 2. *Ac* amplification during replication in one cell. The mechanism of *Ac* amplification proposed to explain the results of *Ac* (*Ac* designated as triangle) transposition observed. The high transposase activity mediated by the enhancer adjacent to the *Ac* promoter in the initial cell after transformation and before further cell division triggers an early excision event from the unreplicated T-DNA site (shaded box) that is revealed by the occurrence of the same excision footprint (open star on chromosome site) in every regenerant. In the same cell cycle, the excised *Ac* reinserts into a second site, which then replicates to generate a second *Ac* copy in the newly replicated strand. In the continued presence of the high *Ac* transposase activity in the same cell, secondary transposition events of both *Ac* copies result in two new unreplicated positions that get duplicated following replication, giving rise to four copies at the two chromosomal positions. The high transposase activity mediates a third cycle of transposition and replication that produces four new *Ac* chromosomal positions (denoted by different shaded lines) in the cell.

Ac knockout mutagenesis

The *Ac* lines we characterized in rice revealed multiple transpositions due to amplifications with more than four inserts per plant that could generate an average of one to two new inserts per progeny. The propagation of these genotypes for three to four generations can generate a population of plants containing four or more *Ac* inserts at different positions in the genome. Using 25,000 lines for three to four generations would generate about 100,000 insertions that are suitable for identifying knockouts for forward as well as reverse genetic strategies.

To analyze the transposon insertion sites in these lines, we isolated and sequenced DNA flanking transposed *Ac* elements and compared these *Ac* insertion tagged site (ITS) sequences with those in public databases. We could position six ITSs in a 70-kb interval of sequenced DNA on chromosome 6 (Fig. 3), demonstrating linked transposition that is useful for targeted tagging. The majority of the *Ac* inserts are oriented in the same way (thin arrows in Fig. 3), suggesting that the transposition process is not random. Most significantly, as this chromosome is being systematically sequenced, the *Ac* lines we have generated will be very useful for obtaining knockout mutants of genes identified in this region.

The ITS sequences revealed a high proportion with homology to predicted genes in the databases. About 29% of the ITSs reveal homology to known proteins and 10% to sequenced ESTs (Greco et al, submitted). To calculate the genome sequence predicted to code for proteins, we assume 30,000 rice genes with an average 2.5-kb coding sequence and 55% predicted genes with similarity to proteins of known function based on *Arabidopsis* estimates (Lin et al 1999, Mayer et al 1999). This transforms to 41.5 Mb (30,000 ¥ 2.5 ¥ 55/100) or 10% of the genome that should show

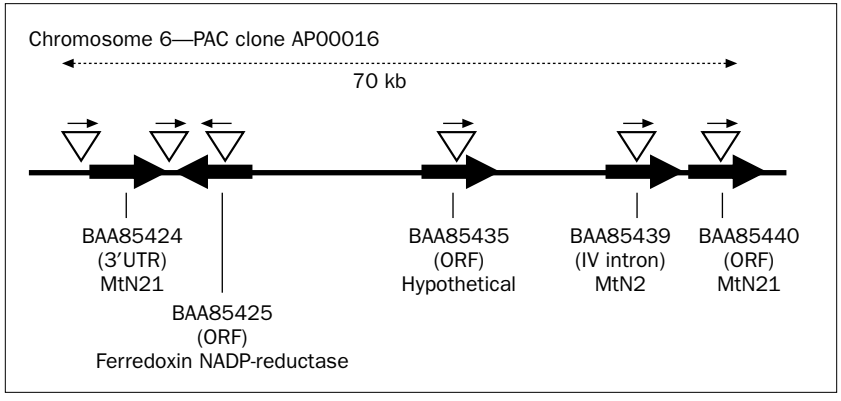


Fig. 3. Local transposition of *Ac* on chromosome 6. The isolation and sequencing of DNA flanking *Ac* inserts in a rice *Ac* line enabled the identification of several tagged sites in a sequenced P1-derived artificial chromosome (PAC) clone from rice chromosome 6. The *Ac* insertions are shown as triangles and the orientation of the *Ac* element in the genes is shown by the thin arrows above the *Ac* inserts. The predicted genes containing insertions are denoted by thick solid arrows and are labeled underneath by accession number, position of insert in the gene, and similarity to known/predicted proteins. ORF = open reading frame, UTR = untranslated region.

homology to known proteins. The transposon inserts are therefore biased about three times higher than expected on a random basis to insert in sequences predicted to code for proteins. To calculate the genome sequence represented by ESTs, with 20,000 different ESTs based on the gene index calculation (Quackenbush et al 2000) and an average size of 400-bp sequence, 8 Mb of unique EST sequence or 1.86% (8/430) of the genome sequence is present in the public databases. We observe 9.6% of the total ITSs with identity to ESTs, meaning that the ITSs are biased for insertions in transcribed genes five times more than randomly expected.

These two calculations reveal a bias of three to five times more insertions in genes and strongly suggest that *Ac* inserts preferentially into genes in rice. This confirms the earlier results of *Ac* insertional preference in rice, in which 4% of the inserts in ESTs were observed (Enoki et al 1999), and suggests that the preferential transposon insertions in genes could be a valuable asset for generating mutants in rice.

Gene detection strategies in rice

Several constructs for gene detection were developed in rice. The general structure of the enhancer trap (ET) and activation tag (AT) constructs is outlined in Figure 4 and

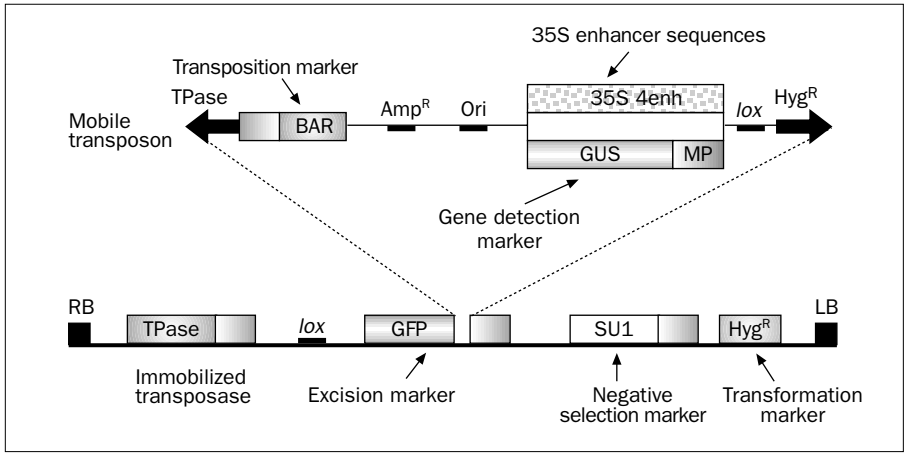


Fig. 4. Model gene detection rice construct. The construct has two major parts: the mobile transposon (*Ds* or I) drawn above and inserted in the resident T-DNA drawn below. The mobile transposon, with transposon borders denoted by outward pointing thick solid arrows, contains a transposition marker (BAR gene) to select transposants as Basta^R (resistant). Plasmid origin and bacterial selection markers help in the recovery of large fragments of genomic-flanking DNA in *E. coli*. The gene detection marker for an enhancer trap consists of a GUS gene with a CaMV 35S minimal promoter (MP) located very near the transposon border. This is replaced by a multiple 35S enhancer (4enh) sequence to create an activation tag construct. To create chromosomal deletions, *lox* sites (*Cre-lox* system) are located in the mobile transposon as well as in the resident T-DNA. The T-DNA contains the immobilized transposase (*Ac* or *En*) under control of a strong promoter. The mobile transposon is inserted in the excision marker green fluorescent protein (GFP) gene to select transpositions. The transformation marker Hyg^R is used for selection of transformants using hygromycin and a negative selection marker, the SU1 gene, used to select for stable transposants. LB = left border, RB = right border, Ori = origin, Amp^R = ampicillin resistance.

Table 2 summarizes the transformation experiments. The *Ac-Ds* construct contains the mobile transposon that can be a *Ds* or an *I* transposon and the corresponding immobile transposase source *Ac* or *En* under control of a strong promoter. To observe transposition early after transformation and help make a choice among transformants, the GFP excision marker can be used to monitor excision. Independent excision events can be selected at the seed or germinated seedling level so that germinal transposants are recovered and grown. To be able to select stable transposed elements, we employed a BAR gene on the *Ds* element conferring resistance to the herbicide Basta ($Basta^R$) and a negative selection marker SU1 (O'Keefe et al 1994) that converts the proherbicide R7402 into the active herbicide and results in shorter plants (Koprek et al 1999). Using a combination of these greenhouse/field selectable markers, progeny of single T-DNA locus transformants can be used to identify stable transposants (BAR^+ $SU1^-$) (Fig. 5), where the *Ds*-BAR transposes from the T-DNA and is segregated away (to another chromosome or after recombination from a linked site).

The enhancer trap construct contains a minimal promoter that can initiate transcription upstream of the GUS marker gene. On insertion near enhancers of host genes in the genome, the GUS gene detector can display the expression pattern of the adjacent gene. This will help identify the adjacent plant gene on the basis of its expression. In plants such as *Arabidopsis*, about 50% of the inserts display expression that is similarly observed in rice. The activation tag construct has a multiple enhancer of the CaMV 35S promoter inserted near the transposon end and can activate adjacent genes. Enhancer trap and activation tag constructs have been transformed and molecular data generated in rice (Table 1). They yield about 50% of the active lines that are being propagated for developing tagging strategies. Some lines also contain multiple *Ds* elements with active transpositions that are useful for generating multiple transposon lines.

Insertional mutagenesis in rice genomics

The rice genome sequence will probably uncover about 30,000 genes, half of which will have no known function. Transposon mutagenesis, using knockout and gene detection insertions, is an important tool for discovering these gene functions by reverse genetics strategies. In *Arabidopsis*, about 100,000 random inserts (Krysan et al 1999) are required for genome saturation and for rice about four times that number. Multiple independent inserts per plant, averaging four in many of the *Ac* and *Ds* lines, will decrease the required number of plants. The insertional preference of *Ac* for genes as described here can reduce the required number further by a factor of 3–5. A population of about 25,000 *Ac* or *Ds* plants, with multiple inserts, would therefore be sufficient to recover knockout or gene detection inserts for most genes.

Transposon populations for single stable *Ds* insertions can also be generated from a minimum of 10 active lines, as shown in the strategy outlined in Figure 5, in four generations after transformation and seed multiplication. This could produce a population of around 150,000 inserts, sufficient for genome saturation assuming preferential transposition of *Ds* in genes. By a concerted international effort, this transposon library is being produced and will be made available to rice researchers worldwide.

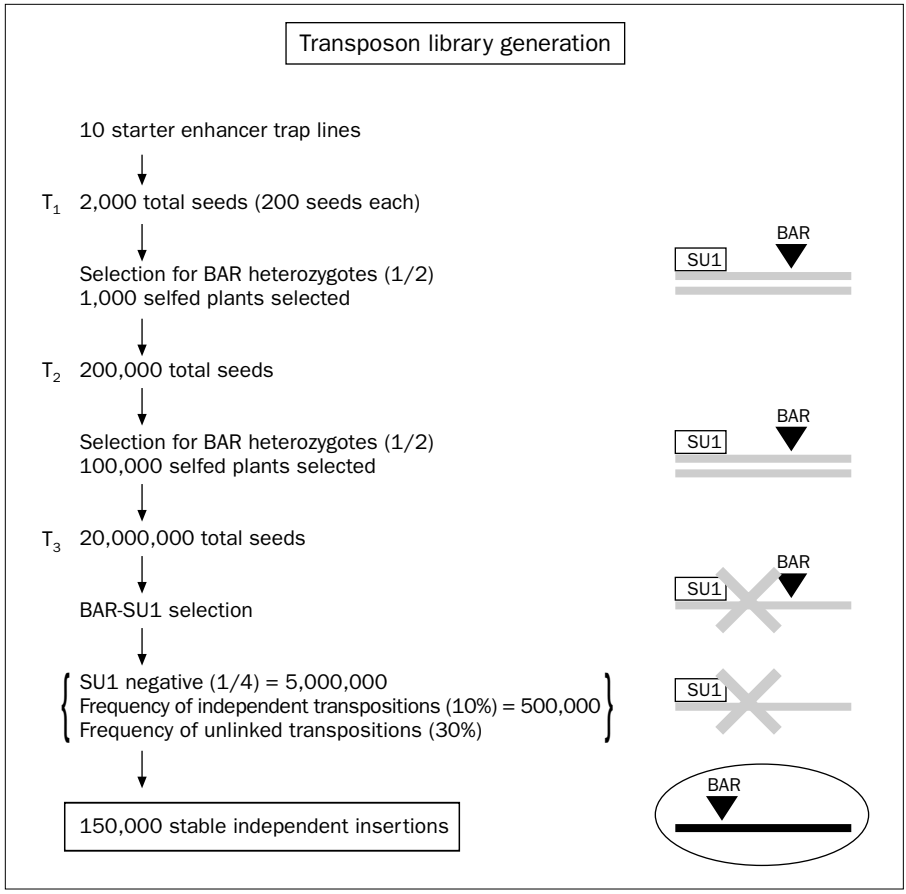


Fig. 5. Transposon library generation. Using a minimum of 10 active single-locus starter enhancer trap lines, with a multiplication rate of 200 seeds per plant (Nipponbare), a total of 200,000 T_2 seeds can be obtained from heterozygous plants that are identified on the basis of progeny-test segregation for Basta^R. Similarly, from the 200,000 plants, half are heterozygotes and are identified by progeny tests in the T_3 generation, in which 20 million seeds are produced from the heterozygous parents and can be used for selection of the population of stable transposants. A sketch of the T-DNA carrying the SU1 and BAR (on transposon) markers is shown alongside. Using Basta^R and SU1 selection, only the plants bearing a Basta^R transposed element (BAR$^+$ SU1$^-$) are selected, whereas all SU1-bearing plants (with transposase) are rejected. The frequency of independent transposition was observed to be 10% and that of unlinked transposition about 30%, which will generate about 150,000 stable independent insertions.

References

Baldwin D, Crane V, Rice D. 1999. A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr. Opin. Plant Biol. 2:96-103.

Baker B, Schell J, Lörz H, Fedoroff N. 1986. Transposition of the maize controlling element 'Activator' in tobacco. Proc. Natl. Acad. Sci. USA 83:4844-4848.

Baker B, Coupland G, Fedoroff N, Starlinger P, Schell J. 1987. Phenotypic assay for excision of the maize controlling element *Ac* in tobacco. EMBO J. 6:1547-1554.

Balcells L, Coupland G. 1994. The presence of enhancers adjacent to the *Ac* promoter increases the abundance of transposase mRNA and alters the timing of *Ds* excision in *Arabidopsis*. Plant Mol. Biol. 24:789-798.

Burns N, Grimwade R, Macdonald PB, Choi E-Y, Finberg K, Roeder GS, Snyder M. 1994. Large-scale analysis of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*. Genes Dev. 8:1087-1105.

Chin HG, Choe MS, Lee S-H, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC, Choi HC, Cho MJ, Han CD. 1999. Molecular analysis of rice plants harboring an *Ac*/*Ds* transposable element-mediated gene trapping system. Plant J. 19:615-623.

Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J. 1996. Engineered GFP as a vital reporter in plants. Curr. Biol. 6:325-330.

Christou O, Ford T, Kofron M. 1991. Production of transgenic rice (*Oryza sativa* L.) from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. Bio/Technology 9:957-962.

Davis SJ, Vierstra RD. 1998. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol. Biol. 36:521-528.

Dove A. 1999. Proteomics: translating genomics into products. Nat. Biotechnol. 17:233-236.

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J, Shimamoto K. 1999. *Ac* as a tool for the functional genomics of rice. Plant J. 19:605-613.

Gale MD, Devos KM. 1998. Plant comparative genetics after 10 years. Science 282:656-659.

Greenblatt IM. 1984. A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element Modulator, in maize. Genetics 108:471-485.

Hiei Y, Ohta S, Komari T, Kumashiro T. 1994. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. 6:271-282.

Hirochika H. 1997. Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol. Biol. 35:231-240.

Izawa T, Miyazaki C, Yamamoto M, Terada R, Iida S, Shimamoto K. 1991. Introduction and transposition of the maize transposable element *Ac* in rice (*Oryza sativa* L.). Mol. Gen. Genet. 227:391-396.

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miyazaki C, Endo T, Iida S, Shimamoto K. 1997. Transposon tagging in rice. Plant Mol. Biol. 35:219-229.

Jones JDG, Carland FM, Lin E, Ralston E, Dooner HK. 1990. Preferential transposition of the maize element *Activator* to linked chromosomal locations in tobacco. Plant Cell 2:701-707.

Koprek T, McElroy D, Louwerse J, Williams-Carrier R, Lemaux PG. 1999. Negative selection systems for transgenic barley (*Hordeum vulgare* L.): comparison of bacterial codA- and cytochrome P450 gene-mediated selection. Plant J. 19:719-726.

Krysan PJ, Young JC, Sussman MR. 1999. T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11:2283-2290.

Lin X and 36 others. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761-768.

Matsumura H, Nirasawa S, Terauchi R. 1999. Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). Plant J. 20:719-726.

Mayer K, Schuller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terryn N, Harris B, Ansorge W, Brandt P, Grivell L, Rieger M, Weichselgartner M, de Simone V, Obermaier B, Mache R, Muller M, Kreis M, Delseny M, Puigdomenech P, Watson M, McCombie WR et al. 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature 402:769-777.

Murai N, Li ZJ, Kawagoe Y, Hayashimoto A. 1991. Transposition of the maize *Activator* element in transgenic rice plants. Nucleic Acids Res. 19:617-622.

O'Keefe DP, Tepperman JM, Dean C, Leto KJ, Erbes DL, Odell JT. 1994. Plant expression of a bacterial cytochrome P450 that catalyzes activation of a sulfonylurea pro-herbicide. Plant Physiol. 105:473-482.

Pereira A. 1998. Heterologous transposon tagging systems. In: Lindsey K, editor. Transgenic plant research. Amsterdam (Netherlands): Harwood Academic Publishers. p 91-108.

Pereira A, Saedler H. 1989. Transpositional behaviour of the maize *En/Spm* element in transgenic tobacco. EMBO J. 8:1315-1321.

Phang S-Z, DeBoer DL, Wan Y, Ye G, Layton JG, Neher MK, Armstrong CL, Fry JE, Hinchee MAW, Fromm ME. 1996. An improved green fluorescent protein gene as a vital marker in plants. Plant Physiol. 112:893-900.

Quackenbush J, Liang F, Holt I, Pertea G, Upton J. 2000. The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. 28:141-145.

Shimamoto K, Miyazaki C, Hashimoto H, Izawa T, Itoh K, Terada R, Inagaki Y, Iida S. 1993. *Trans*-activation and stable integration of the maize transposable element *Ds* cotransfected with the *Ac* transposase gene in transgenic rice plants. Mol. Gen. Genet. 239:354-360.

Skarnes WC. 1990. Entrapment vectors: a new tool for mammalian genetics. Bio/Technology 8:827-831.

Sundaresan V, Springer P, Volpe T, Haward S, Jones JDG, Dean C, Ma H, Martienssen R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev. 9:1797-1810.

Swinburne J, Balcells L, Scofield SR, Jones JDG, Coupland G. 1992. Elevated levels of *Activator* transposase mRNA are associated with high frequencies of *Dissociation* excision in *Arabidopsis*. Plant Cell 4:583-595.

Trethewey RN, Krotzky AJ, Willmitzer L. 1999. Metabolic profiling: a Rosetta Stone for genomics. Curr. Opin. Plant Biol. 2:83-85.

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dixon RA, Harrison MJ, Lamb CJ, Yanofsky MF, Chory J. 2000. Activation tagging in *Arabidopsis*. Plant Physiol. 122:1003-1013.

Wilson K, Long D, Swinburne J, Coupland G. 1996. A *Dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*. Plant Cell 8:659-671.

Notes

Authors' addresses: R. Greco, A. Pereira, Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands; P.B.F. Ouwerkerk, J.H.C. Hoge, Institute of Molecular Plant Sciences, Leiden University, P.O. Box 9505, 2300 RA Leiden, The Netherlands; C. Sallaud, E. Guiderdoni, CIRAD-AMIS, BP 5035, 34032 Montpellier, France; A. Kohli, P. Christou, John Innes Centre, Norwich NR4 7UH, UK; C. Favalli, L. Colombo, E. Pè, University of Milan, Via Celoria 26, 20133 Milan, Italy; T. Beguiristain, P. Puigdomènech, CID-CSIC, Jordi Girona 18, 08034 Barcelona, Spain, E-mail: A.Pereira@plant.wag-ur.nl.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Retrotransposons of rice as a tool for the functional analysis of genes

H. Hirochika, A. Miyao, M. Yamazaki, S. Takeda, K. Abe, R. Hirochika, G.K. Agrawal, T. Watanabe, K. Sugimoto, T. Sasaki, K. Murata, K. Tanaka, K. Onosato, A. Miyazaki, Y. Yamashita, and N. Kojima

Five endogenous active retrotransposons have been found in rice. Among them, the most active one, called *Tos17*, was characterized in detail. *Tos17* is silent under normal conditions and becomes active only under tissue culture conditions. Five to 30 transposed *Tos17* copies were found in each plant regenerated from culture. *Tos17* was shown to transpose preferentially into low-copy-number, gene-rich regions, indicating that *Tos17* can be used as an efficient insertional mutagen. A collection of 32,000 regenerated rice lines carrying about 256,000 insertions was generated, and these lines are being used for forward and reverse genetic analyses. By using a transposon-tagging strategy, causative genes for viviparous, dwarf, semidwarf, brittle culm, pale green, and narrow leaf mutations, among others, have been cloned. For reverse genetic studies, two strategies are being employed. One is the polymerase chain reaction (PCR) screening of mutants of the gene of interest. We screened 12,000 lines and found mutants of 15 genes, including MAPK, MADS-box, and P450 genes, among the 47 genes analyzed. This suggests that at least 37,000 lines are required for saturation mutagenesis. Another important strategy is the random sequencing of mutated genes by isolating the sequences flanking transposed *Tos17*. The flanking sequences are amplified by TAIL (thermal asymmetric interlaced)- and suppression-PCR and directly sequenced. Until now, 7,376 independent flanking sequences from 2,134 lines have been determined and mutants of different classes of genes have been identified.

Because of an explosion in the number of sequenced genes of rice with unknown function revealed by large-scale analysis of expressed sequence tags (ESTs) (Sasaki et al 1994, Yamamoto and Sasaki 1997) and genome sequencing (Sasaki and Burr 2000), development of a systematic method suitable for discovering the biological functions of these genes becomes extremely important. To determine gene function, gene inactivation is a powerful tool. Among the gene inactivation strategies, insertional mutagenesis seems most suitable for a systematic functional analysis of a large number of genes. In *Arabidopsis*, whose entire genomic sequencing will be completed soon, several insertional mutagens have been used for efficient mutagenesis.

These include T-DNA (Azpiroz-Leehan and Feldmann 1997, Krysan et al 1999) and maize transposable elements *Ac/Ds* (Parinov et al 1999) and *En/Spm* (Speulman et al 1999, Tissier et al 1999). Mutant populations induced by these mutagens are being used for forward and reverse genetics. Since these lines induced by the exogenous insertional mutagens are transgenic plants, it is necessary to grow them under the regulation of recombinant DNA experiments. Because rice plants are relatively large, it is not easy to handle thousands of transgenic rice plants in a greenhouse. Therefore, it is desirable to use endogenous insertional mutagens. *Ac/Ds*-, *En/Spm*-, and *Mu*-like elements (Hirochika and Fukuchi 1992, Motohashi et al 1996, R. Ishikawa and M. Freeling, personal communication) have been identified by molecular analysis in rice. Most of them, however, appear to be inactive due to accumulation of mutations. Consistent with these results, none of the spontaneous waxy mutations were induced by transposable elements (Okagaki and Wessler 1988). This finding does not necessarily mean that active transposable elements do not exist in the rice genome. We have shown that the retrotransposon of rice, named *Tos17*, is highly active only during tissue culture and that the activation of *Tos17* is responsible for tissue-culture-induced mutations (Hirochika et al 1996). Unique features of *Tos17* suggest that it can be used for forward and reverse genetic studies (Hirochika 1997, 1999). Here, we report features of *Tos17* as an insertional mutagen and the current status of forward and reverse genetic studies of rice using *Tos17*.

Isolation of active retrotransposons of rice

Although more than 40 families of long terminal repeats (LTR)-retrotransposons have been found in rice (*Tos1*-*Tos32*: Hirochika et al 1992, 1996, unpublished; *RIRE1*-*RIRE8*: Noma et al 1997, Ohtsubo et al, 1999, H. Ohtsubo et al, personal communication; *Rrt1*-*Rrt23*: Wang et al 1997), only five of them were shown to be active. *Tos10*, *Tos17*, *Tos19*, *Tos25*, and *Tos27*, which were isolated by using polymerase chain reaction (PCR) or reverse transcript (RT)-PCR, are silent during normal growth conditions and become active only during tissue culture (Hirochika et al 1996, Hirochika et al, unpublished). The PCR method is based on the conservation of certain amino acid sequences in the reverse transcriptase domain among *Ty1-copia* group retrotransposons. Four out of the five active retrotransposons were isolated by using RT-PCR, indicating that this method is very effective for isolating active retrotransposons. This method is based on the fact that most inactive plant retrotransposons are not transcribed and active retrotransposons are transcribed only under certain stress conditions (Hirochika 1993). Among the five active retrotransposons, *Tos17* has been shown to be the most active (Fig. 1, Hirochika et al 1996, Hirochika, unpublished). As the transcript of *Tos17* was only detected under tissue culture conditions, the transposition of *Tos17* must be regulated mainly at the transcriptional level as in the case of tobacco retrotransposons such as *Tnt1* (Pouteau et al 1991, Grandbastien 1998) and *Tto1* (Hirochika 1993, Takeda et al 1998). To compare the activities of *Tos17* and *Tos19* under the same culture conditions, the DNA gel blot used in Figure 1 was rehybridized with a *Tos19* probe. An increase in copy number ranging from one to three was found only in those plants regenerated from 16-mo-old culture (data not shown), indicating

that *Tos17* is the major insertional mutagen among the rice retrotransposons examined.

As shown in Figure 2, *Tos17* is 4,114 bp long (Okamoto and Hirochika, unpublished) and carries one open reading frame (ORF) corresponding to Gag and Pol

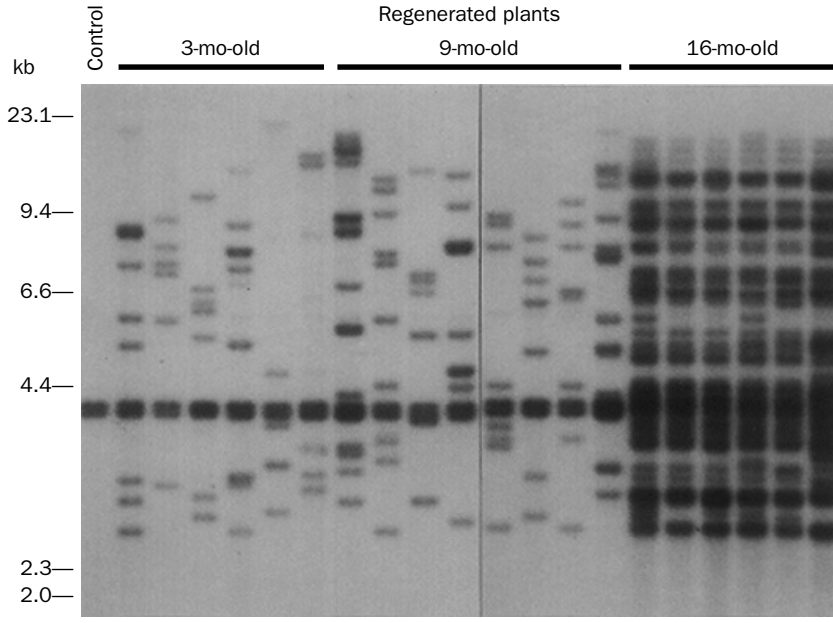


Fig. 1. Activation of *Tos17* by tissue culture. DNAs were prepared from leaves of normally propagated rice plants (control) and plants regenerated from 3-, 9-, and 16-mo-old cultures and analyzed by DNA blot hybridization with a *Tos17*-specific probe after digestion with *Xba*I. (This figure is adapted from Hirochika et al 1996.)

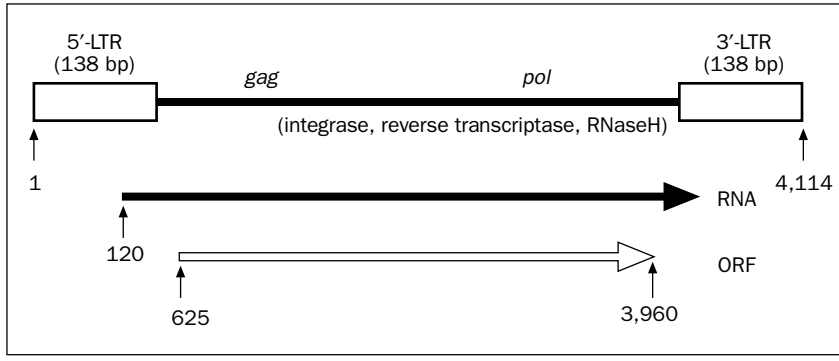


Fig. 2. Structure deduced from the complete nucleotide sequence and transcription of *Tos17*. Long terminal repeats (LTRs) are boxed. The thick arrow indicates the transcript of *Tos17* and the open arrow indicates the longest open reading frame (ORF) starting with the ATG. The *pol* gene encodes integrase, reverse transcriptase, and RNaseH.

polyproteins and two identical 138-bp LTRs. The existence of one uninterrupted ORF suggests that the cloned *Tos17* is an autonomous element.

Features of *Tos17* as an insertional mutagen

The original copy number of *Tos17* is two in the japonica variety Nipponbare and five to 30 transposed *Tos17* copies were detected in all of the plants regenerated from tissue cultures (Fig. 1), including transgenic plants (Hirochika et al 1996). Each plant regenerated from 3- to 9-mo culture showed a different pattern of hybridization with the *Tos17* probe (Fig. 1), indicating that independent transposition events occurred in each cell of the tissue culture. In contrast, the patterns are almost the same in the plants regenerated from 16-mo-old culture. This suggests that a specific cell type was selected during prolonged culture. These results indicate that a proper culture period must be chosen to maintain independent mutations induced in the populations for efficient mutagenesis, although the number of mutations induced by *Tos17* increased with a prolonged culture period.

Considering the frequency of transposition, *Tos17* is most likely to be involved in tissue-culture-induced mutations. Direct evidence that *Tos17* insertions lead to mutations of genes was obtained by analyzing the target sites of *Tos17* transposition (Hirochika et al 1996). Eight target sites were amplified as *Tos17*-flanking sequences by using inverse PCR (IPCR) and sequenced. Regenerated plants with *Tos17*-insertions in the phytochrome A gene and the S-receptor kinase-related gene were identified. This result indicates that activation of *Tos17* is an important cause of tissue-culture-induced mutations. To further confirm the result, more than 100 sequences flanking newly transposed *Tos17* copies were characterized. Only one sequence out of 123 investigated showed a 3-bp deletion at the 3′ end of *Tos17*, indicating that rearrangement of *Tos17* ends during transposition is rarely induced. Forty-two out of a total of 123 flanking sequences showed significant homology to known genes or hypothetical genes, showing at least one-third of transposed *Tos17* integrated into genes. Hybridization studies with amplified flanking sequences indicated that *Tos17* preferentially integrates into low-copy-number regions of the genome. These results indicate that *Tos17* prefers gene-rich, low-copy regions as integration targets. Twenty-nine loci of newly transposed copies were identified by restriction fragment length polymorphism (RFLP) linkage analysis using recombinant inbred lines or by PCR-based physical mapping onto the yeast artificial chromosomes (YACs) and plotted onto the rice genetic map (Harushima et al 1998). Integration target loci were shown to be widely distributed over the chromosomes. Considering this feature and the preferential integration into gene-rich, low-copy regions, saturation mutagenesis with *Tos17* should be feasible for tagging and reverse genetic studies.

To make rice lines mutagenized with *Tos17*, calli were induced from scutella by culturing dehusked mature seeds in a liquid medium for 3 to 5 mo. Five to 15 transposed copies were detected in each plant regenerated from these cultures. A collection of 32,000 regenerated rice lines carrying about 256,000 insertions was generated. Most of the mutagenized lines were derived from variety Nipponbare, which has been chosen as a common resource for genome sequencing (Sasaki and Burr 2000).

Forward and reverse genetic studies using retrotransposons

Features of *Tos17* useful for forward and reverse genetic studies are summarized as follows. (1) Transposition can be regulated since *Tos17* is activated by tissue culture and becomes silent in regenerated plants. (2) Highly mutagenic during tissue culture, *Tos17* transposes preferentially into gene-rich, low-copy regions and about eight loci on average are disrupted in each plant regenerated from 5-mo-old culture. (3) Integration target loci were widely distributed over the chromosomes so that random insertion for saturation mutagenesis is feasible. (4) Induced mutations are stable. (5) The original copy number is quite low, one to four depending on varieties, so that it is easy to identify the transposed copy responsible for the specific mutation. (6) The transposon is endogenous so that screening and characterization of mutants in the field are possible without any regulation of recombinant DNA experiments. This must be important because many agronomically important traits are best characterized in the field.

Retrotransposons as a tool for transposon-tagging

The feasibility of transposon-tagging in rice was suggested by using *Ac* and *Ac/Ds* elements (Izawa et al 1997); however, no clear evidence for tagging has been reported so far. Here, we demonstrate the feasibility of gene-tagging using *Tos17*. We first screened R_1 (M_2) generations of regenerated rice for mutants based on the phenotypes in the paddy field. About 30% of 2,300 lines examined showed many kinds of visible mutant phenotypes, such as dwarf, sterile, yellow, albino, virescent, viviparous, brittle, and spotted leaf. These lines were also subjected to flanking sequence analysis for cataloguing genes disrupted by insertion of *Tos17* (see the following section). Some of the mutants were picked out and further subjected to cosegregation analysis to determine whether the mutations were caused by *Tos17* insertions. Genetic analysis showed that all these mutations are recessive, although some did not segregate in a 1:3 ratio. Because the original copy number of *Tos17* is quite low (one to four depending on varieties) in contrast to other plant retrotransposons such as *Tnt1* (several hundred copies, Grandbastien et al 1989) and *Tto1* (30 copies, Hirochika 1993), it is possible to identify the specific *Tos17* copy cosegregating with the mutant phenotype by DNA gel blot analysis. In the mutants, five to 30 transposed *Tos17* copies were detected and perfect cosegregation of one *Tos17* copy was observed in 5% to 10% of the mutant lines examined. These results strongly suggest that the cosegregating mutations are caused by *Tos17* insertion. Genomic sequences corresponding to mutant loci were amplified by IPCR or TAIL-PCR (Liu and Whittier 1995) and partially sequenced. Some mutants exhibited phenotypes expected from biochemical functions deduced from disrupted gene sequences. For example, in one viviparous mutant, the zeaxanthin epoxidase gene was disrupted. Because this gene, originally isolated in *Nicotiana plumbaginifolia* by transposon-tagging using the maize *Ac* (Marin et al 1996), has been shown to be involved in the abscisic acid (ABA) biosynthetic pathway, it is reasonable to expect a viviparous phenotype. In the rice mutant, the induced ABA level was dramatically reduced. Finally, tagging was confirmed by characterizing an allelic mutant that was identified by PCR screening among the *Tos17*-induced mutant population (see the next section). As expected, the allelic mutant showed the viviparous and wilty phenotype. Another example is a gibberellin (GA)-sensitive dwarf

mutant. The mutation in the *ent*-kaurene synthase gene cosegregated with the mutant phenotype and another allelic mutant showed the same dwarf phenotype, demonstrating the causal relationship between the mutation and the observed mutant phenotype. One semidwarf mutant exhibited insensitivity to brassinolide, suggesting that the mutant is deficient in brassinolide-signaling. An allelic mutant identified by PCR screening showed a similar phenotype. The causative gene showed no homology to known genes. These data clearly demonstrate that transposon-tagging with *Tos17* is quite feasible and useful for cloning of important genes. Table 1 summarizes examples of successful tagging. In addition to listed genes, many other genes are being analyzed to confirm the causal relationship.

Although the above results demonstrated that transposon-tagging using *Tos17* is feasible, we must consider the problem of low tagging efficiency. A similar problem was noted when T-DNA and *Ac/Ds* elements were used (Bancroft et al 1993, Long et al 1993, Azpiroz-Leehan and Feldmann 1997). In these cases, nontagged mutations are thought to be caused by abortive insertion events that result in a rearrangement at the target site without integration of the elements (Azpiroz-Leehan and Feldmann 1997, Bancroft et al 1993). A rearrangement caused by imprecise excision would be another important factor, as has been discussed in rice, in which the *Ac/Ds* system was employed (Izawa et al 1997). Mutations not tagged with *Tos17* may be due to insertions of unknown transposable elements activated by tissue culture, whereas tissue-culture-induced mutations not related to transposable elements (Larkin and Scowcroft 1981, Brettell et al 1986, Dennis et al 1987) may also be involved. To increase the efficiency of transposon-tagging, it is important to find new transposable elements responsible for tissue-culture-induced mutations and/or reduce mutation frequency not related to transposable elements. As discussed above, transposition of plant retrotransposons including *Tos17* is mainly regulated at the transcriptional level. Thus, it would be possible to avoid the problems associated with tissue culture by modifying the promoter of *Tos17* or by finding new conditions under which *Tos17* can be activated.

Retrotransposons as a tool for reverse genetic studies

Two strategies are employed to screen mutants for reverse genetic studies. One is the PCR screening of mutants. Mutants of the gene of interest can be identified in a large mutant population by PCR. Another important strategy is the random sequencing of

Table 1. Genes cloned by tagging methods.

Mutant phenotype[a]	Causative genes
Dwarf (gibberellin-deficient)	*ent*-kaurene synthase
Semidwarf (brassinolide-insensitive)	No homology
Viviparous (ABA-deficient)	Zeaxanthin epoxidase
Viviparous (ABA-deficient)	Transcription factor?
Pale green	*TatC*-homologue
Brittle	Cellulose synthase (*OsCesA7*)
Narrow leaf	No homology

[a]ABA = abscisic acid.

mutated genes by isolating the sequences flanking transposed *Tos17*. Loss-of-function phenotypes shown by identified mutants can then be studied to determine the biological function of the gene.

Screening of mutants of a specific gene by using PCR. This method was first developed in *Drosophila* (Ballinger and Benzer 1989, Kaiser and Goodwin 1990) and has since been used in *Caenorhabditis elegans* (Zwaal et al 1993), petunia (Koes et al 1995), and maize (Bensen et al 1995, Das and Martienssen 1995). Figure 3A shows the principle of PCR screening of mutants. Four combinations of two gene-specific

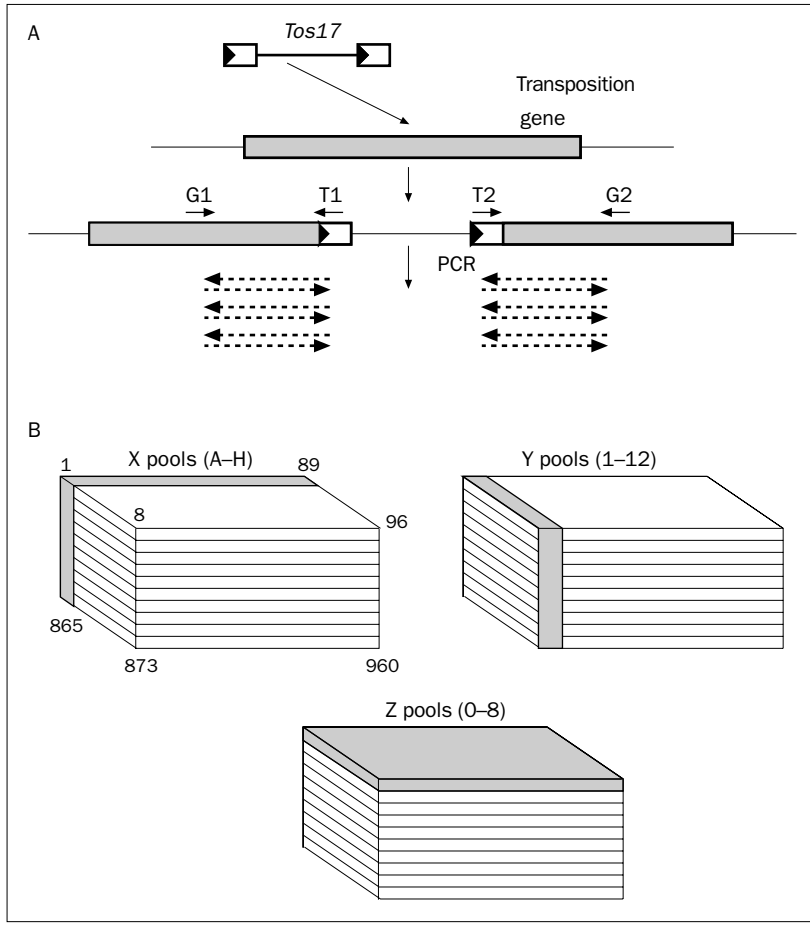


Fig. 3. (A) Use of polymerase chain reaction (PCR) to detect *Tos17* insertion mutants of a specific gene. Four combinations of two gene-specific primers (G1, G2) and two *Tos17*-specific primers (T1, T2) are used in separate reactions to detect *Tos17* insertions irrespective of their position and orientation. (B) Three-dimensional pooling of DNAs of rice mutant lines. A population of 960 lines was arrayed in a 96-well format. Seeds were pooled in a three-dimensional matrix and DNA was extracted from shoots derived from each pool of seeds. A total of 270 pooled DNAs were extracted from nine populations (a total of 8,640 lines).

primers (G1, G2) and two *Tos17*-specific primers (T1, T2) were used in separate reactions to detect *Tos17* insertions irrespective of their position and orientation. Two primer combinations produce PCR products if mutants carry insertions between the T1 and T2 primer sites. If mutants carry insertions outside of this region, only one primer combination produces PCR products. To carry out PCR screening efficiently, a two- or three-dimensional DNA-pooling system can be used. To screen a small population, a two-dimensional system can be used. A mutant of the homeobox gene (*OSH15*) was found in a small population consisting of 529 lines (Sato et al 1999). Based on the phenotype of the mutant, *OSH15* was shown to be involved in the development of internodes.

Figure 3B shows a three-dimensional pooling system. A population of 960 mutant lines was arrayed in a 96-well format. Seeds of the lines were pooled in a three-dimensional matrix and DNA was extracted from shoots derived from each pool of seeds. From nine populations (a total of 8,640 lines), 270 pooled DNAs were extracted. A total of 8,640 lines can be screened by 72 ($= 2 \times 4 \times 9$) PCR reactions using a super pool consisting of four X-pool DNAs. Figure 4 shows one example of screening. Eighteen mutant alleles of *OSERECTA* were identified. Considering the saturation level (see below) of the mutant population used for PCR screening, the *OSERECTA* gene must be a hot spot for *Tos17* insertion. This gene is homologous to *ERECTA* of *Arabidopsis* (Torii et al 1996); it is thus named *OSERECTA*. If a mutant of a target gene is found, Y- and Z-pool DNAs from the positive population undergo PCR screening to identify the mutant line. A total of 11,809 lines carrying 84,975 insertions have been screened for mutants of 47 genes and mutants of 15 genes have been identified. These include mutants of genes involved in signal transduction and genes for transcription factors. Some mutants are being characterized for their phenotypes. Among those, mutants of *OSMPK2* (MAP kinase gene), *OSDWARF3* homologous to the maize *DWARF3* gene (Winkler and Helentjaris 1995), and *OSERECTA* showed clear phenotypes.

The results described above demonstrate the feasibility of PCR screening of mutants. To make this system generally applicable to any gene of interest, mutations induced by *Tos17* insertion must be saturated. The success rate (15/47) of PCR screening suggests that at least 37,000 lines are required for saturation mutagenesis. The number of lines required for saturation mutagenesis can also be estimated by using the equation of Clarke and Carbon (1976). By assuming that (1) *Tos17* transposes randomly throughout the chromosomes, (2) the average size of rice genes is 3.0 kb, and (3) the number of transposed *Tos17* copies per regenerated plant is eight, then a total of 58,875 mutant lines are required to have a 99% chance of finding a mutant of any gene. Because *Tos17* prefers low-copy, gene-rich regions as integration targets, fewer mutant lines may be enough for saturation. Considering the ease of tissue culture and regeneration of rice, saturation mutagenesis with *Tos17* is quite feasible.

Systematic screening of mutants by sequencing Tos17 insertion sites. Although PCR screening seems to be the most efficient approach for reverse genetics, it may not be suitable for the analysis of a large number of genes. Considering large-scale functional genomics in the postsequencing era, a systematic approach to find mutants for a large number of genes is needed. One of the possible approaches is cataloguing

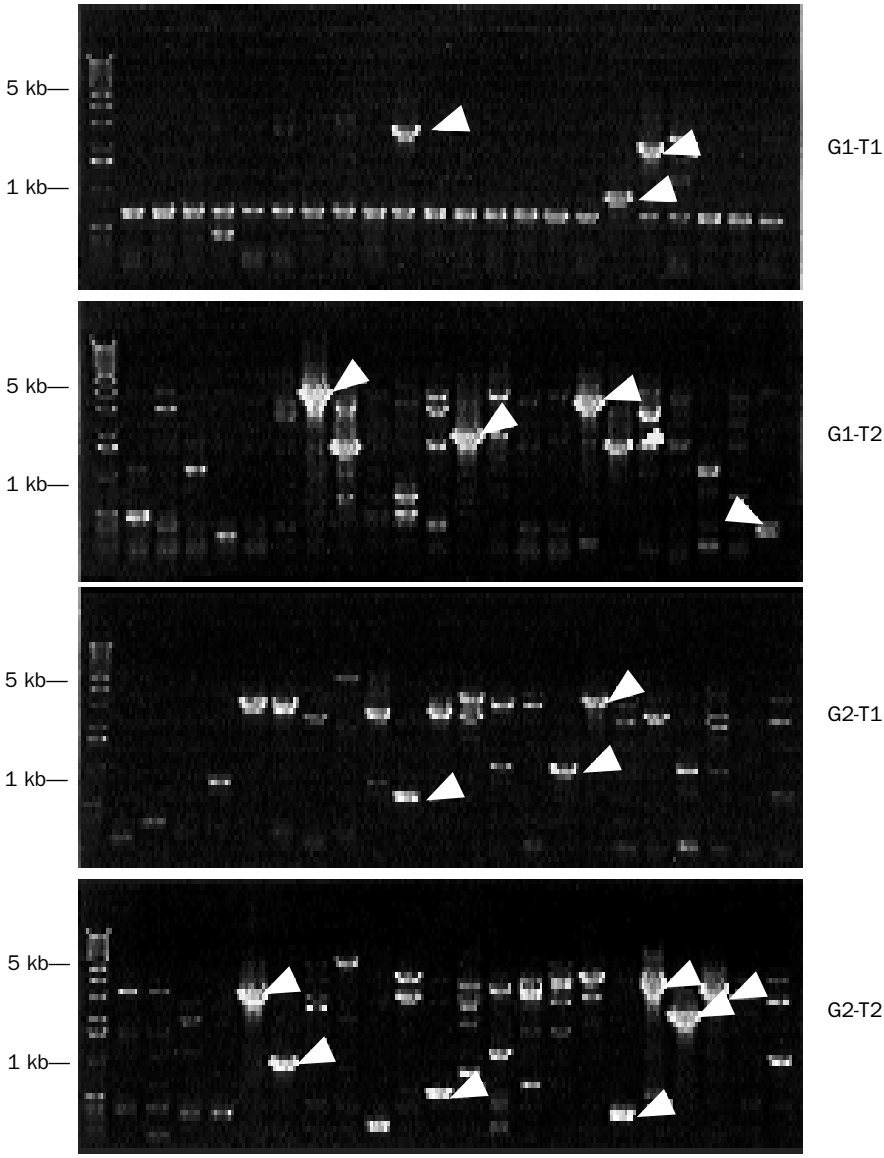


Fig. 4. Screening for mutants of *OSERECTA*. Twenty-two superpools underwent polymerase chain reaction using four combinations of two gene-specific primers (G1, G2) and two *Tos17*-specific primers (T1, T2). Positive bands (indicated by arrowheads) were detected by DNA blotting with a gene-specific probe. A total of 10,944 lines carrying 72,000 insertions were screened.

of insertion mutants by sequencing the genomic DNA sequence flanking insertions. As we reported previously, mutants of interesting genes were identified by sequencing only eight sequences flanking *Tos17* insertions (Hirochika et al 1996). In this case, IPCR was used to amplify the flanking sequences. To carry out large-scale sequencing of the flanking sequences, more simple PCR methods such as TAIL- and suppression-PCR (Siebert et al 1995) were adopted (Miyao et al 1998) and PCR products were directly sequenced. Two different PCR methods were employed to increase the efficiency of amplifying flanking sequences. By combining these two methods, about 95% of flanking sequences were obtained. Sequences obtained underwent a similarity search using the BLASTX program. More than 20,000 flanking sequences from 2,134 lines have been determined. These are classified into 7,376 independent flanking sequences and about 30% of the sequences showed homology to known genes. Different classes of genes, such as genes for transcription factors, genes involved in the signal transduction, genes with similarity to disease-resistance genes, and genes involved in metabolism, have been shown to be disrupted by *Tos17* insertion (Table 2). Insertions into DNA-type transposable elements and retrotransposons including *Tos17* have also been found. This strategy should be useful not only for identification of mutants but also for discovery of new genes. Some 36,000 ESTs of rice have been sequenced but these correspond to only one-third to one-half of the rice genes, considering the redundancy (Sasaki 1997). This suggests that at least one-half of the mutant genes identified by random sequencing of the *Tos17*-flanking sequences are new genes. If the entire genomic sequencing is completed, the disrupted genes can be readily identified and the insertion sites can be mapped by comparing the flanking sequences with the genomic sequence. Thus, flanking-sequence databases will become a powerful tool for reverse genetic studies.

Table 2. Categorization of disrupted genes identified by sequencing *Tos17*-flanking sequences.

Category[a]	Number
Cell growth/division	136
Cell structure	35
Disease or defense mechanism	402
Energy	32
Intracellular traffic	7
Metabolism	212
Protein destination and storage	58
Protein synthesis	24
Secondary metabolism	39
Signal transduction	274
Transcription	114
Transporters	124
Transposons (*Tos17*)	255 (16)
Unclear classification	995
Total	2,707

[a]A total of 7,376 independent flanking sequences from 2,134 lines were analyzed using the BLASTX program. Sequences showing significant similarity (E-value less than e^{-4}) to known proteins in the database are assigned to functional categories as described (Bevan et al 1998).

Mutants identified by this method are being characterized for their phenotypes and some showed clear phenotypes. For example, one mutant line was shown to carry the *Tos17* insertion in the gene with a significant homology to a RING finger gene. This gene is new in rice and is thus named *OSRING1*. Diverse important functions of RING finger genes have been reported in yeast, nematode, *Drosophila*, and mammals (Saurin et al 1996), such as oncogenesis, signal transduction, and biogenesis of peroxisome. However, because the functions of plant RING finger genes are not known except for *COP1* (Torii et al 1998), we started the functional analysis of the *OSRING1* gene. The mutant showed deficiency in lateral root development. Introducing the 5.3-kb genomic fragment containing the entire ORF driven by the 35S promoter could complement the mutant phenotype, confirming that it is caused by the disruption of *OSRING1*. Further analysis showed that *OSRING1* negatively regulates ethylene production and thus controls cell elongation during lateral root development.

In summary, these results indicate that the strategy described in this section is feasible and applicable to the genome-wide screening of mutants. Another important result obtained by the analysis of the flanking sequences is that the ends of transposed *Tos17* copies are intact. In contrast, rearrangements of terminal regions of T-DNA are induced quite often (Castle et al 1993, Krysan et al 1999). No induction of rearrangements of the *Tos17* ends makes it easy to isolate flanking sequences and screen for mutants by PCR.

References

Azpiroz-Leehan R, Feldmann KA. 1997. T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. Trends Genet. 13:152-156.

Ballinger DG, Benzer S. 1989. Targeted gene mutations in *Drosophila*. Proc. Natl. Acad. Sci. USA 86:9402-9406.

Bancroft I, Jones JDG, Dean C. 1993. Heterologous transposon tagging of the *DRL1* locus in *Arabidopsis.* Plant Cell 5:631-638.

Bensen RJ, Johal GS, Crane VC, Tossberg JT, Schnable PS, Meeley RB, Briggs SP. 1995. Cloning and characterization of the maize *An1* gene. Plant Cell 7:75-84.

Bevan M and 67 others. 1998. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391:485-488.

Brettell RIS, Dennis ES, Scowcroft WR, Peacock WJ. 1986. Molecular analysis of a somaclonal mutant of maize alcohol dehydrogenase. Mol. Gen. Genet. 202:235-239.

Castle LA, Errampalli D, Atherton TL, Franzmann LH, Yoon ES, Meinke DW. 1993. Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis.* Mol. Gen. Genet. 241:504-514.

Clarke L, Carbon J. 1976. A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. Cell 9:91-99.

Das L, Martienssen R. 1995. Site-selected transposon mutagenesis at the *hfc109* locus in maize. Plant Cell 7:287-294.

Dennis ES, Brettell RIS, Peacock WJ. 1987. A tissue culture induced *Adh* null mutant of maize results from a single base change. Mol. Gen. Genet. 210:181-183.

Grandbastien M-A. 1998. Activation of plant retrotransposons under stress conditions. Trends Plant Sci. 3:181-187.

Grandbastien M-A, Spielman A, Caboche M. 1989. *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376-380.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

Hirochika H. 1993. Activation of tobacco retrotransposons during tissue culture. EMBO J. 12:2521-2528.

Hirochika H. 1997. Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol. Biol. 35:231-240.

Hirochika H. 1999. Retrotransposons of rice as a tool for forward and reverse genetics. In: Shimamoto K, editor. Molecular biology of rice. Tokyo (Japan): Springer-Verlag. p 43-58.

Hirochika H, Fukuchi A. 1992. Transposable elements in rice plants. Jpn. Agric. Res. Q. 25:230-237.

Hirochika H, Fukuchi A, Kikuchi F. 1992. Retrotransposon families in rice. Mol. Gen. Genet. 233:209-216.

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. USA 93:7783-7788.

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miyazaki C, Endo T, Iida S, Shimamoto K. 1997. Transposon tagging in rice. Plant Mol. Biol. 35:219-229.

Kaiser K, Goodwin SF. 1990. "Site-selected" transposon mutagenesis of *Drosophila*. Proc. Natl. Acad. Sci. USA 87:1686-1690.

Koes R, Souer E, van Houwelingen A, Mur L, Spelt C, Quattrocchio F, Wing J, Oppedijk B, Ahmed S, Maes T, Gerats T, Hoogeveen P, Meesters M, Kloos D, Mol JNM. 1995. Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. Proc. Natl. Acad. Sci. USA 92:8149-8153.

Krysan PJ, Young JC, Sussman MR. 1999. T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11:2283-2290.

Larkin PJ, Scowcroft WR. 1981. Somaclonal variation: a novel source of variability from cell culture for plant improvement. Theor. Appl. Genet. 60:197-214.

Liu Y-G, Whittier RF. 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragment from P1 and YAC clones for chromosomal walking. Genomics 25:674-681.

Long D, Martin M, Sundberg E, Swinburne J, Puangsomlee P, Coupland G. 1993. The maize transposable element system *Ac/Ds* as a mutagen in *Arabidopsis*: identification of an *albino* mutation induced by *Ds* insertion. Proc. Natl. Acad. Sci. USA 90:10370-10374.

Marin E, Nussaume L, Quesada A, Gonneau M, Sotta B, Hugueney P, Frey A, Marion-Poll A. 1996. Molecular identification of zeaxanthin epoxidase of *Nicotiana plumbaginifolia*, a gene involved in abscisic acid biosynthesis and corresponding to the *ABA* locus of *Arabidopsis thaliana*. EMBO J. 15:2331-2342.

Miyao A, Yamazaki M, Hirochika H. 1998. Systematic screening of mutants of rice by sequencing retrotransposon-insertion sites. Plant Biotechnol. 15:253-256.

Motohashi R, Ohtsubo E, Ohtsubo H. 1996. Identification of *Tnr3*, a *Suppressor-Mutator/Enhancer*-like transposable element from rice. Mol. Gen. Genet. 250:148-152.

Noma K, Nakajima R, Ohtsubo H, Ohtsubo E. 1997. *RIRE1*, a retrotransposon from wild rice *Oryza australiensis*. Genes Genet. Syst. 72:131-140.

Ohtsubo H, Kumekawa N, Ohtsubo E. 1999. *RIRE2*, a novel gypsy-type retrotransposon from rice. Genes Genet. Syst. 74:83-91.

Okagaki RJ, Wessler SR. 1988. Comparison of non-mutant and mutant *waxy* genes in rice and maize. Genetics 120:1137-1143.

Parinov S, Sevugan M, Ye D, Yang WC, Kumaran M, Sundaresan V. 1999. Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. Plant Cell 11:2263-2270.

Pouteau S, Huttner E, Grandbastien MA, Caboche M. 1991. Specific expression of the tobacco *Tnt1* retrotransposon in protoplasts. EMBO J. 10:1911-1918.

Sato Y, Sentoku N, Miura Y, Hirochika H, Kitano H, Matsuoka M. 1999. Loss-of-function mutations in the rice homeobox gene *OSH15* affect the architecture of internodes resulting in dwarf plants. EMBO J. 18:992-1002.

Sasaki T. 1997. The rice genome project in Japan. Proc. Natl. Acad. Sci. USA 95:2027-2028.

Sasaki T, Burr B. 2000. International rice genome sequencing project: the effort to completely sequence the rice genome. Curr. Opin. Plant Biol. 3:138-141.

Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimura K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkaka I, Minobe Y. 1994. Towards cataloguing all rice genes: large-scale sequencing of randomly chosen cDNAs from a callus cDNA library. Plant J. 6:615-624.

Saurin AJ, Borden KLB, Boddy MN, Freemont PS. 1996. Does this have a familiar RING? Trends Biochem. Sci. 21:208-214.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. 1995. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. 23:1087-1088.

Speulman E, Metz PL, van Arkel G, te Lintel Hekkert B, Stiekema WJ, Pereira A. 1999. A two-component enhancer-inhibitor transposon mutagenesis system for functional analysis of the *Arabidopsis* genome. Plant Cell 11:1853-1866.

Takeda S, Sugimoto K, Otsuki H, Hirochika H. 1998. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Mol. Biol. 36:365-376.

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD. 1999. Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. Plant Cell 11:1841-1852.

Torii KU, McNellis TW, Deng XW. 1998. Functional dissection of *Arabidopsis COP1* reveals specific roles of its three structural modules in light control of seedling development. EMBO J. 17: 5577-5587.

Torii KU, Mitsukawa N, Oosumi T, Matsuura Y, Yokoyama R, Whittier RF, Komeda Y. 1996. The *Arabidopsis ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell 8:735-746.

Wang S, Zhang Q, Maughan PJ, Saghai Maroof MA. 1997. *Copia*-like retrotransposons in rice: sequence heterogeneity, species distribution and chromosomal locations. Plant Mol. Biol. 33:1051-1058.

Winkler RG, Helentjaris T. 1995. The maize *Dwarf3* gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis. Plant Cell 7:1307-1317.

Yamamoto K, Sasaki T. 1997. Large-scale EST sequencing in rice. Plant Mol. Biol. 35:135-144.

Zwaal RR, Broeks A, van Meurs J, Groenen JT, Plasterk RH. 1993. Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon mutant bank. Proc. Natl. Acad. Sci. USA 90:7431-7435.

Notes

Authors' addresses: H. Hirochika, A. Miyao, M. Yamazaki, S. Takeda, K. Abe, R. Hirochika, G.K. Agrawal, T. Watanabe, K. Sugimoto, T. Sasaki, A. Miyazaki, Department of Molecular Genetics, National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305-8602, Japan; K. Murata, K. Tanaka, K. Onosato, Y. Yamashita, N. Kojima, Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305-0854, Japan.

Acknowledgments: This study was supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences, a grant for the enhancement of Center-of-Excellence, special coordination funds for promoting science and technology in Japan, and a project grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Rice Genome Project MP-2101, 2102).

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Bioinformatics and the rice genome

B.A. Antonio, K. Sakata, and T. Sasaki

As rice genomics data continue to accumulate at a rapid rate, databases are becoming more valuable for storing and providing access to large and rigorous data sets. This chapter gives an overview of available resources on rice bioinformatics and their role in elucidating and propagating biological and genomic information on rice. Of particular focus here is the informatics infrastructure developed at the Rice Genome Research Program (RGP) following an extensive rice genome analysis. The database named INE (INtegrated Rice Genome Explorer) integrates genetic and physical mapping information with the genome sequence being generated in collaboration with the International Rice Genome Sequencing Project (IRGSP). Database links are initially evaluated using a query tool to explore and compare data across the rice and maize genome databases and for potential application to multiple-crop database querying. A proposed logistics for interlinking these resources is presented to integrate, manipulate, and analyze information on the rice genome. One of the biggest challenges of rice bioinformatics lies in the emerging role of rice as a model system among grass crop species. In view of the importance of comparative genomics in the formulation of new knowledge on plant genome structure and function, bioinformatics remains an essential strategy for gaining new insights into the needs and expectations of rice genomics.

Bioinformatics is a new field that emerged in parallel with the advances achieved in genomic analysis. Improved techniques in molecular biology played a key role in catalyzing large-scale sequencing of expressed sequence tags (ESTs), construction of whole genetic maps with specified markers, physical mapping with large insert-size libraries, whole genome sequencing, and transcriptional profiling (Benton 1996). This scenario of rapid technology development combined with mass production of genomic data led to a vital need to transform massive information into more manageable forms by way of bioinformatics. Advances in computer technology including the emergence of the World Wide Web and the Internet, now dominating every aspect of

human endeavor, greatly contributed to the revolution in bioinformatics (Lim and Butt 1998). Initially designed as a worldwide computer network to support education and research (Lee et al 1998), the Internet is now viewed as the most ideal vehicle for delivering genomic data to the scientific community.

Bioinformatics is defined as the study of information content and information flow in biological systems and processes. This definition overlaps with computational biology, which encompasses the use of algorithmic tools to facilitate biological analyses, and bioinformation infrastructure comprising the entire body of information management systems, analysis tools, and communication networks supporting biology (Lim and Butt 1998). For our purposes, bioinformatics incorporates both of these capabilities into a broad interdisciplinary science that involves both conceptual and practical tools for the understanding, generation, processing, and propagation of biological information (Spengler 2000). The core of bioinformatics is the database (Rolland 1998). Historically, databases have been largely used as repositories of data accumulated from genome-wide investigations. The immediate accessibility of resources and scientific data through databases disseminated via the Internet provides an unlimited data-mining opportunity for investigators. Furthermore, computational tools and approaches through bioinformatics may be used for evaluating and manipulating resources from databases, leading to the formation of new findings or theories even though these may require adequate validation.

Following the successful launching of the human genome project, a new era of whole genome science has emerged (Collins et al 1998). Refinements of a wide array of molecular biology techniques and the advent of sophisticated information technology have redefined the future direction of genome-wide analyses of several organisms. The complete genome sequence of the model plant *Arabidopsis*, likely to be released soon, is anticipated to revolutionize plant biology. A similar effort being undertaken in rice will directly transform the feasibility and utility of sequencing large genomes for improvement of cereal crops that provide the bulk of the world's food supply. Our main goal at the Rice Genome Research Program (RGP) is directed toward a comprehensive investigation on the structure of the rice genome and the elucidation of all its genes and functions (Sasaki 1999). In the process, we hope to establish the foundation for structural, functional, and applied genomics of rice.

In this chapter, we review rice bioinformatics resources that are available to the research community and briefly describe database projects encompassing biological and genomic information in rice. Of particular focus is the bioinformatics infrastructure developed at RGP principally covering integrated information on structural analysis of the genome as well as the elucidated sequences from the International Rice Genome Sequencing Project (IRGSP). Suggested measures to help improve the current status of rice bioinformatics will be presented.

Rapid growth of genomic data

Biological data are increasing at a phenomenal rate with the advent of high-throughput genome projects. The growing number of biological data on rice is reflected in the

total number of sequences submitted to the International Nucleotide Sequence Database Collaboration, where rice ranks fifth in number of accessions among all organisms and second among plants. Informatics infrastructure on rice is evidenced in several databases offering both general and specific information online. Key research centers involved in rice research have also made their resources available on the Web, including distribution of germplasm. More recently, the establishment of the international rice genome sequencing consortium has mobilized the participating groups to create individual repositories for their accumulated sequence data.

Much of the information on rice genomics has been brought about by the large-scale rice genome projects in the last decade. RGP has contributed greatly in developing the fundamental tools for rice genome analysis. Genomic information currently available on the Internet includes nucleotide sequences, genetic maps, physical maps, molecular markers, expressed sequence tags (ESTs), and morphological markers (Table 1). As of July 2000, IRGSP had already submitted a total of about 20 Mb of rice genome sequences to GenBank, European Molecular Biology Laboratory (EMBL), and DNA Data Bank of Japan (DDBJ). This collaboration of public sequencing centers from 10 countries is expected to make a major contribution to rice genomics (Sasaki and Burr 2000). In addition to the genome sequences, bacterial artificial chromosome (BAC)-end sequences that will be useful as sequence tag connectors are also available.

Table 1. Summary of rice genomic resources available on the World Wide Web.

Resource[a]	Number
DNA sequence accessions	153,153
Nucleotides (bp)	103,128,543
Genetic maps	9
Physical maps	2
Molecular markers	19,840
Morphological markers	185
QTLs	342
ESTs	51,000
BAC-end sequences	110,438
Worldwide strains	126,369
Tester and mutant lines, etc.	2,515
Landraces/improved varieties	6,293
Wild species	1,609
References	1,500

[a]Accessions and nucleotides are based on DDBJ release 42, July 2000. Map information including molecular and morphological markers was derived from INE, Oryzabase, and RiceGenes. Other sources of information are BAC-end sequences from the CUGI site; ESTs from the TIGR Rice Gene Index; and germplasm information from NIG, MAFF, NPGS-GRIN, CGIAR-SINGER, and NPGRC Web sites. QTLs = quantitative trait loci, ESTs = expressed sequence tags, BAC = bacterial artificial chromosome, INE = INtegrated Rice Genome Explorer, MAFF = Ministry of Agriculture, Forestry, and Fisheries, TIGR = The Institute for Genomic Research, NPGS-GRIN = National Plant Germplasm System-Germplasm Resources Integrated Network, NIG = National Institute of Genetics, CGIAR-SINGER = Consultative Group on International Agricultural Research-Systemwide Information Network for Genetic Resources, NPGRC = National Plant Genetic Resources Center.

Equally important as the genomic information are the germplasm resources that will provide the basis for direct application of molecular tools. These resources range from tester lines and mutant lines to cultivated varieties, landraces, and wild species maintained in various crop germplasm centers worldwide. The database servers provide information on acquisition, preservation, evaluation, and distribution of all available rice strains. The literature on different aspects of rice research has also been catalogued for searching on the Web for the user community.

Informatics for the rice community

More than 20 databases and Web sites are currently available online for rice genomics (Table 2). Most of these databases focus on specialized and detailed information covering a broad range of research on rice genomics or rice biology, whereas some are designed for ordinary users and offer general information on rice. An overview of these databases is briefly given here.

Rice genome databases

One of the earliest databases that have been serving the rice community since 1993 is RiceGenes (http://genome.cornell.edu/rice/). Supported by the USDA-ARS Center for Bioinformatics in partnership with Cornell University, this database focuses on comparative genetic maps, molecular markers, morphological markers, quantitative trait loci (QTLs), and germplasm data. Newly added features include microsatellite marker data and microsatellite maps of the 12 rice chromosomes. RiceGenes is currently being converted to a relational database management system to allow for increased data integration and greater flexibility.

Table 2. Major rice genome and germplasm databases.

Database/ organization[a]	URL	Information categories
INE	http://rgp.dna.affrc.go.jp/	Genetic map, YAC physical map, EST map, PAC contigs, chr. 1 and 6 PAC/ BAC clone sequence/annotation, IRGSP sequencing status
RiceGenes	http://genome.cornell.edu/rice/	Genetic maps, molecular markers, morphological markers, QTLs, comparative maps, rice variety releases, references
Oryzabase	http://www.shigen.nig.ac.jp/rice/oryzabase/	Genetic maps, integrated maps, genetic stocks, gene dictionary, mutant images, general information, references

continued on next page

Table 2 continued.

Database/ organization[a]	URL	Information categories
U.S. Rice Genome Sequencing	http://www.usricegenome.org/	Chr. 3, 10, and 11 sequencing status
TIGR Rice Database	http://www.tigr.org/tdb/rice/	Sequencing status, chr. 3, 10 BAC sequence data, annotation
Clemson Univ. Genomics Institute	http://www.genome.clemson.edu/	Rice BAC libraries, BAC-end sequence, rice BAC fingerprints, contig maps, chr. 3 and 10 sequencing status, rice blast STCs
Cold Spring Harbor Laboratory	http://nucleus.cshl.org/riceweb/	Chr. 10 sequencing status, BAC (cv. Teqing and Lemont) sequence/ annotation
Plant Genome Initiative at Rutgers	http://www.rutgers.edu/	Chr. 10 sequencing status, BAC sequence data, annotation
Korea Rice Genome Database	http://bioserver.myongji.ac.kr/	Rice ESTs, genetic maps, chloroplast genome sequence, chr. 1 (151.4–160.0 cM region), sequencing status
National Center for Gene Research, CAS	http://www.ncgr.ac.cn/index.html	Chr. 4 sequencing status, BAC contig map, BAC sequence data (indica)
Plant Genome Center Academia Sinica	http://genome.sinica.edu.tw/	Chr. 5 sequencing status, PAC sequence data, annotation
Thai Rice Genome Program	http://jasmine.kps.ku.ac.th/	Chr. 9 sequencing status
Rice Gene Index	http://www.tigr.org/tdb/ogi/	Rice ESTs, cDNA libraries, gene products
Rice Repeat Database	http://www.tigr.org/tdb/rice/blastsearch.html	Rice repeat sequences, mobile elements
RiceBlastDB	http://ascus.cit.cornell.edu/blastdb/	Genetic/physical maps, markers, population structure, DNA fingerprints, pathotype, mating type related to rice blast
MAFF DNA Bank	http://bank.dna.affrc.go.jp/	Rice cDNA clones, genetic map, clones/filters for distribution
MAFF Genebank	http://www.gene.affrc.go.jp/	Genetic stocks, evaluation data
GRIN	http://www.ars-grin.gov/npgs	Genetic stocks, evaluation data
SINGER	http://www.cgiar.org/ http://noc1.cgiar.org/	Genetic stocks maintained at IRRI, IITA, and WARDA, evaluation data
NPGRC	http://www.npgrc.tari.gov.tw/	Genetic stocks, evaluation data
Riceweb	http://www.riceweb.org/	General information

[a]TIGR = The Institute for Genomic Research, MAFF = Ministry of Agriculture, Forestry, and Fisheries, GRIN = Germplasm Resources Integrated Network, SINGER = Systemwide Information Network for Genetic Resources, NPGRC = National Plant Genetic Resources Center, YAC = yeast artificial chromosome, EST = expressed sequence tag, PAC = P1-derived artificial chromosome, BAC = bacterial artificial chromosome, IRGSP = International Rice Genome Sequencing Project, QTLs = quantitative trait loci, STCs = sequence tagged connectors, IRRI = International Rice Research Institute, IITA = International Institute of Tropical Agriculture, WARDA = West Africa Rice Development Association, CAS = Chinese Academy of Sciences.

Oryzabase is one of the most recently established rice genome databases developed at the National Institute of Genetics in Mishima, Japan, as part of a huge collection of biological germplasm databases called SHIGEN (SHared Information of GENetic Resources, http://www.shigen.nig.ac.jp/). It encompasses a wide range of information from general facts about rice and classical genetics to recent advances in rice genomics. The rice maps consist of the classical genetic map with phenotypic markers, molecular maps, and an integrated map with restriction fragment length polymorphism (RFLP) markers and phenotypic markers for cross-referencing of marker positions on each map. The database also features a compilation of rice strain information; a dictionary of all identified genes in rice with corresponding references; basic information on rice classification, morphology, and cultivation; and a mutant collection with corresponding images.

The rice genome analysis projects of various research institutions are also documented in several databases. The genome initiative of the Japanese Ministry of Agriculture, Forestry and Fisheries (MAFF) is represented at the RGP Web site (http://rgp.dna.affrc.go.jp) and the MAFF DNA Bank (http://bank.dna. affrc.go.jp) affiliated with the National Institute of Agrobiological Resources. The latter features the sequence analysis of all rice cDNA clones generated from RGP. The Korea Rice Genome Database (http://bioserver.myongji.ac.kr) is maintained by Myongji University as part of the Korea Rice Genome Research Program. It features a rice EST database, genetic maps, and chloropla2st genome sequences. The Thai Rice Genome Project Web site (http://jasmine.kps.ku.ac.th) focuses on rice genome projects in Thailand, including genome sequencing and gene analysis.

Genetic resources databases

The genetic resources databases principally cover important resources on the source of materials for rice research. These databases also offer essential information on the specific attributes of these resources, evaluation data, and their availability for distribution. The MAFF Gene Bank (http://www.gene.affrc.go.jp) features a total of 8,671 genetic stocks, including evaluation data such as morphological characters, resistance to stress, yield, quality, and ingredients. The National Institute of Genetics (Mishima, Japan) provides information on about 11,000 genetic stocks maintained in various universities and research stations in Japan (http://www.shigen.nig.ac.jp). These resources include marker gene testers, mutant lines, isogenic lines, autotetraploid lines, primary trisomics, reciprocal translocation homozygote lines, cytoplasm substitution lines, and cell-cultured lines. The Systemwide Information Network for Genetic Resources (SINGER) is an information exchange network of research centers affiliated with the Consultative Group on International Agricultural Research (CGIAR, http://www.cgiar.org). It contains key data on the identity, source, characteristics, and transfers to users of more than 106,000 accessions of rice genetic resources. The National Plant Genetic Resources Center in Taiwan (http://www. npgrc.tari.gov.tw) has a total of 3,980 strains. Rice germplasm data are also available via GRIN (Germplasm Resources Integrated Network, http://www.ars-grin.gov/npgs), the USDA germplasm database comprising pedigrees and trait data.

Specialized databases

In contrast to generalized databases, some specialized databases for rice offer in-depth information on a specific biological function or aspect of rice research. Participating members of IRGSP present the sequencing progress of their assigned chromosome(s) of the rice genome via independent databases. These databases incorporate the sequence of P1-derived artificial chromosome (PAC) or BAC clones for individual chromosomes, including the annotation of the sequence. The National Center for Gene Research, Chinese Academy of Sciences (http://www.ncgr.ac.cn), features the sequencing progress of chromosome 4 for the indica rice variety. The Plant Genome Center of Academia Sinica (http://genome.sinica.edu.tw/) provides the sequencing progress of chromosome 5. The United States Rice Genome Sequencing (http://www.usricegenome.org) integrates the sequencing efforts of the four participating U.S. groups: The Institute for Genomic Research (TIGR), the Clemson University/Cold Spring Harbor Laboratory / Washington University consortium (CCW), the Plant Genome Initiative at Rutgers (PGIR), and the University of Wisconsin. Each group also has independent sequencing databases covering specific regions of chromosomes 3, 10, and 11. The sequencing efforts of other participating groups such as Korea (chr. 1 region), UK (chr. 2), Canada (chr. 2), Thailand (chr. 9), India (chr. 11), and France (chr. 12) are expected to be integrated into their respective databases in the near future.

Other specialized databases on rice include the Rice Gene Index and the Rice Repeat Database, both of which are maintained by TIGR. The Rice Gene Index (http://www.tigr.org/tdb/ogi/) is part of the TIGR Gene Indices integrating ESTs from international sequencing projects and consists of 51,000 ESTs and ETS (expressed transcript sequences) from rice. These sequences are clustered and assembled using stringent overlap criteria to produce a high-quality consensus sequence. The advantage of this approach is that it provides a mechanism to link to candidate orthologues in other species (Quackenbush et al 2000). The Rice Repeat Database (http://www.tigr.org/tdb/rice/blastsearch.html) features a compilation of repeats and mobile elements making up 50% of the rice genome. It provides a search engine using BLASTN and an anonymous FTP link to download rice repeat sequences. Rice blast DB (http://genome. cornell.edu/riceblast/) developed by the RiceGenes group features various types of information on the rice blast fungus *Magnaporthe grisea*.

RGP informatics

The RGP Web site encompasses the activities of RGP, the sequencing status of IRGSP, and an integrated database of the rice genome. The informatics effort of RGP centers on the database named INE (INtegrated Rice Genome Explorer), which literally means "rice plant" in the Japanese language (Sakata et al 2000). This database was developed primarily to serve as a repository for all the information accumulated from the first phase of the rice genome project and to integrate the genetic and physical mapping data with the sequence of the rice genome that will be elucidated as part of the second phase of the program. In addition, it also functions as a repository of rice genome sequences from the international sequencing collaboration.

Database structure

A characteristic feature of INE is the integrated genetic and physical map of rice. The high-density linkage map currently contains 2,275 DNA markers (Harushima et al 1998) and an updated version with about 3,000 markers will be released soon. Information on the markers includes the probe size, the RFLP pattern as a result of parental screening, the sequence data, as well as the results of similarity searches. The markers on the genetic map are traceable to the ordered YAC (yeast artificial chromosome) clones covering 70% of the whole genome (Saji et al 2001). Information on the insert size of the clones is also provided. At present, approximately 4,500 ESTs are being assembled in the database. These ESTs were mapped by polymerase chain reaction (PCR) screening of our YAC library and positioned using the YAC physical map as a base (Wu et al 2000). Together with the genetic markers, these ESTs are subsequently used for ordering the PAC clone contigs to construct a sequence-ready physical map that serves as the template for genome sequencing. Integrating the structural information of the rice genome allows efficient use of available data particularly in the functional and applied aspects of genomics.

As part of the sequencing effort of RGP, the sequence data for chromosomes 1 and 6 are incorporated in INE. The ordered PAC clones in these chromosomes are displayed and linked to the sequence of the clone and the annotation of the sequence using various similarity searches and gene prediction programs. The annotation page consists of the annotation map for each clone and tables summarizing the predicted genes, including the results of the similarity searches and gene prediction programs. Furthermore, a list of low-quality sequence data is provided in a separate link as a requirement of the IRGSP to maintain an accuracy of less than one error in 10,000 base pairs (greater than 99.99%).

Navigation and visualization

The integrated maps for each chromosome facilitate a general overview of the genomic information in a particular chromosome. The maps can be manipulated by zoom in/out, which enables browsing at detail-oriented levels. Another distinct feature of INE is the rapid display of integrated maps. This has been achieved by programming the viewer in Java language using an application GIOT (Genome Information display Orderly Tool) developed by Mitsubishi Space Software Co. Ltd. This attribute contributes to smooth navigation of specific information associated with each data set. Clicking on a particular marker opens a Java applet window containing details such as screening data, image sets, sequences, etc., as well as detailed contig maps in the case of sequenced regions of the chromosome. These features allow getting all specific information together with an overall view of the distribution of the markers and clones in a short time. In addition, different chromosomes can be displayed simultaneously for direct comparisons between or among chromosomes.

The integrated map allows users to correlate various genomic information such as DNA markers, large insert clones, and associated ESTs. The genome sequence can also be directly traced to the physical and genetic maps. This will undoubtedly provide valuable information for positional cloning of target genes and for analysis of

more complicated traits such as QTLs. For instance, a richly informative map can be established for a particular trait by integrating the genetic and EST maps. With the information from the YAC-based physical map and PAC contigs, specific clones carrying markers tightly linked to the gene of interest can provide the template for either developing more markers or directly cloning the gene. Linking these data with the annotation of the sequence may provide necessary information on the identification of the genes involved in a particular trait of interest.

Interoperable query system for rice and maize

In collaboration with the Missouri Maize Project, an interoperable query system will be incorporated to INE to explore multiple crop databases (Antonio et al 2000). MaizeDB (http://www.agron.missouri.edu/) is a comprehensive public database of phenotypic, genetic, and genomic information on maize such as genetic maps, loci, sequences, phenotypes, QTLs, and relevant references. Cross-database querying and display of text objects between INE and MaizeDB are being implemented using a Web-based object-oriented query system called the OPM (Object-Protocol Model) data management tools of Gene Logic Inc. These tools are unique in their capacity to impose a uniform object-oriented data model on an existing relational database framework where users can explore and assemble biological information from heterogeneous databases (Chen 2000). This query system promotes direct analysis of colinearity at the nucleotide level in rice and maize species that may also be applied for exploring multiple crop databases.

Future prospects

Bioinformatics support is essential for implementing the rice genome research project. As INE will be used as the central repository of international sequencing collaboration, a system that will allow pulling of sequence data from the various sequencing centers is being incorporated. This will facilitate easy monitoring of the sequencing progress as well as maintenance of quality standards. A standardized rice annotation approach will be developed as well to permit gene discovery in a systematic, comprehensive, and consistent manner. This may require an automated analysis suite of tools and a platform that will allow reannotation on a regular basis.

Functional genomics at RGP is currently progressing with accumulated data derived from gene expression monitoring, insertional mutagenesis, and map-based cloning. The ESTs used in the microarray analyses are incorporated in the genetic and EST maps where direct linking of gene profiles can be done. A database of gene expression information will provide insights into the function and specificity of all genes in rice as well as extensive information on the profiles of genes related to growth and development and in response to various environmental changes. Eventually integrating information on rice structural and functional genomics will provide an overall view of the network of genes involved in complex biological responses.

We are also developing methods to further improve the interface of our database to facilitate efficient data mining and improve facile access to the database for all

types of users. An annotation database is also being constructed as we proceed with the annotation of the rice genome to categorize all the rice genes that have been identified so far. As rice genome data will increase exponentially, adequate tools for input, integration, and query will become necessary.

Rice informatics in the next decade

The major challenge for rice bioinformatics is to establish a comprehensive database that will allow integration of genomic information with present and future expectations in biological and agricultural research. Rice genome databases that evolve from rigorous and systematic sequencing efforts should not merely function as storehouses for thousands of bases or amino acids. Of particular importance is the ability to attach substantial genomic information to the sequence. These databases should therefore provide the framework to allow postsequencing analysis such as identifying genes and predicting the proteins they encode, determining when and where the gene proteins are expressed and how they interact, and how these expression and interaction profiles are modified in response to environmental signals. This may lead to a deeper understanding of biological processes that may be used for rice improvement. Meeting the demands and challenges of an ideal crop improvement strategy remains a matter of combining traditional breeding concepts and genomic tools through rigorous phases of experimentation. Therefore, emphasis on the underlying value of genotypic and genomic elements must be balanced with a phenocentric approach with emphasis on discovering genes that are most important in determining the phenotypes required in agriculture (Miflin 2000). Data representations for rice genomics should be able to cope with these innovations. One way to address this need is to interlink the resources of various types of information such as genomic data, phenotypic or expression data, and genetic resources (Fig. 1). For a given gene, the database would horizontally link sequence, structure, and map position and would connect related elements of the same type pertaining to the expression profile, proteins, and phenotypes. All this information should be defined in terms of the genetic resources available for rice. Logical connections to other information will enhance the intrinsic value of the genomic data to facilitate the formation of new biological discoveries and simulate approaches for an effective rice improvement program.

Rice, being a model system for other grass crops, should establish an informatics infrastructure designed to interlink database resources on rice genomics to better serve a more focused research community using this system in contrast to a larger user community. The proposed interlinking of database resources as outlined above hopes to address this need for the rice research community. This should be extended to databases of other cereal crops as well. One of the more serious challenges of specialized or expert domain databases, best represented by model organism databases, is to balance the needs between the broader scientific community and the specialized focused groups (Gelbart 1998).

As comparative genomics is now viewed as a key to extending existing knowledge on plant genomes and genes, bioinformatics remains an essential strategy for

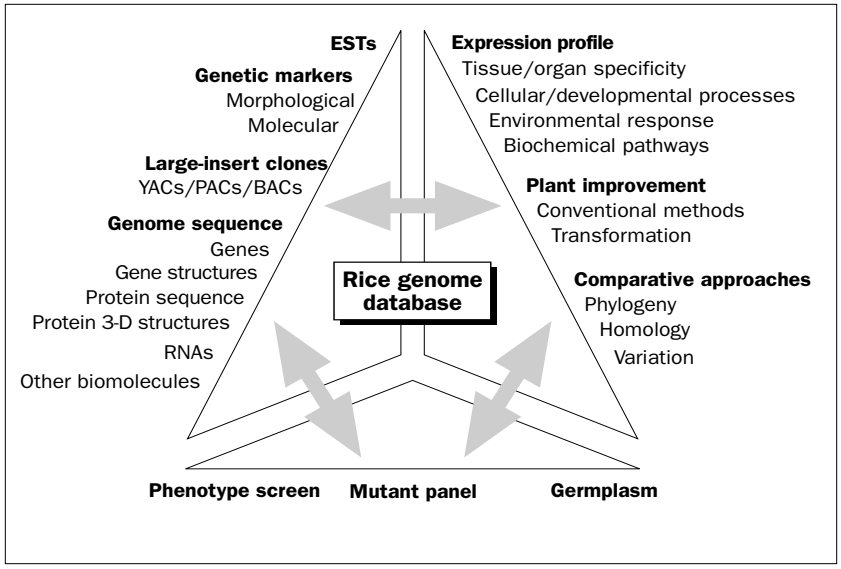


Fig. 1. Integrating genomic information and genetic resources in rice to facilitate the formation of new biological discoveries and simulate approaches for an effective rice improvement program. EST = expressed sequence tag, YAC = yeast artificial chromosome, PAC = P1-derived artificial chromosome, BAC = bacterial artificial chromosome.

pursuing the needs and expectations of rice genomics. Comparative bioinformatics offers possibilities to link various cereal crops through their genomes and will provide keys to understanding how genes and genomes are structured, how they function, and how they evolved. Genetic mapping as well as some preliminary sequence data show the extent of synteny among cereal crops (Bennetzen 1999). Through identification of synteny, it will be possible to isolate genes from crop plants with large genomes using information of homologous genes in related crops with smaller genomes. As the creation of links between different databases may foster interoperability (Karp 1996, Macauley et al 1998), linkages and interactions should also be promoted between databases of rice and other organisms.

An era of biological revolution has begun during which a tremendous amount of information on plant genetics will be accumulated over the next ten years. The need to transform these data into information that the rice community and biologists in general can query and use properly should be given high priority. As in any informatics effort, the essential factors necessary to achieve the desired success in rice bioinformatics involve developing the necessary tools for interlinking information or whole databases. As we move from the pregenomic to the postgenomic era, rice bioinformatics should also be able to adapt to the next challenge of analyzing biological and agricultural problems in multiple dimensions.

References

Antonio BA, Fang Z, Sanchez H, Polacco M, Chen SS, Sakata K, Ohta I, Mukai Y, Coe Jr. E, Sasaki T. 2000. An interoperable query system for INE and MaizeDB: expanding genomic resources for cereal crops. In: Proceedings of the 6th International Congress of Plant Molecular Biology, Quebec, Canada, 18-24 June 2000. Supplement to Reporter 18:2.

Bennetzen JL. 1999. Plant genomics takes root, branches out. Trends Genet. 15:85-87.

Benton D. 1996. Bioinformatics: principles and potential of a new multidisciplinary tool. Trends Biotechnol. 14:261-272.

Chen IA. 2000. OPM data management tools employed for plant genome databases. In: Proceedings of the 7th International Rice Genome Meeting, Tsukuba, Japan, 8-10 February 2000.

Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. 1998. New goals for the U.S. human genome project: 1998-2003. Science 282:682-689.

Gelbart WM. 1998. Databases in genomic research. Science 282:659-661.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:1-16.

Karp PD. 1996. Database links are a foundation to interoperability. Trends Biotechnol. 14:273-279.

Lee LE, Chin P, Mosser DD. 1998. Biotechnology and the Internet. Biotechnol. Adv. 16:949-960.

Lim HA, Butt TR. 1998. Bioinformatics takes charge. Trends Biotechnol. 16:104-107.

Macauley J, Wang H, Goodman N. 1998. A model system for studying the integration of molecular biology databases. Bioinformatics 14:575-582.

Miflin B. 2000. Crop improvement in the 21st century. J. Exp. Bot. 51:1-8.

Quackenbush J, Liang F, Holt I, Pertea G, Upton J. 2000. The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. 28:141-145.

Rolland FD. 1998. The essence of databases. London (UK): Prentice Hall Europe. 226 p.

Saji S, Umehara Y, Antonio BA, Yamane H, Tanoue H, Baba T, Aoki H, Ishige N, Wu J, Koike K, Matsumoto T, Sasaki T. 2001. A physical map with yeast artificial chromosome (YAC) clones covering 63 percent of the 12 rice chromosomes. Genome 44:32-37.

Sakata K, Antonio BA, Mukai Y, Nagasaki H, Sakai Y, Makino K, Sasaki T. 2000. INE: a rice genome database with an integrated map view. Nucleic Acids Res. 28:97-101.

Sasaki T, Burr B. 2000. International rice genome sequencing project: the effort to completely sequence the rice genome. Curr. Opin. Plant Biol. 3:138-141.

Sasaki T. 1999. Rice genome research in Japan as a frontrunner in crop genomics. Sci. Technol. Jpn. 18:18-21.

Spengler JS. 2000. Bioinformatics in the Information Age. Science 287:1221-1223.

Wu J, Yamamoto S, Maehara T, Harada C, Shimokawa T, Ono N, Mukai Y, Takazaki Y, Yazaki J, Koike K, Shomura A, Fujii F, Yano M, Ando T, Kono I, Sasaki T, Matsumoto T. 2000. A comprehensive map of rice expressed sequence tags (ESTs): its characteristics and application to genome analysis. In: Proceedings of the 6th International Congress of Plant Molecular Biology, Quebec, Canada, 18-24 June 2000. Supplement to Reporter 18:2.

Notes

Authors' address: Rice Genome Research Program, National Institute of Agrobiological Resources/Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305, Japan.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Gene isolation and function

Enhancing deployment of genes for blast resistance: opportunities from cloning a resistance gene/avirulence gene pair

B. Valent, G.T. Bryan, Y. Jia, L. Farrall, S.A. McAdams, K.N. Faulk, and M. Levy

Pi-ta-mediated resistance in rice controls strains of the rice blast fungus that express avirulent alleles of *AVR-Pita* in a gene-for-gene manner. Map-based cloning of *AVR-Pita* identified a gene predicted to encode a neutral zinc metalloprotease. We also cloned *Pi-ta*, a centromere-linked resistance gene on rice chromosome 12, using a map-based cloning strategy. It encodes a predicted 928-amino acid cytoplasmic receptor with a centrally localized nucleotide-binding site. *Pi-ta* is a single-copy gene and it shows constitutive low-level expression in both resistant and susceptible lines of rice. Susceptible rice lines contain *pi-ta*$^-$ alleles encoding proteins that share in common one amino acid difference relative to the Pi-ta resistance protein, serine instead of alanine at position 918. Direct interaction of the Pi-ta and AVR-Pita proteins was demonstrated using the yeast two-hybrid system and in vitro binding assays. Our current hypothesis is that the Pi-ta protein is an intracellular receptor that binds to the mature AVR-Pita protease inside the host cell, initiating *Pi-ta*-mediated defense responses. Understanding how pathogen recognition occurs provides the opportunity to manipulate recognition specificity. Population analysis of *AVR* gene structure and dynamics will aid in the deployment of resistance genes. The relationship between the *Pi-ta* and *Pi-ta*^{2} genes is discussed.

Genetic analysis of the rice blast fungus, *Magnaporthe grisea* (Hebert) Barr [anamorph *Pyricularia grisea* (Rossman et al 1990)], has defined avirulence (*AVR*) genes in the pathogen that function together with corresponding resistance (R) genes in rice to initiate a hypersensitive resistance response (HR) and block infection (Flor 1971, Kiyosawa 1971, Silué et al 1992, Zeigler et al 1994). It has been an important goal to understand the molecular mechanisms underlying this gene-for-gene relationship in plant-pathogen interactions. One particularly attractive model, the ligand and receptor model, suggests that the R gene encodes a receptor that recognizes an elicitor (or ligand) produced by expression of the corresponding *AVR* gene. In this model, *AVR* genes might encode pathogen molecules that are directly recognized by the R gene

product, or they might encode enzymes involved in the production of small molecule ligands that serve as recognition factors. The efficacy of any particular R gene in the field depends on the population biology of the corresponding *AVR* gene from the fungus. In particular, it depends on the role the *AVR* gene plays in the pathogenicity process itself, and on the potential for mutation of the *AVR* gene to avoid recognition without affecting pathogen fitness in the field.

Individual R genes can be distinguished only by the resistance they confer to corresponding avirulent strains of the pathogen (Flor 1971). It has been difficult to compare rice blast R genes identified in different countries due to regulatory prohibitions against sharing pathogen strains between geographically separated rice production regions. R gene identification is also complicated by the common occurrence of multiple *AVR* genes in single field isolates of the pathogen, any one of which is sufficient to trigger resistance in the presence of its corresponding R gene. Thus, the number of distinct blast R genes is uncertain and strategies for pyramiding R genes for blast control are hampered by imprecise R gene characterization. Molecular cloning and characterization of R genes and their corresponding *AVR* genes will provide the necessary tools to determine the number and nature of R genes available for blast disease control.

We describe the first molecular characterization of an R gene/*AVR* gene pair from the rice blast system. The resistance gene *Pi-ta*, previously referred to as *Pi-62(t)* (Wu et al 1996), encodes a putative cytoplasmic receptor (Bryan et al 2000), and the avirulence gene *AVR-Pita*, formerly known as *AVR2-YAMO*, encodes a putative metalloprotease (Orbach et al 2000). We report progress in understanding the molecular events by which the Pi-ta protein and the AVR-Pita protein act together to trigger resistance and block disease (Jia et al 2000). An understanding of R gene action in pathogen recognition is a first step toward targeted evolution of broad-spectrum R genes that will provide durable resistance. Cloning and sequencing of *Pi-ta* have identified ideal molecular markers for marker-assisted selection, and have enabled transgenic strategies for pyramiding *Pi-ta* with other R genes. Understanding the population biology of the *AVR-Pita* gene will aid in deploying *Pi-ta* using lineage exclusion strategies.

Molecular characterization of *Pi-ta* and *AVR-Pita*

Genetic identification of an R gene/*AVR* gene pair

Genetic analysis identified an *AVR* gene derived from the Chinese field isolate O-137, collected in 1985 at the China National Rice Research Institute in Hangzhou. This gene, originally named *AVR2-YAMO*, determined avirulence toward rice variety Yashiro-mochi, a Japanese differential variety for *Pi-ta* (Yamada et al 1976). A fungal progeny strain, 4360-R-62, from the O-137 pedigree contained the *AVR2-YAMO* gene in isolation from other *AVR* genes segregating in these crosses, and it was used to identify the corresponding R gene, originally named *Pi-62(t)*, using doubled-haploid and F_2 mapping populations from reciprocal crosses between rice varieties Yashiro-

mochi and Tsuyuake (Wu et al 1996). This *AVR* gene/R gene pair was shown to correspond to the previously identified *Pi-ta* gene, rather than to an undetected R gene in Yashiro-mochi, by the following results: (1) the R gene mapped to the centromeric region of chromosome 12 in the expected location for *Pi-ta*; (2) the second *Pi-ta* differential rice variety, K1 (Kiyosawa 1984), showed an infection spectrum identical to that of Yashiro-mochi in inoculations with fungal progeny within the O-137 pedigree; and (3) Japanese pathogen strains (avirulent strains Ken54-20, Ina72, and Ken54-04; virulent strains Ina168, Ken53-33, and P-2b) used in the original identification of *Pi-ta* (Kiyosawa 1971) gave consistent infection results in our rice mapping populations. We concluded that we are analyzing the *AVR-Pita* and *Pi-ta* genes defined by Kiyosawa and colleagues (Kiyosawa 1967, 1971), and we have adopted these gene names for consistency.

Pi-ta encodes a putative cytoplasmic receptor

Random amplified polymorphic DNA (RAPD) markers and bulked segregant analysis were used to identify molecular markers linked to *Pi-ta* (Wu et al 1996). The *Pi-ta* gene resides on chromosome 12 between flanking RAPD markers designated SP4B9 and SP9F3, both of which show one recombination event relative to *Pi-ta* in a mapping population of 990 individuals. The cosegregating restriction fragment length polymorphism (RFLP) marker p7C3 was used to initiate chromosome walking (Wu et al 1996, Bryan et al 2000). We assembled an overlapping set of rice bacterial artificial chromosome (BAC) clones covering approximately 850 kb. The left border marker SP4B9 was identified in the assembled BAC contig. However, SP9F3 has not yet been identified, possibly due to decreased recombination in the vicinity of a centromere. We adopted a sample sequencing strategy in order to identify R gene candidates in the BAC contig. BAC clones in this region are relatively high in repetitive DNA sequences, including the centromere-specific sequence RCE1 (Dong et al 1998, Singh et al 1996). This result is consistent with a previous report that the region near *Pi-ta* is high in repetitive DNA sequences (Nakamura et al 1997).

The *Pi-ta* candidate gene was identified by sequences from BAC142E8 showing homology to the nucleotide-binding site (NBS) region of R genes cloned from other systems, and this candidate R gene was subsequently confirmed as *Pi-ta* by stable transformation of rice (Bryan et al 2000). *Pi-ta* encodes a predicted cytoplasmic protein (Fig. 1A) with an NBS region characteristic of the largest class of plant R genes (Bent 1996, Hammond-Kosack and Jones 1997). *Pi-ta* lacks either a leucine zipper motif or the Toll/Interleukin-1 receptor homology reported as key features of dicot R genes. The C-terminal portion of the *Pi-ta* protein contains a leucine-rich domain (LRD), which also lacks features of the leucine-rich repeats reported for other R gene products. *Pi-ta* is a single-copy gene that is constitutively expressed in both resistant and sensitive varieties.

We compared *Pi-ta* sequences from several resistant and sensitive rice varieties (Bryan et al 2000). Differential rice varieties Yashiro-mochi (Yamada et al 1976) and

K1 (Kiyosawa 1984) obtained *Pi-ta* from different sources. While *Pi-ta* in variety K1 was introgressed from the indica variety Tadukan, the origin of *Pi-ta* in Yashiro-mochi appears to be an upland variety, Okaine (Rybka et al 1997). The DNA sequences of the genes cloned from Tadukan and K1 are identical to those of *Pi-ta* from Yashiro-mochi. In contrast, the Yashiro-mochi *Pi-ta* coding sequence differs by seven nucleotides from the sequence present in the sensitive japonica varieties Tsuyuake, Nipponbare, and Sariceltik. The seven base pair differences in the coding region result in five amino acid differences between the two proteins (Table 1). The amino acid sequence encoded by the susceptible indica C101A51 allele differs from the resistant protein in only one of these five amino acid differences, serine at residue 918. Thus, direct comparison of DNA sequence for *Pi-ta* homologues from resistant and sensi-

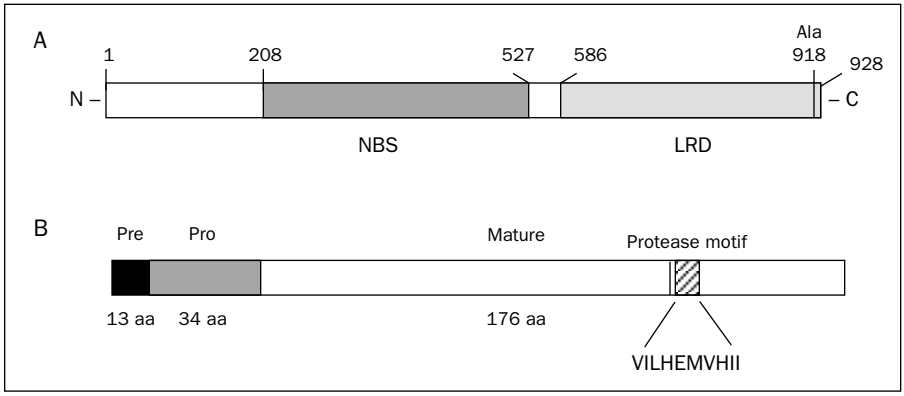


Fig. 1. (A) The predicted 928-amino acid Pi-ta protein. Only the nucleotide-binding site (NBS)-containing region has homology to other R genes. The C-terminal leucine-rich domain (LRD) of the Pi-ta protein, with critical residue 918, has been implicated in binding to AVR-Pita. (B) Working hypothesis on the structure of the putative AVR-Pita protease. A gene that encodes the predicted mature form of the protein was constructed and named *AVR-Pita*$_{176}$.

Table 1. Amino acid sequence differences among Pi-ta proteins from varieties carrying the *Pi-ta* and *Pi-ta*^{2} genes, and varieties with neither of these genes.

Variety	R gene	Resistant to *AVR-Pita* fungus	Amino acid[c] position				
			6	148	158	176	918
Yashiro-mochi[a]	*Pi-ta*	Yes	I	R	H	D	A
Tetep[a]	*Pi-ta*^{2}	Yes	I	R	H	D	A
C101A51	*Pi-2*	No	I	R	H	D	S
Tsuyuake[b]	*Pi-km*	No	S	S	Q	V	S

[a]Additional *Pi-ta*-containing variety K1 and additional *Pi-ta*^{2}-containing varieties Tadukan, Reiho, and Katy contain sequences identical to those of Yashiro-mochi and Tetep. [b]Additional susceptible japonica varieties Nipponbare and Sariceltik contain sequences identical to those of Tsuyuake. [c]Amino acid abbreviations are I = isoleucine, R = arginine, H = histidine, D = aspartic acid, A = alanine, Q = glutamine, S = serine, V = valine.

tive indica rice varieties supports the hypothesis that a single amino acid substitution, serine for alanine at position 918, eliminates *Pi-ta* function.

AVR-Pita encodes a putative metalloprotease

AVR-Pita was cloned by a map-based approach (Orbach et al 2000) based on its location adjacent to a telomere (Tel 5) on chromosome 3 of the integrated map of the blast fungus (Nitta et al 1997, Sweigard et al 1993). The *AVR-Pita* gene encodes a polypeptide with 223 amino acids and features that are characteristic of fungal-secreted metalloproteases. Amino acids 173–182 form a characteristic motif of a neutral zinc metalloprotease and mutational analysis of *AVR-Pita* indicates that maintenance of the protease motif is essential for avirulence (G.T. Bryan and B. Valent, unpublished results). Homology between the AVR-Pita protein and the metalloprotease NpII from *Aspergillus oryzae* is confined to the C-terminal 176 amino acids, corresponding to the mature region of the NPII prepro-protein (Tatsumi et al 1991). Thus, we hypothesize that AVR-Pita also contains a prepro-region that is processed to an active protease (Fig. 1B). Biochemical demonstration of protease activity, however, remains to be demonstrated for *AVR-Pita*.

A model for *Pi-ta*-mediated resistance

Transient expression of *Pi-ta* and *AVR-Pita* in plant cells induces resistance

If the ligand/receptor model for gene-for-gene interactions holds for the *Pi-ta*/*AVR-Pita* system, the Pi-ta receptor might bind either the AVR-Pita protein itself or a peptide released from a substrate of the putative AVR-Pita protease. Vacuum infiltration or other apoplastic applications of various forms of recombinant AVR-Pita protein (Fig. 1B) into *Pi-ta*-containing leaf tissue failed to induce hypersensitive resistance (HR) (G.T. Bryan and B. Valent, unpublished results). These data, along with the putative cytoplasmic location of *Pi-ta*, suggested that recognition of AVR-Pita may occur inside the plant cell.

A transient expression assay provides direct evidence that expression of *AVR-Pita* inside rice cells containing *Pi-ta* is sufficient to induce HR (Bryan et al 2000, Jia et al 2000). For this assay, 1-wk-old intact rice seedlings are biolistically transformed with a GUS reporter gene (*E. coli uidA* linked to the constitutive *35S* promoter from cauliflower mosaic virus). Bombarded cells accumulate β-glucuronidase enzyme (GUS), which is assayed histochemically for accumulation of blue pigment. GUS activity in a transformed cell indicates a "healthy" cell that is not undergoing HR, and reduced GUS activity indicates a cell undergoing HR. Thus, particle bombardment transformation of rice seedlings results in multiple spots of blue pigment in the absence of HR (Fig. 2).

We engineered the full-length *AVR-Pita* gene and *AVR-Pita*$_{176}$ encoding the predicted mature protease (Fig. 1B) for direct expression in plant cells, and introduced these constructs along with the GUS reporter gene in the transient assay (Jia et al 2000). In these assays, GUS activity was inhibited whenever *AVR-Pita*$_{176}$ and *Pi-ta* were expressed together in rice cells (Fig. 2). This was true whether the rice variety

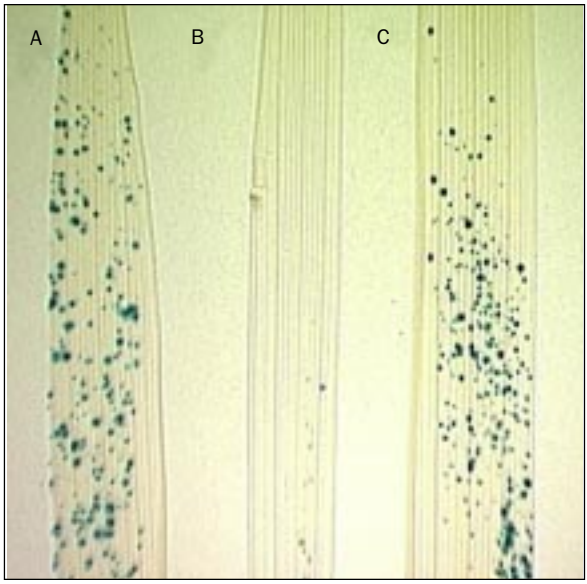


Fig. 2. Rice leaves showing localized blue pigment resulting from GUS enzyme activity after co-bombardment of the GUS reporter gene along with (A) *Pi-ta* alone, (B) *Pi-ta* and avirulent *AVR-Pita*, and (C) *Pi-ta* and virulent *avr-pita*$^{-}$. GUS activity is high in the absence of hypersensitive resistance (HR) and low in cells undergoing HR.

contained the native *Pi-ta* gene or whether a *Pi-ta* transgene was co-transformed, along with the GUS reporter and the *AVR-Pita*$_{176}$ genes, into rice varieties lacking *Pi-ta*. Transient expression of full-length *AVR-Pita* showed low levels of HR-inducing activity compared with *AVR-Pita*$_{176}$, supporting the hypothesis that the AVR-Pita protein is processed to a mature active form. Mutant *avr-pita*$_{176}$ genes that no longer confer avirulence to the pathogen in fungal infection assays also fail to induce HR in the transient assay (Bryan et al 2000, Jia et al 2000). The sensitive *pi-ta*$^{-}$ allele from C101A51 fails to induce HR in combination with *AVR-Pita*. These results suggest that AVR-Pita and Pi-ta proteins act together inside plant cells to induce defense responses. In addition, these experiments demonstrate that AVR-Pita is the only pathogen protein that is required for inducing *Pi-ta*-mediated resistance.

The AVR-Pita protein interacts directly with the Pi-ta protein

We have demonstrated that the *Pi-ta* LRD region binds to the avirulent AVR-Pita$_{176}$ protein in the yeast two-hybrid system (Jia et al 2000). In this assay, the DNA binding domain and the transcriptional activation domain of a transcription factor are cloned separately into vectors that produce fusion proteins between each domain and one of the proteins being tested. Binding between the two proteins of interest links the DNA-binding domain to the transcriptional activation domain and reconstitutes an active

transcription factor, resulting in activation of reporter gene expression in yeast. DNA fragments encoding either the full-length AVR-Pita or the predicted mature AVR-Pita$_{176}$ proteins were cloned into activation domain vectors and tested against a DNA-binding domain fusion, either to the full-length Pi-ta protein or to a DNA fragment encoding the *Pi-ta* LRD (Fig. 1A). Expression of the AVR-Pita$_{176}$ fusion protein together with the Pi-ta LRD fusion protein produced a functional transcription factor that allowed expression of two reporter genes, a histidine biosynthetic gene required for prototrophic growth and the *lacZ* reporter gene. This suggested that the AVR-Pita$_{176}$ protein binds directly to the LRD region from Pi-ta. Interaction, defined by expression of both reporter genes, was not detected when the AVR protein was fused to the DNA-binding domain and the LRD protein was fused to the activation domain. Interaction was also not detected with the sensitive pi-ta^{-} LRD^{A918S}, in which serine had been substituted for alanine at position 918. Furthermore, a virulent *avr-pita*$_{176}$ mutant with an amino acid substitution in the putative metalloprotease active site (E177D) failed to interact with either form of the LRD. Thus, the specificity of binding is consistent with the *in vivo* specificity in the *Pi-ta* system, and it seems reasonable to conclude that binding between AVR-Pita$_{176}$ and the leucine-rich domain of Pi-ta plays a role in inducing the plant defense response.

Biochemical confirmation of binding between the AVR-Pita$_{176}$ and Pi-ta LRD polypeptides was obtained by a modified "far-western analysis" (Jia et al 2000). For this procedure, recombinant forms of AVR-Pita polypeptides, which are insoluble as expressed in *E. coli*, were purified, separated using SDS-PAGE gels, transferred to a nitrocellulose membrane, and allowed to refold in the presence of $ZnCl_2$. Membranes containing refolded AVR-Pita proteins were incubated with extracts of *E. coli* proteins containing soluble S-tagged Pi-ta or LRD polypeptides and rinsed extensively. Pi-ta polypeptides that were bound to AVR-Pita polypeptides on these membranes were detected with the S-antibody (Novagen). Again, the LRD domain of Pi-ta bound specifically to the avirulent form of AVR-Pita$_{176}$. That is, LRD from the resistant allele bound to AVR-Pita$_{176}$, but not to longer forms of the AVR protein, and the LRD^{A918S} from the sensitive allele did not bind under the same conditions. Thus, specific Pi-ta/AVR-Pita binding has been demonstrated in two separate assays.

These data, along with the results of the transient expression assays, support the hypothesis that Pi-ta functions as a receptor that binds to the AVR-Pita$_{176}$ protein in the cytoplasm of the invaded plant cell (Fig. 3). In addition, they suggest that the elicitor-binding domain of Pi-ta protein resides within the leucine-rich C-terminal region and that a single amino acid substitution (A918S) eliminates R gene function.

AVR-Pita population analysis

Sasaki first discovered physiological races of the rice blast pathogen in 1922 (Yamada 1985) and, since that time, rice pathogens have been classified into hundreds of physiological races, or pathotypes, that differ in their ability to cause disease on different rice varieties. According to the gene-for-gene hypothesis, the pathotype of a strain of the fungus is determined by its *AVR* genes, which control sensitivity to various rice R

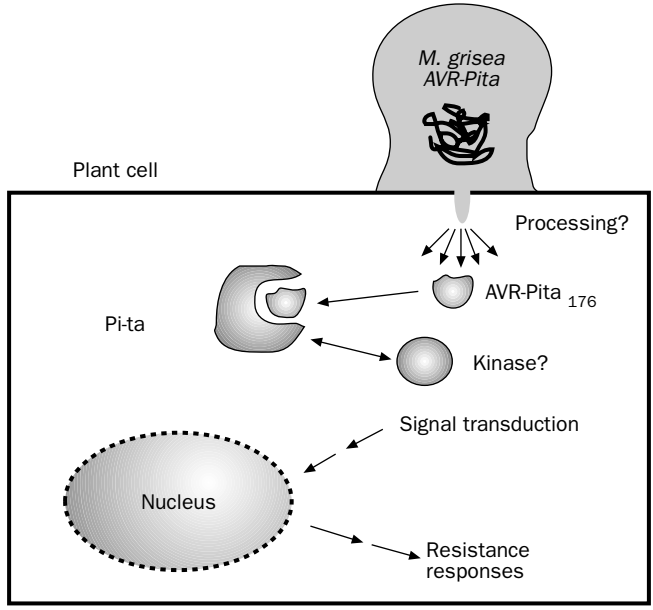


Fig. 3. Model for resistance mediated by Pi-ta. We hypothesize that the predicted mature AVR-Pita$_{176}$ protein binds to the Pi-ta receptor in the cytoplasm of the invaded plant cell and initiates the defense response. The timing and mechanism for processing of AVR-Pita and its delivery into the plant cell cytoplasm have not yet been addressed.

genes. Differential rice varieties, especially those with single, mapped R genes, have helped in some cases to define *AVR* gene composition within pathotypes (Atkins et al 1967, Kiyosawa 1984, Yamada et al 1976). However, pathotype characterization has been limited by the occurrence of multiple *AVR* genes within single field isolates of the pathogen, by the inability of researchers to share pathogen strains among different rice-growing regions of the world, and by the imprecise knowledge of R gene composition in individual rice varieties.

DNA fingerprinting analyses using the dispersed, middle-repetitive DNA sequence MGR586 have shown that blast fungal populations in major rice-growing areas are predominantly asexual and composed of dozens of lineages, with each lineage apparently derived through clonal propagation from a common ancestor (Levy et al 1991, 1993, Xia et al 1993, Zeigler et al 1994). Clearly, pathotypes have evolved independently within different lineages—some lineages are associated with a single pathotype and others show a more complex pathotype structure. All lineages show uniform avirulence toward some R gene(s), raising the possibility that members of a given lineage may be limited in ability to modify *AVR* recognition corresponding to these R genes. Lineage exclusion breeding strategies based on pyramiding of "lineage-specific" R genes, effective against all strains of a particular clonal population, may

facilitate breeding durable resistance in rice (Levy et al 1993, Zeigler et al 1994). In any case, this simplified view of the rice pathogen population as sets of discrete lineages provides a genealogical framework for evaluating pathogen evolution and potential for variation.

The first examples of population analysis of *AVR-Pita* gene structure validate lineage structure as defined by neutral DNA markers such as MGR586. DNA gel blot analysis using the *AVR-Pita* gene as a probe determined that restriction fragment patterns were conserved within single genetic lineages of the fungus in Colombia (Montenegro-Chamorro 1997) and in the Philippines (Zeigler et al 1995). *AVR-Pita* restriction fragment patterns generally varied between lineages. Particular *AVR-Pita* restriction fragments correlated with avirulence and others correlated with virulence. A few lineages lacked homology to *AVR-Pita*.

Montenegro-Chamorro (1997) sequenced *AVR-Pita* alleles from Colombian lineages and demonstrated striking conservation of the *AVR-Pita* coding sequence relative to the gene from the Chinese rice pathogen O-137. Uniformly avirulent lineages SRL-1 and SRL-2 contain genes encoding AVR-Pita proteins that differ by only two or three amino acids from the O-137 protein. Virulent lineages SRL-3 and ALT-7 contain *avr-pita*$^{-}$ genes encoding proteins that uniformly have seven amino acid differences relative to AVR-Pita proteins that confer avirulence. Again, these avr-pita^{-} proteins differ from each other by only one or two amino acids. The *avr-pita*$^{-}$ genes also contain a putative protease motif that differs by a single allowed amino acid substitution within the motif region (Montenegro-Chamorro 1997, G.T. Bryan, S. Kang, and B. Valent, unpublished results). Thus, it is possible that evolution of *AVR-Pita* alleles may involve modifications of a protein resulting in preservation of metabolic function while altering its recognition as an avirulence factor.

An opportunity to test the lineage exclusion hypothesis

MGR-DNA fingerprinting of blast fungus in southern Brazil, Argentina, and Uruguay has identified sister lineages, BZ-A, ARG-1, and URU-1, to Colombian lineages SRL-1 and SRL-2, which contain an avirulent allele of *AVR-Pita* (M. Levy, A. Livore, S. Avila, and B. Valent, unpublished results). The Southern Cone lineages are statistically indistinguishable and more than 85% similar to SRL-1 and SRL-2 by MGR-DNA fingerprinting, and they retain the *AVR-Pita* gene structure (common RFLPs) and avirulence to *Pi-ta* characteristic of SRL-1 and SRL-2. Moreover, these Southern Cone lineages are the only components of the diverse blast population in each country that agressively infect tropical indica varieties, such as El Paso 144, constituting the major rice germplasm grown in this region. The lineage exclusion hypothesis predicts that introgression of *Pi-ta* into El Paso 144 and related varieties will protect these varieties from the only genetic families currently known to infect them. Transgenic strategies to test this hypothesis are under way.

The relationship between *Pi-ta* and *Pi-ta²*

The *Pi-ta²* gene, reported as being allelic or tightly linked to *Pi-ta* (Kiyosawa 1971, Rybka et al 1997), has a broader disease control spectrum than *Pi-ta*. We sequenced the *Pi-ta* allele from *Pi-ta²*-containing varieties Tetep, Reiho, and Katy, and found that these varieties all contain the resistant *Pi-ta* allele (Table 1). This is interesting in view of reports that *Pi-ta* is required for the function of *Pi-ta²* (Kiyosawa 1967, 1971, Rybka et al 1997, Silué et al 1993). A viable working hypothesis is that *Pi-ta²* consists of *Pi-ta* plus another R gene, presumably located in the chromosome 12 centromeric region.

Infection assays summarized in Table 2 support our sequencing results that *Pi-ta²* varieties contain *Pi-ta* (Table 1). Strain O-137 with *AVR-Pita* is avirulent on both *Pi-ta* and *Pi-ta²* rice varieties, but the spontaneous O-137 *avr-pita⁻* mutant CP3337 has gained the ability to infect both types of rice. This result suggests that *AVR-Pita* in strain O-137 is responsible for triggering resistance in *Pi-ta²* and in *Pi-ta* rice, and that O-137 lacks other *AVR* genes effective toward the tested *Pi-ta²* varieties. Another Chinese field isolate, O-135, lacks homology to the *AVR-Pita* gene and fails to trigger *Pi-ta*-mediated resistance in Yashiro-mochi and K1. However, O-135 is avirulent on *Pi-ta²* varieties, suggesting that it contains an *AVR* gene that does not occur in strain O-137. A spontaneous mutant of O-135, named CP753, was selected for virulence on *Pi-ta²*-containing variety Reiho. This mutant had gained the ability to infect all *Pi-ta²* varieties tested, while retaining avirulence toward C101A51 with a different R gene. Taken together, these results suggest that the broader spectrum *Pi-ta²* gene is a combination of at least two R genes: *Pi-ta* plus a second, presumably linked, R gene. These results underscore the difficulty in precisely defining the R gene composition of commonly used rice varieties using uncharacterized fungal pathogen strains.

Hybridization studies indicate that *Pi-ta* is single-copy gene in rice and does not identify homologues as candidate genes for *Pi-ta²*. Although sample genomic sequencing of 750–800 kb of the assembled *Pi-ta*-chromosomal region has not identified a *Pi-ta²* candidate, genetic distance will give a misleading indication of physical distance in the region of the centromere. Other R genes, including *Pi-4ᵃ(t)*, *Pi-4ᵇ(t)*, *Pi-6(t)*, and *Pi-12(t)*, have been mapped to the centromeric region of chromosome 12

Table 2. Avirulent fungus/mutant pairs identify *Pi-ta* and a second R gene in rice varieties with *Pi-ta²*.

Variety	O-137 *AVR-Pita*	CP3337[a] *avr-pita⁻*	*Pi-ta* present	O-135 *AVR-Reih*	CP753[b] *avr-reih⁻*	2nd R gene present
Yashiro-mochi[c]	AVR	vir	+	vir	vir	–
Tetep[c]	AVR	vir	+	AVR	vir	+
C101A51	vir	vir	–	AVR	AVR	–
YT16	vir	vir	–	vir	vir	–

[a]CP3337 is a spontaneous mutant of O-137 that has lost the *AVR-Pita* function. [b]CP753 is a spontaneous mutant of O-135 recovered as a rare lesion on *Pi-ta²*-containing variety Reiho. [c]Additional *Pi-ta*-containing variety K1 has the same specificity as Yashiro-mochi and additional *Pi-ta²*-containing varieties Tadukan, Reiho, and Katy show the same specificity as Tetep.

(Inukai et al 1994, Yu et al 1996, Zheng et al 1996). The *Pi-ta*^{2}-containing variety Katy, from the southern United States, has been reported to contain a tightly linked cluster of at least seven R genes that map in the same region as *Pi-ta* and *Pi-ta*^{2} (Chao et al 1999, Moldenhauer et al 1992). Genome sequencing now under way in rice will facilitate the cloning and characterization of all R genes in this complex chromosomal region.

Future prospects

Cloning *Pi-ta* will facilitate its incorporation into agronomically useful crop varieties, either through breeding using marker-assisted selection or through transgenic strategies. Transgenic strategies are especially valuable for R genes linked to centromeres, since reduced recombination near centromeres will exacerbate linkage drag associated with R gene introgression by classical breeding. Understanding the field biology of the *AVR-Pita* resistance trigger within lineages of the rice blast fungus is already identifying germplasm that would benefit from pyramiding *Pi-ta* with native R genes. The conservation of the *AVR-Pita* structure among field isolates of the pathogen from China and Colombia contrasts with the genetic instability we see for the telomeric *AVR-Pita* locus in laboratory studies (Orbach et al 2000). The full impact of telomeric location of *AVR* genes for the dynamics of rice blast disease in the field remains to be determined.

This research provides the first direct evidence concerning molecular interactions between an *AVR* gene and its corresponding R gene in the rice blast system. The current hypothesis is that a mature AVR-Pita protease triggers *Pi-ta*-mediated resistance by binding directly to the cytoplasmic Pi-ta receptor protein. Future directions include biochemical demonstration of protease activity for AVR-Pita, defining its natural substrate, and understanding its role in the infection process. If protease activity is confirmed for AVR-Pita, it is important to determine what relationship this activity has to the protein's avirulence function in triggering *Pi-ta*-mediated resistance. The *AVR-Pita* system presents an opportunity for understanding the presumed dual functions an *AVR* gene plays in the lifestyle of the pathogen, in contributing to the pathogenicity process and in triggering R-gene-mediated defense responses in the presence of its corresponding R gene. It is also of interest to know when and how the AVR-Pita protein enters the plant cell during the infection process, and where and how this protein gets processed. Important challenges lie in defining the complete signal transduction pathway downstream of the Pi-ta protein. Understanding the molecular basis for *AVR* gene triggering of resistance will lead to the development of novel strategies for broad-spectrum, durable disease control using the native plant defense systems.

References

Atkins JG, Robert AL, Adair CR, Goto K, Kozaka T, Yanagida R, Yamada M, Matsumoto S. 1967. An international set of rice varieties for differentiating races of *Piricularia oryzae*. Phytopathology 57:297-301.

Bent AF. 1996. Plant disease resistance genes: function meets structure. Plant Cell 8:1757-1771.

Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B. 2000. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12:2033-2045.

Chao CT, Moldenhauer KAK, Ellingboe AH. 1999. Genetic analysis of resistance/susceptibility in individual F_3 families of rice against strains of *Magnaporthe grisea* containing different genes for avirulence. Euphytica 109:183-190.

Dong F, Miller JT, Jackson SA, Wang GL, Ronald PC, Jiang J. 1998. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc. Natl. Acad. Sci. USA 95:8135-8140.

Flor HH. 1971. Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. 9:275-296.

Hammond-Kosack KE, Jones JDG. 1997. Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:575-607.

Inukai T, Nelson RJ, Zeigler RS, Sarkarung S, Mackill DJ, Bonman JM, Takamure I, Kinoshita T. 1994. Allelism of blast resistance genes in near-isogenic lines of rice. Phytopathology 84:1278-1283.

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J. 19:4004-4014.

Kiyosawa S. 1967. Inheritance of resistance of the rice variety Pi No. 4 to blast. Jpn. J. Breed. 17:165-172.

Kiyosawa S. 1971. Genetical approach to the biochemical nature of plant disease resistance. Jpn. Agric. Res. Q. 6:73-80.

Kiyosawa S. 1984. Establishment of differential varieties for pathogenicity test of rice blast fungus. Rice Genet. Newsl. 1:95-97.

Levy M, Correa-Victoria FJ, Zeigler RS, Xu S, Hamer JE. 1993. Genetic diversity of the rice blast fungus in a disease nursery in Colombia. Phytopathology 83:1427-1433.

Levy M, Romao J, Marchetti MA, Hamer JE. 1991. DNA fingerprinting with a dispersed repeated sequence resolves pathotype diversity in the rice blast fungus. Plant Cell 3:95-102.

Moldenhauer KAK, Bastawisi AO, Lee FN. 1992. Inheritance of resistance in rice to races IB-49 and IC-17 of *Pyricularia grisea* rice blast. Crop Sci. 32:584-588.

Montenegro-Chamorro MV. 1997. Allelic diversity of an avirulence gene in Colombian field isolates of the rice blast fungus. MS thesis. Purdue University, West Lafayette, Ind., USA.

Nakamura S, Asakawa S, Ohmido N, Fukui K, Shimizu N, Kawasaki S. 1997. Construction of an 800-kb contig in the near-centromeric region of the rice blast resistance gene *Pi-ta*^{2} using a highly representative rice BAC library. Mol. Gen. Genet. 254:611-620.

Nitta N, Farman ML, Leong SA. 1997. Genome organization of *Magnaporthe grisea*: integration of genetic maps, clustering of transposable elements and identification of genome duplications and rearrangements. Theor. Appl. Genet. 95:20-32.

Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. 2000. The fungal avirulence gene *AVR-Pita* determines efficacy for the rice blast resistance gene *Pi-ta*. Plant Cell 12:2019-2032.

Rossman AY, Howard RJ, Valent B. 1990. *Pyricularia grisea*, the correct name for the rice blast fungus. Mycologia 82:509-512.

Rybka K, Miyamoto M, Ando I, Saito A, Kawasaki S. 1997. High resolution mapping of the *indica*-derived rice blast resistance genes. II. *Pi-ta*^{2} and *Pi-ta* and a consideration of their origin. Mol. Plant-Microbe Interact. 10:517-524.

Silué D, Notteghem JL, Tharreau D. 1992. Evidence for a gene-for-gene relationship in the *Oryza sativa-Magnaporthe grisea* pathosystem. Phytopathology 82:577-580.

Silué D, Tharreau D, Notteghem JL. 1993. Genetic control of *Magnaporthe grisea* avirulence to rice cultivar Pi-No4 and K1. Abstracts, International Congress of Plant Pathology, 6th, Montreal, Canada. p 159.

Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS. 1996. Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). Proc. Natl. Acad. Sci. USA 93:6163-6168.

Sweigard JA, Valent B, Orbach MJ, Walter AM, Rafalski A, Chumley FG. 1993. Genetic map of the rice blast fungus *Magnaporthe grisea*. In: O'Brien SJ, editor. Genetic maps. Plainview, NY (USA): Cold Spring Harbor Laboratory Press. p 3.112-3.117.

Tatsumi H, Murakami S, Tsuji RF, Ishida Y, Murakami K, Masaki A, Kawabe H, Arimura H, Nakano E, Motai H. 1991. Cloning and expression in yeast of a cDNA clone encoding *Aspergillus oryzae* neutral protease II, a unique metalloprotease. Mol. Gen. Genet. 228:97-103.

Wu KS, Martínez C, Lentini Z, Tohme J, Chumley FG, Scolnik PA, Valent B. 1996. Cloning a blast resistance gene by chromosome walking. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Los Baños (Philippines): International Rice Research Institute. p 669-674.

Xia JQ, Correll JC, Lee FN, Marchetti MA, Rhoads DD. 1993. DNA fingerprinting to examine microgeographic variation in the *Magnaporthe grisea* (*Pyricularia grisea*) population in two rice fields in Arkansas. Phytopathology 83:1029-1035.

Yamada M. 1985. Pathogenic specialization of rice blast fungus in Japan. Jpn. Agric. Res. Q. 19:178-183.

Yamada M, Kiyosawa S, Yamaguchi T, Hirano T, Kobayashi T, Kushibuchi K, Watanabe S. 1976. Proposal of a new method for differentiating races of *Pyricularia oryzae* Cavara in Japan. Ann. Phytopathol. Soc. Jpn. 42:216-219.

Yu ZH, Mackill DJ, Bonman JM, McCouch SR, Guiderdoni E, Notteghem JL, Tanksley SD. 1996. Molecular mapping of genes for resistance to rice blast (*Pyricularia grisea* Sacc.). Theor. Appl. Genet. 93:859-863.

Zeigler RS, Cuoc LX, Scott RP, Bernardo MA, Chen DH, Valent B, Nelson RJ. 1995. The relationship between lineage and virulence in *Pyricularia grisea* in the Philippines. Phytopathology 85:443-451.

Zeigler RS, Leong SA, Teng PS, editors. 1994. Rice blast disease. Wallingford (UK): CAB International. 626 p.

Zheng KL, Zhuang JY, Lu J, Qian HR, Lin HX. 1996. Identification of DNA markers tightly linked to blast resistance genes in rice. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct 1995, Manila, Philippines. Los Baños (Philippines): International Rice Research Institute. p 565-569.

Notes

Authors' addresses: B. Valent, G.T. Bryan, L. Farrall, S.A. McAdams, K.N. Faulk, DuPont Agricultural Enterprise, P.O. Box 80402, Wilmington, DE 19880-0402, USA; G.T. Bryan, AgResearch Limited, Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North, New Zealand; Y. Jia, Dale Bumpers National Rice Research Center, Stuttgart, AR 72160, USA; M. Levy, Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Molecular signaling in disease resistance of rice

K. Shimamoto, A. Takahashi, and T. Kawasaki

We study signaling pathways involved in the interaction of rice and the rice blast fungus, *Magnaporthe grisea*, by using molecular and genetic methods. We recently demonstrated that the rice homologue of the small GTPase Rac is an important molecular switch for resistance gene–mediated disease resistance against rice blast infection. Furthermore, it was shown to activate the production of reactive oxygen species as well as cell death, both of which are often observed during the resistance response to avirulent pathogens in many plant systems. In another approach, we study lesion-mimic mutants of rice showing increased resistance to the rice blast fungus. Biochemical analysis of these mutants allowed us to identify genes that may play an important role in disease resistance of rice.

A better understanding of disease resistance reactions occurring in plants will help us develop crops with improved resistance to major pathogens. Rice is no exception. Blast, bacterial blight, sheath blight, and tungro virus cause billions of dollars in losses. Recently, major progress has been made in the molecular biological study of rice-pathogen interactions (reviewed in Ronald 1997, Wang and Leung 1999). Genes with resistance against bacterial blight (Song et al 1995, Yoshimura et al 1998) and rice blast fungus (Wang et al 1999, Jia et al 2000) were isolated and lesion-mimic mutants were recently characterized at the molecular level (Takahashi et al 1999, Yin et al 2000). These studies suggest that we have begun to understand the molecular biology of disease resistance in rice and that rice will become a useful model plant for studies in plant-microbe interactions in the near future. Despite this recent progress in our understanding of the molecular biology of rice-pathogen interactions, little is known about the signaling pathways underlying disease resistance in rice.

We have been taking molecular biological approaches to study the signaling pathways involved in rice-blast interactions and some of the recent results are described here.

Role of the small GTPase Rac in disease resistance of rice

OsRac gene family in rice

Rac belongs to the Rho subfamily of small GTPase and constitutes a gene family in plants (Valster et al 2000). Although some of these genes were shown to play a role in the control of pollen tube growth (Lin and Yang 1997), the functions of most of these genes are not known. Because Rac plays an important role in the regulation of the NADPH oxidase in mammalian phagocytic cells, we sought rice expressed sequence tags (ESTs) having homology with human Rac and found three such sequences called OsRac (Kawasaki et al 1999). They had deduced amino acid sequences approximately 60% identical with those of human Rac proteins. More recently, we isolated four more genes in rice that are homologous with the human Rac genes. Rice has four groups of Rac genes, OsRac1–4, and OsRac2 has four closely related members. Thus, in total, rice has at least seven Rac-related genes and most are expressed in both leaves and roots.

ROS production by the constitutive active OsRac1

The small GTPase such as Rac takes two forms within a cell. The GTP-bound form is an active form and it transmits signals to downstrean effectors and the GDP-bound inactive form is not active. These two forms cycle in the cell depending on the presence or absence of various signals (Fig. 1).

We made a constitutive active form of OsRac1 by substituting glycine at position 19 with valine and fused it with the CaMV 35S promoter (Fig. 2). The construct was introduced into seed-derived calli of the wild type (cv. Kinmaze) and a lesion-mimic mutant of rice, *sl* (Sekiguchi lesion) (Kiyosawa 1970), by *Agrobacterium*-mediated transformation. In the transformed cell lines of the wild type and the *sl* mutant, H_2O_2 production was demonstrated by staining with diaminobenzidine (DAB). The ob-

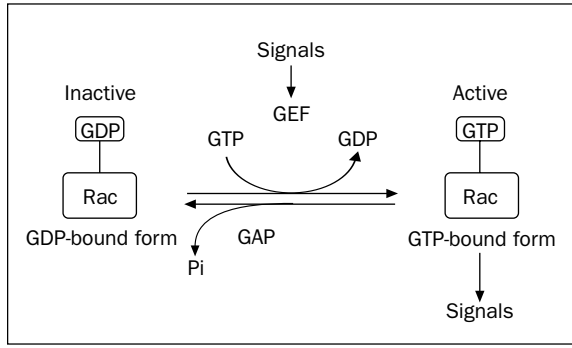


Fig. 1. Two forms of the small GTPase Rac within a cell. The inactive GDP (guanosine diphosphate)-bound form is converted to the active GTP (guanosine triphosphate)-bound form by the help of GEF (guanine nucleotide exchange factor). The active form is changed to the inactive form by the help of GAP (GTPase-activating protein).

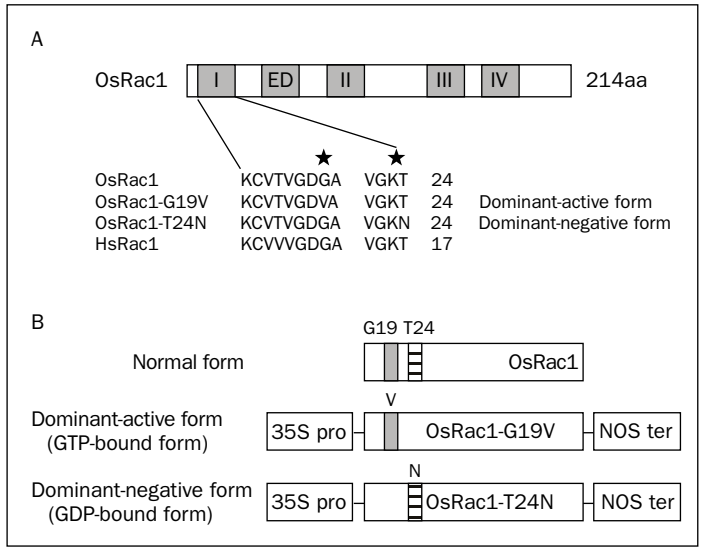


Fig. 2. Construction of vector plasmids carrying constitutive active and dominant-negative OsRac1. (A) Constitutive active and dominant-negative OsRac1 can be produced by single amino acid changes in the conserved region I required for GTPase activity. The ED (effector domain) and II-IV regions are highly conserved between rice and human Rac genes. (B) Constitutive active and dominant-negative OsRac1 used for rice transformation. OsRac1 was fused with the CaMV 35S promoter and NOS terminator for constitutive expression in rice plants and cells.

served H_2O_2 production was inhibited by diphenylene iodonium (DPI), an inhibitor of the neutrophil NADPH oxidase. These findings suggest that an NADPH oxidase similar to the neutrophil enzyme is involved in Rac-induced H_2O_2 production. Furthermore, DAB staining of leaf sheath cells of transgenic wild-type plants showed H_2O_2 production, which appeared to be localized at the intercellular space.

Cell death induced by the constitutive active OsRac1

Terminal deoxynucleotidyltransferase-mediated UTP end labeling (TUNEL) signals indicative of nuclear DNA cleavage were observed in the transformed *sl* cells but not in untransformed *sl* cells. Electron microscopy of the transformed *sl* cells indicated that cell death occurs in the transformed cells and that the observed cell death exhibits a set of morphological changes found in apoptosis in mammalian cells (Shimamoto et al 1999). Furthermore, necrotic lesions were found in the leaves of transgenic wild-type plants at a young stage. During maturation, discrete lesions developed in the leaves and they were frequently observed at the junction of the blade and the sheath of the leaves. These results indicate that the constitutive active OsRac1 can induce cell death in transgenic rice plants as well as in cultured cells.

Suppression of ROS production and cell death by the dominant-negative OsRac1

We produced a dominant-negative form of OsRac1 by changing threonine at the 24th position to asparagine, fused with the 35S promoter (Fig. 3), and introduced it into the *sl* mutant. The mutant dominant-negative form is able to suppress activities of all Rac genes within the same cell. In untransformed *sl* cell cultures, H_2O_2 production was induced by calyculin A, a protein phosphatase inhibitor, and the induction was inhibited by DPI. In contrast, carbonic anhydrase (CA)-induced H_2O_2 production was inhibited in the *sl* cells transformed with OsRac1-T24N, indicating that OsRac was required for activation of H_2O_2 production in the *sl* cells. To examine whether the dominant-negative OsRac1 suppresses lesion formation in *sl* plants, leaves of untransformed and transformed *sl* plants were inoculated with the rice blast fungus, *Magnaporthe grisea*. The results showed that lesion formation induced by the fungus was strongly inhibited in transformed plants. These results strongly suggest the role of OsRac in induction of ROI production and cell death in rice.

Role of OsRac1 for R gene–mediated blast resistance in rice

To test whether OsRac1 is involved in R gene–mediated disease resistance in rice, we chose the rice blast fungus *M. grisea,* since a series of R genes corresponding to several races has been identified (Notteghem et al 1994) and a gene-for-gene relationship has been established in this plant-pathogen interaction (Valent and Chumley 1994). More recently, *Pi-b* and *Pi-ta* genes, the R genes for blast resistance, were isolated and shown to be members of the NBS-LRR class of plant R genes (Wang et al 1999, Jia et al 2000).

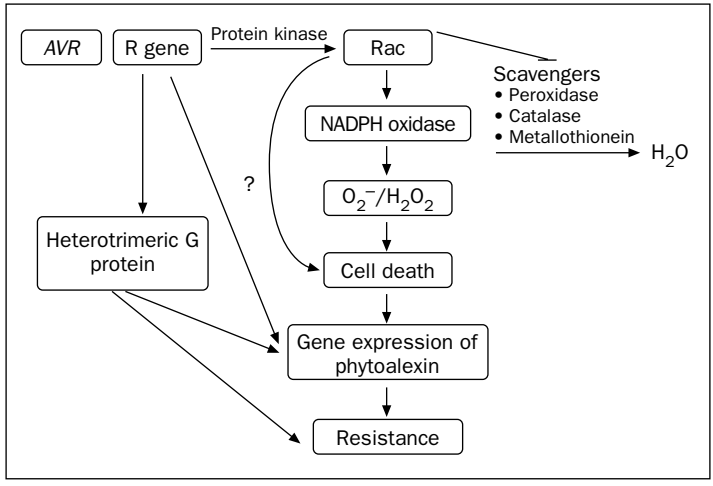


Fig. 3. Model of the signaling pathways for R gene–mediated disease resistance of rice.

Japonica rice variety Kinmaze, which was used to produce transgenic plants, carries the *Pi-a* resistance gene that is incompatible with race 31 of the blast fungus. We infected the leaf blade of transgenic rice plants expressing the constitutive active OsRac1 with race 7 of the blast fungus by the punch infection method. Because race 7 is compatible with var. Kinmaze, infected leaves of untransformed control plants developed typical disease symptoms. In contrast, transgenic plants were highly resistant or showed a reduced level of symptoms. These results indicate that rice plants transformed with the constitutive active OsRac1 acquire resistance to a compatible race of the rice blast fungus and exhibit hypersensitive resistance (HR)-like responses.

We next asked whether HR caused by an avirulent race of the rice blast fungus is suppressed by expression of the dominant-negative OsRac1. The leaf blades of transgenic plants expressing the dominant-negative OsRac1 were infected with an avirulent race, 31, of the blast fungus. Typical HR accompanied by cell death was observed in untransformed control plants. In contrast, in transgenic lines examined, HR was clearly suppressed in the leaf blade and suppression was observed in repeated experiments, indicating that the dominant-negative OsRac1 suppresses R gene-specific resistance against rice blast fungus. Taken together, our results strongly suggest that OsRac1 is an essential component of a resistance-signaling pathway for the blast fungus acting downstream of R genes.

Activation of phytoalexin production in transgenic rice expressing the constitutive active OsRac1

We examined the amounts of momilactone A, a major phytoalexin of rice (Cartwright et al 1977), in the leaves of transgenic rice. In all seven transgenic lines examined, the levels of momilactone A were highly elevated and the increases were 19–180-fold higher than the levels of the untransformed control plants. Transgenic lines expressing the dominant-negative OsRac1 did not show increased levels of momilactone A. These results indicate that constitutive active OsRac1 activates phytoalexin synthesis to confer resistance to the blast fungus.

Alteration of defense-related gene expression in transgenic rice expressing the constitutive active OsRac1

To further elucidate the basis for the blast resistance observed in transgenic rice expressing the constitutive active OsRac1, we next examined the expression of two genes that are closely associated with disease resistance in rice. Expression of D9 was highly activated in all lines examined. D9 was isolated from lesion-mimic mutants of rice (Takahashi et al 1999) resistant to rice blast and found to be up-regulated in these mutants. Its deduced amino acid sequence is highly homologous with that of terpenoid synthase and may be involved in momilactone A synthesis (unpublished results). One of the rice peroxidase genes, POX22.3, which was originally isolated as a gene activated during the resistance reactions of rice against *Xanthomonas oryzae* pv. *oryzae* (Chittoor et al 1997), was shown to be strongly down-regulated in the transgenic rice expressing the constitutive active OsRac1 that acquired resistance to blast disease. This observation suggests that OsRac1 may also function to down-regulate H_2O_2 scav-

enging activities at the level of transcription and that this function may be independent of its role in the activation of NADPH oxidase.

Model of the signaling pathways for R gene–mediated disease resistance of rice

Figure 3 shows a model for the signaling pathway for R gene–mediated disease resistance of rice. All the defense reactions are initiated by recognition of the *AVR* gene product or some signals derived from the pathogens. Very recently, direct interaction of the products of the blast resistance gene, *Pi-ta*, and the corresponding *AVR* gene, *AVR-Pi-ta*, was demonstrated (Jia et al 2000); however, whether the observed direct interaction between products of the R gene and *AVR* gene is a general phenomenon in rice-pathogen interactions remains to be studied. After perception of the signal from the pathogen, a cascade of protein phosphorylation occurs and it leads to activation of NADPH oxidase through the small GTPase Rac. The biochemical nature of the hypothetical NADPH oxidase is not known. Rac amplifies the signal from the pathogen to activate ROS production and cell death. These two cellular events initiate activation of defense genes and production of phytoalexin and eventually lead to disease resistance. The second function of Rac in disease resistance may be to transiently suppress expression of scavengers at the transcriptional level to strengthen the ROS signal. Since this suppression is quickly released after oxidative burst ceases, cells will not be damaged for a long time after the perception of signals from pathogens. It is also possible that R gene products activate some of the defense genes independent of Rac action.

Apart from the Rac signaling pathway, our recent studies indicate that heterotrimeric G protein appears to function downstream of the R gene. It may directly activate various defense genes or it may contribute to resistance by other unknown pathways.

Lesion-mimic mutants of rice

Mutants that show spontaneous cell death in the absence of pathogens are called disease lesion mimics and have been known to occur in several cereals including maize (Johal et al 1995), rice (Kiyosawa, 1970), and barley (Wolter et al 1993). They are recessive or dominant, and almost all mutants show a unique lesion phenotype with respect to the size of the lesion, timing of lesion formation, and color of the lesion (Johal et al 1995). Mutant phenotypes are often influenced by environmental conditions such as light and temperature (Johal et al 1995).

cdr lesion-mimic mutants of rice

We screened 93 rice mutants that exhibited spontaneous lesions for resistance to blast fungus infection and found eight mutants with increased resistance to blast infection. We called these mutants *cdr* (*c*ell *d*eath and *r*esistance). Characteristics of three such mutants, *cdr1–3*, are summarized in Table 1. The three mutants showed many small brownish lesions that were uniformly spread over the entire leaf surface (Fig. 4). In *cdr1*, lesions became visible around 20 d after sowing. In the *cdr2* mutant, many

Table 1. Characteristics of lesion-mimic mutants analyzed.

Line	Response to blast[b]		Genetic data	Timing[c] (d)
	Lesion[a]	Race 007		
WT	–	S	–	–
cdr1	–	S	Recessive	<20
	+	R		
cdr2	–	S	Recessive	<50
	+	R		
Cdr3	–	R	Recessive	<20[d]
	+	R		

[a]+ indicates plants with lesion and – shows younger plants in which lesions were not visible. [b]R = resistant, S = susceptible. [c]Timing indicates the number of days after sowing when lesions became visible. [d]Homozygous plants.

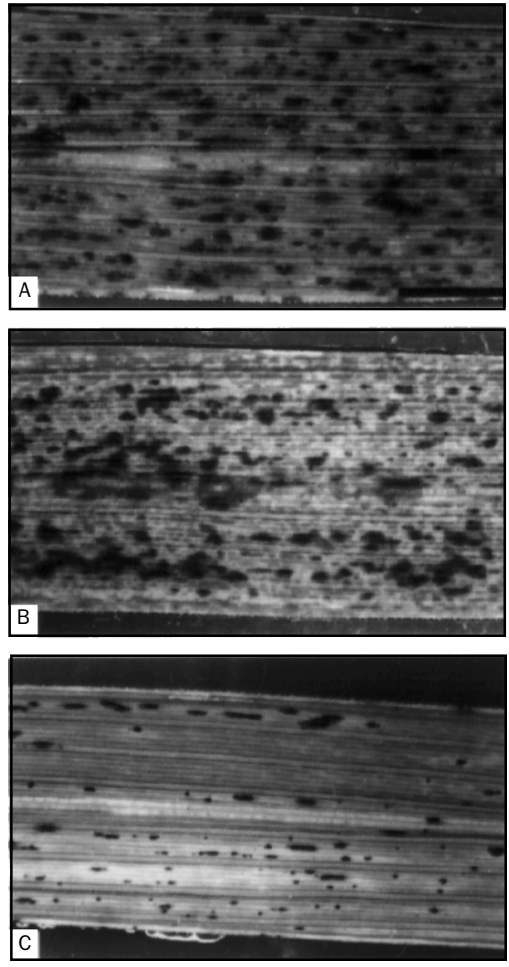


Fig. 4. Phenotypes of the *cdr* mutants.

lesions appeared over the entire leaf around 50 d after sowing and often contained yellowish lesions overlapped with brownish lesions. *Cdr3* was a dominant mutation and, when homozygous, plants exhibited numerous lesions approximately 20 d after sowing.

Constitutive activation of defense reactions in the *cdr* mutants

Several histological markers for the defense response have been observed in plants challenged with avirulent pathogens (reviewed in Hammond-Kosack and Jones 1996). We first examined autofluorescence and callose deposition, both of which are well-established cellular markers observed in plants exhibiting disease resistance. Autofluorescence is strongly correlated with resistance to the blast fungus (Koga 1994). Autofluorescence was detected in leaf sheath cells of the three *cdr* mutants of rice, which were undergoing cell death in the absence of pathogens, whereas no signal was detected in the wild type. Callose deposition was detected in leaf blade cells, which were dying or dead, by aniline blue staining in the three mutants but not in the wild type. These results indicate that these two histological markers for the defense response were constitutively activated in the leaves of the three *cdr* mutants of rice.

We next analyzed the expression of two defense-related genes, *PBZ1* and *PR1* (Fig. 5). The *PBZ1* gene encodes an intracellular PR protein and is activated by treatment of rice plants with probenazole (3-allyloxy-1,2-benzisothiazole-1,1-dioxide),

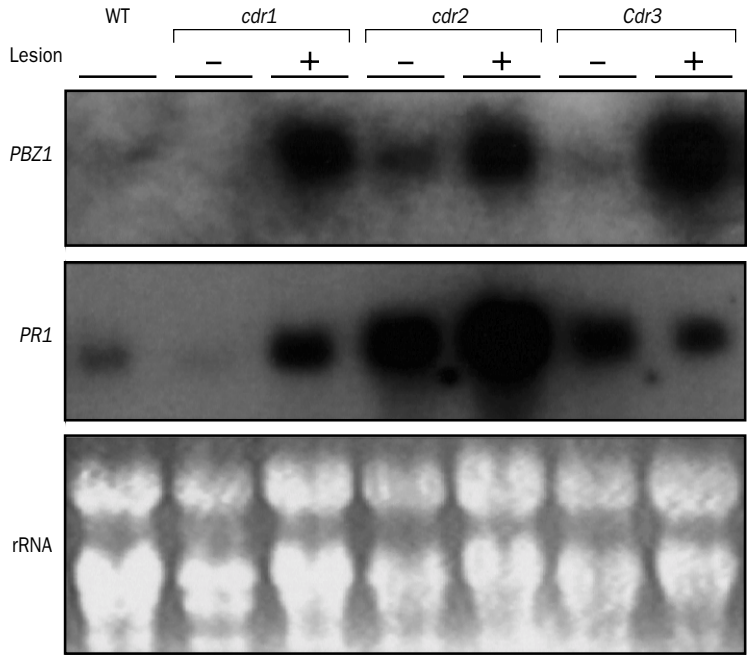


Fig. 5. Expression of *PBZ1* and *PR1* genes in the *cdr* mutants of rice. The (–) indicates leaves without lesions and (+) indicates leaves with lesions.

an effective inducer of host resistance against blast infection (Midoh and Iwata 1996). Expression of *PBZ1* and *PR1* was highly activated in leaves of the three *cdr* mutants when the lesion was developed. In contrast, very low expression of these two genes was detected in leaves of the wild type. In mutant leaves showing no visible lesions, *PBZ1* expression was much lower. Thus, *PBZ1* expression was closely correlated with development of the lesion. These results indicate that defense-related genes are highly induced in leaves of the *cdr* mutants exhibiting lesion formation.

The level of momilactone A in the leaves of the *cdr* mutants that showed lesions was increased approximately 100-fold in *cdr1* and *Cdr3* and more than 400-fold in *cdr2*. In leaves that showed only a few lesions, much lower levels of momilactone A were detected. These results clearly indicate that the level of this major phytoalexin of rice in these mutants is much higher than that found in the wild type.

Enhanced production of ROS in response to calyculin A, an inhibitor of protein phosphatase, in the *cdr* mutants

It was previously shown that calyculin A, an inhibitor of protein phosphatase 1, induces ROS in cell suspension cultures of rice (Kuchitsu et al 1995). Thus, we examined its effect on H_2O_2 production in cell suspension cultures of the mutants and the wild type. The wild-type cells generated H_2O_2 after treatment with calyculin A. The observed H_2O_2 production by calyculin A was inhibited by DPI, a potent inhibitor of the NADPH oxidase, suggesting that the plasma membrane NADPH oxidase was involved in the H_2O_2 production in the cell cultures of rice. When suspension cell cultures of *cdr1* and *cdr2* were treated with calyculin A, increased production of H_2O_2 was observed and this enhanced production of H_2O_2 was inhibited by DPI. No H_2O_2 production was observed if they were not treated with the protein phosphatase inhibitor. In contrast to *cdr1* and *cdr2*, no enhancement of H_2O_2 production was observed in suspension cell cultures of *Cdr3*.

The results obtained from the studies using the cell cultures suggest that the two mutants, *cdr1* and *cdr2*, have alterations in the signaling cascade leading to activation of NADPH oxidase. Since these two mutations are both recessive, they may be mutations of factors that normally suppress NADPH oxidase in the absence of signals from pathogens. In contrast, *Cdr3* may have alterations downstream of NADPH oxidase.

Identification of proteins whose phosphorylation levels were increased in *cdr* mutants

To study proteins whose phosphorylation levels were increased by the *cdr* mutations, we compared profiles of ^{32}P-labeled protein spots separated by two-dimensional gel electrophoresis. We identified four protein spots that were phosphorylated by calyculin A treatment and the degree of phosphorylation was increased in the *cdr* mutants. One spot was shown to be more phosphorylated in the *cdr1* than in the wild type. For the other three protein spots, phosphorylation increased equally in both *cdr1* and *cdr2* compared with the wild type. We attempted to sequence these polypeptides after protease treatment and obtained information on the amino acid sequences. Based on the

limited amino acid sequence information, we were able to identify one protein species and obtained its cDNA. Detailed analysis of this protein is in progress and we plan to identify the possible function of this protein in disease resistance and induction of cell death in rice.

References

Cartwright D, Langcake P, Pryce DP, Leworthy DP, Ride JP. 1977. Chemical activation of host defence mechanisms as a basis for crop protection. Nature 361:153-156.

Chittoor JM, Leach JE, White FF. 1997. Differential induction of a peroxidase gene family during infection of rice by *Xanthomonas oryzae* pv. *oryzae*. Mol. Plant-Microbe Interact. 10:861-871.

Hammond-Kosack KE, Jones JDG. 1996. Resistance gene-dependent plant defense responses. Plant Cell 8:1773-1791.

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. 2000. Direct interaction of resistance gene and avirulence gene products confer rice blast resistance. EMBO J. 19:4004-4014.

Johal GS, Hulbert SH, Briggs SP. 1995. Disease lesion mimics of maize: a model for cell death in plants. BioEssays 17:685-692.

Kawasaki T, Henmi K, Ono E, Hatakeyama S, Iwano M, Satoh H, Shimamoto K. 1999. The small GTP-binding protein Rac is a regulator of cell death in plants. Proc. Natl. Acad. Sci. USA 96:10922-10926.

Kiyosawa S. 1970. Inheritance of a particular sensitivity of the rice variety, Sekiguchi Asahi, to pathogens and chemicals, and linkage relationship with blast resistance genes. Bull. Natl. Inst. Agric. Sci. (Jpn.) Ser. D. Physiol. Genet. 21:61-71.

Koga H. 1994. Hypersensitive death, autofluorescence and ultrastructural changes in cells of leaf sheaths of susceptible and resistant near-isogenic lines of rice *(Pi-z^t)* in relation to penetration and growth of *Pyricularia oryzae*. Can. J. Bot. 72:1463-1477.

Kuchitsu K, Kosaka H, Shiga T, Shibuya N. 1995. EPR evidence for generation of hydroxyl radical triggered by *N*-acetylchitooligosaccharide elicitor and a protein phosphatase inhibitor in suspension-cultured rice cells. Protoplasma 188:138-142.

Lin Y, Yang Z. 1997. Inhibition of pollen tube growth elongation by microinjected anti-Rop 1Ps Ps antibodies in the control of tip growth. Plant Cell 9:1647-1659.

Midoh N, Iwata M. 1996. Cloning and characterization of a probenazole-inducible gene for an intracellular pathogenesis-related protein in rice. Plant Cell Physiol. 37:9-18.

Notteghem JL, Tharreau D, Silue D, Roumen E. 1994. Present knowledge of rice resistance genetics and strategies for *Magnaporthe grisea* pathogenicity and avirulence gene. In: Ziegler RS, Leong SA, Teng PS, editors. Rice blast disease.Wallingford (UK): CAB International. p 155-165.

Ronald PC. 1997. The molecular basis of disease resistance in rice. Plant Mol. Biol. 35:179-186.

Shimamoto K, Takahashi A, Henmi K, Ono E, Hatakeyama S, Iwano M, Kawasaki T. 1999. Programmed cell death in plants. Plant Biotechnol. 16:49-53.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald PC. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804-1806.

Takahashi A, Kawasaki T, Henmi K, Shii K, Kodama O, Satoh H, Shimamoto K. 1999. Lesion-mimic mutants of rice with alterations in early signaling events of defense. Plant J. 17:535-545.

Valent B, Chumley FG. 1994. Avirulence genes and mechanisms of genetic instability in the rice blast fungus. In: Ziegler RS, Leong SA, Teng PS, editors. Rice blast disease. Wallingford (UK): CAB International. p 111-134.

Valster AH, Hepler PK, Chernoff J. 2000. Plant GTPases: the Rhos in bloom. Trends Cell Biol. 10:141-146.

Wang GL, Leung H. 1999. Molecular biology of host-pathogen interactions in rice diseases. In: Shimamoto K, editor. Molecular biology of rice. Tokyo (Japan): Springer. p 201-232.

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T. 1999. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J. 19:55-64.

Wolter M, Hollricher K, Salamini F, Schulze-Lefert P. 1993. The *mlo* resistance alleles to powdery mildew infection in barley trigger a developmentally controlled defence mimic phenotype. Mol. Gen. Genet. 239:122-128.

Yin Z, Chen J, Zeng L, Goh M, Leung H, Khush GS, Wang GL. 2000. Characterizing rice lesion mimic mutants and identifying a mutant with broad-spectrum resistance to rice blast and bacterial blight. Mol. Plant-Microbe Interact. 13:869-876.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T. 1998. Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc. Natl. Acad. Sci. USA 17:1663-1668.

Notes

Authors' address: Laboratory of Plant Molecular Genetics, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0101, Japan, phone: +81 743 72 5500, fax: +81 743 72 5501, E-mail: simamoto@bs.aist-nara.ac.jp.

Acknowledgments: We thank Megumi Iwano, Hikaru Satoh, Katsuhiko Shii, Morifumi Hasegawa, and Osamu Kodama for collaboration and Kenji Henmi, Eiichiro Ono, Hann Ling Wong, and Satoru Hatakeyama for various contributions to the work described here. We also thank members of the Plant Molecular Genetics Lab at NAIST for numerous suggestions.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Structure, function, and evolution of disease resistance genes in rice

Guo-Liang Wang, Hei Leung, and P. Ronald

The recent cloning and characterization of several rice genes with resistance to pathogens represent a breakthrough in our understanding of the molecular basis of disease resistance and also provide a starting point for dissecting the resistance pathway in rice. The first resistance gene cloned in rice was *Xa21*, a gene introgressed from the wild rice *Oryza longistaminata*. It encodes a putative receptor-like kinase consisting of leucine-rich repeats (LRRs) in the extracellular domain and serine/threonine kinase in the intracellular domain. Sequence analysis of seven members of the gene family at the locus suggests that duplication, recombination, and transposition have occurred during the evolution of this gene family. Experiments with a truncated member indicate that the LRR domain determines race-specific recognition and is subject to adaptive evolution. To identify additional components in the *Xa21*-mediated resistance pathway, both the yeast two-hybrid screen and mutagenesis approaches are being used. Several defense-related genes were found to interact with the *Xa21* protein in yeast when the kinase domain was used in the screen. Using diepoxybutane and fast-neutron mutagenesis, we recovered 31 mutants that have changed from resistant to fully susceptible (10) or partially susceptible (21) to nine races of the bacterial blight pathogen in the Philippines. All fully susceptible mutants showed changes at the *Xa21* locus as detected by polymerase chain reaction and Southern hybridization. For the partially susceptible mutants, no detectable changes were found at the *Xa21* locus, suggesting that these mutations occur at other loci controlling the *Xa21*-mediated defense pathway.

The use of resistant cultivars is the most preferred method for disease control in crop plants. This method relies mainly on the incorporation of resistance genes from wild species or traditional cultivars into improved cultivars. In recent years, the successful cloning of about 25 disease resistance genes has dramatically advanced our understanding of the molecular basis of disease resistance and provided new ways to engineer resistant crop plants (reviewed by Staskawicz et al 1995, Baker et al 1997, Ellis

et al 2000). Sequence analysis of the predicted proteins reveals that resistance genes of diverse origin and pathogen specificity share similar structural motifs, indicative of the common surveillance strategy adopted by plant species to detect invading pathogens. Based on the similarity in predicted protein structures, the cloned resistance genes can be grouped into four main classes. The first class includes the tomato gene *Pto*, which confers resistance to *Pseudomonas syringae* pv. *tomato* containing the avirulence gene *avrPto* (Martin et al 1993). *Pto* encodes a serine/threonine kinase that interacts with the *avrPto* gene product of the pathogen (Scofield et al 1996, Tang et al 1996). Genes with resistance against diverse pathogens from *Arabidopsis* (*RPS2*, *RPP5*, and *Rpm1*), potato (*Rx*), rice (*Xa1*, *Pib*, and *Pita*), tobacco (*N*), flax (*L6* and *M*), and tomato (*Prf*, *Bs2*, *Sw5*, *Mi*, and *I2*) constitute the second class of plant resistance genes. All of them contain a nucleotide-binding site (NBS) and leucine-rich repeats (LRR) (see review by Ellis et al 2000). The *Cf9*, *Cf2*, and *Cf4* genes, which mediate resistance to the fungal pathogen *Cladosporium fulvum* in tomato (Jones et al 1994, Dixon et al 1996, Thomas et al 1997), form a third class of resistance genes. These *Cf* genes encode putative membrane-anchored proteins with the LRR motif in the extracellular domain and a short C-terminal tail in the intracellular domain. The fourth class of disease resistance genes is represented by a single member, *Xa21*, a rice gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* (Song et al 1995). The fifth class is represented by the *Xa21* gene family member, *Xa21D*, which encodes a presumed extracellular LRR lacking a transmembrane and kinase domains (see below).

Cloning and characterizing the bacterial blight resistance gene *Xa21*

Bacterial blight, caused by the bacterial pathogen *Xanthomonas oryzae* pv. *oryzae*, or *Xoo*, is a severe disease of rice that causes significant yield losses annually. Genetic studies on the inheritance of genes with resistance to bacterial blight began about three decades ago in Japan and at the International Rice Research Institute (Mew 1987). Major genes from different sources of resistant donors have been identified. As of now, 24 bacterial blight resistance genes have been found on different chromosomal loci (Kinoshita 1995, Lin et al 1998). Some of these have been used widely in rice breeding programs and some of them are targets of cloning efforts in several laboratories. Thus far, only two resistance genes, *Xa21* and *Xa1*, have been isolated using a map-based cloning strategy (Song et al 1995, Yoshimura et al 1998). *Xa1* encoded a 1,802-amino acid protein with high similarity to the deduced polypeptide domains of other resistance genes (e.g., *RPS2*, *RPM1*, *N*, and *L6*) and has two motifs of NBS and an LRR domain. In contrast to the constitutive expression seen in all cloned resistance genes, *Xa1* expression is induced upon inoculation with *Xoo* or by wounding.

Through wide hybridization, *Xa21* was transferred from the wild species *Oryza longistaminata* to the susceptible rice cultivar IR24, generating the near-isogenic line IRBB21 (Ikeda et al 1990, Khush et al 1990). In the test for disease resistance spec-

trum, IRBB21 conferred a high level of resistance to many diverse *Xoo* strains from India and the Philippines (Khush et al 1990). Using the restriction fragment length polymorphism (RFLP) mapping technique, the gene was found to be linked to the RFLP marker RG103 on chromosome 11 (Ronald et al 1992). To better understand the molecular basis of the *Xa21* resistance, Song et al (1995) cloned a gene at the *Xa21* locus using a map-based cloning strategy. *Xa21* is a member of a small multigene family with at least seven members. Most of these members are linked at a region of 270 kb. The deduced amino acid sequence of *Xa21* encodes a receptor kinase-like protein carrying LRRs in the putative extracellular domain, a single-pass transmembrane domain, and a serine/threonine kinase intracellular domain. Compared with other cloned plant resistance genes, the structure of *Xa21* is unique in carrying both the receptor domain LRR, presumably for recognition, and the kinase domain for subsequent signal transduction (Ronald 1997).

The LRR domain determines race-specific recognition and is subject to adaptive evolution

Genetic analysis of disease resistance has shown that resistance genes occur at many loci and genes at each locus are often highly polymorphic for resistance specificities. Recent work on the *L* and *M* loci of flax and *Cf* genes of tomato demonstrated that alterations in the LRR domain play a significant role in the evolution of resistance genes and production of new recognition specificities (Anderson et al 1997, Thomas et al 1997, Ellis et al 1999). Within the LRR domain, amino acid differences distinguishing *Cf4* and *Cf9* are confined to their N termini, delimiting a region that determines the recognition specificity of ligand binding. The majority of these differences are in residues interstitial to those of the LRR consensus motif (Thomas et al 1997). Recently, Ellis et al (1999) analyzed the sequences of 13 alleles (*L*, *L1* to *L11*, and *LH*) from the flax *L* locus, which encode Toll/interleukin-1 receptor homology–nucleotide-binding site–leucine-rich repeat (TIR-NBS-LRR) rust resistance proteins and found that specificity differences between alleles can be determined by both the LRR and TIR regions.

At least eight RG-103-hybridizing bands are present at the *Xa21* locus in IRBB21 when digested with restriction enzyme *Hin*dIII. Distinct amino acid differences are found in the LRRs of the *Xa21* gene family members, suggesting that the LRR region is responsible for different specificities. To test this hypothesis, Wang et al (1998) investigated the inheritance of resistance patterns in transgenic rice plants carrying family members A1, A2, C, D, E, and F to eight *Xoo* races of the Philippines. It was found that members A1, A2, C, E, and F conferred no observable resistance in transgenic plants whereas *Xa21* class member D, designated *Xa21D*, conferred the same resistance spectrum as *Xa21*. However, the resistance level in the *Xa21D* transgenic plants was intermediate compared with that observed for *Xa21* (5–7-cm lesion length). Sequence analysis indicated that the presumed open reading frame (ORF) of *Xa21D* was prematurely truncated because of the presence of a stop codon in the last LRR preceding the transmembrane domain. The truncated ORF encodes a

612-amino acid protein, including the signal peptide, and GC-rich and LRR domain. Up to *Retrofit* insertion, *Xa21D* is 99.1% identical to *Xa21* in nucleotide sequence. This result supported the hypothesis that the *Xa21D* LRR domain is involved in pathogen recognition.

To gain insight into the function and evolution of particular coding domains, analysis of the rates of synonymous (no amino acid alteration) and nonsynonymous (altered amino acids) nucleotide substitutions has been used to investigate the nucleotide substitution patterns in members of the *Xa21* gene family. The type of selection acting on a gene family can be inferred by comparing synonymous and nonsynonymous substitution rates per synonymous/nonsynonymous site (Hughes and Nei 1988). In the absence of selection pressure, the nonsynonymous/synonymous ratio is predicted to be 1. The nonsynonymous/synonymous ratio in the *Cf4* and *Cf9* LRR region was greater than 1, indicating positive selection for diversification of the gene family (Thomas et al 1997). Comparison of nucleotide substitutions in the LRR coding regions of *Xa21* and *Xa21D* revealed that, although *Xa21* and *Xa21D* share 99.1% sequence identity, nonsynonymous substitutions occur significantly more frequently than synonymous substitutions in the LRR. Thus, the results further support the hypothesis that LRRs play a role in ligand binding and that the LRR domain is subject to adaptive evolution (Wang et al 1998).

Evolution of the *Xa21* gene family

It has long been speculated that DNA alterations play a key role in the evolution of disease resistance loci. These changes will allow plants to generate new resistances to match the changing pattern of pathogen virulence. To investigate the evolution of the *Xa21* gene family, seven *Xa21* gene family members, designated A1, A2, B, C, D, E, and F, were cloned and sequenced (Song et al 1997). They are grouped into two classes based on sequence similarity. The *Xa21* class contains *Xa21*, as well as members D and F. The A2 class contains members A1, A2, C, and E. Although the nucleotide sequences within each class are nearly identical (98.0% average identity for the members of the *Xa21* class, 95.2% average identity for the members of the A2 class), sequence identity between members of the two classes is low (63.5% identity between *Xa21* and A2) (Song et al 1997). A highly conserved (HC) 233-bp sequence was identified among the seven family members such that recombination at the HC region between family members resulted in a precise swapping of the promoter regions. Large sequence duplications were generated by a presumed unequal crossover event in the intergenic regions.

Furthermore, sequence analysis of the seven members revealed that 17 transposon-like elements were identified in the 5′ and 3′ flanking regions and introns of these members (Song et al 1995). Recently, He et al (2000) found another transposon-like element (*Rim2*) in the *Xa21* locus. Its transcript accumulates in response to *Magnaporthe grisea* and its predicted protein shares similarity with TNP2-like proteins encoded by *CACTA* transposons. These elements are diverse, showing similarity to maize *Ds*, *CACTA*, and miniature inverted repeat-like elements, as well as novel

elements. Two of them were located in the ORFs of the members D and E, suggesting that these elements may play a role in the diversification of the *Xa21* family members. Based on these results, we propose that duplication, recombination, and transposition contribute to the amplification and diversification of the *Xa21* gene family. A proposed evolution model of the *Xa21* locus is presented in Figure 1.

Spectrum of resistance to 32 diverse bacterial blight isolates

The high level of resistance observed in IRBB21 to multiple *Xoo* isolates may be due to the single gene *Xa21* or a cluster of tightly linked genes, each of which recognizes a unique isolate-specific determinant. To test these hypotheses, 32 isolates from eight countries were used to inoculate *Xa21* transgenic plants. The transgenic plants expressing the cloned *Xa21* gene conferred resistance to 29 out of 32 diverse isolates, but were susceptible to three isolates from Korea. The resistance spectrum of the engineered line was identical to that of the donor line, indicating that the single cloned gene is sufficient to confer multi-isolate resistance (Wang et al 1996). It is worth noting that some of the *Xa21* transgenic lines showed increased resistance to *Xoo* compared with the donor lines. This increased resistance may be due to the high copy number of the transgene in these plants and this has great application for engineering high-level resistance to *Xoo* in rice cultivars.

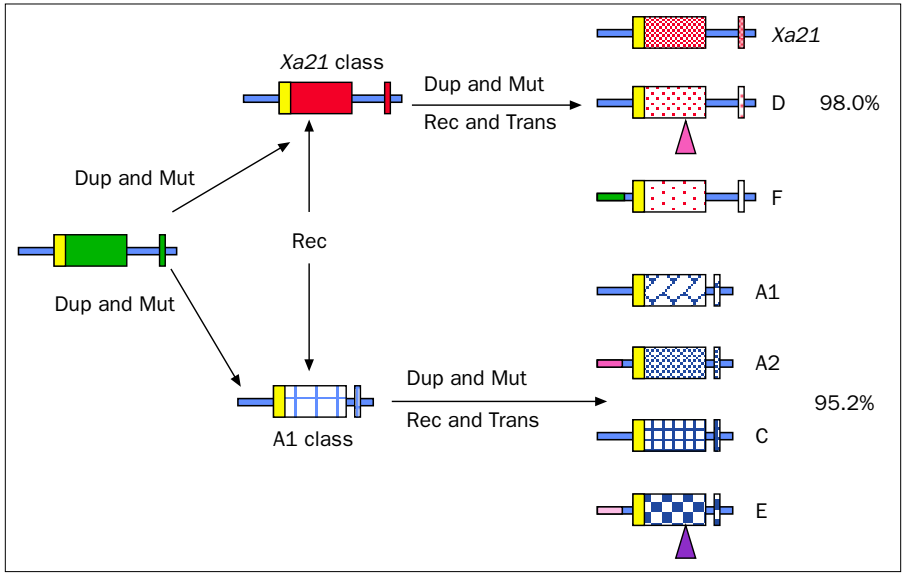


Fig. 1. Proposed model for the evolution of *Xa21* family members. Boxes represent the open reading frames of *Xa21* family members; horizontal bars represent the flanking regions and introns. The vertical bars indicate gene-coding regions. The same color indicates a high level of sequence homology. The highly conserved region is shown in yellow. Evolutionary events are indicated as follows: Dup = duplication, Mut = mutation, Rec = recombination, Trans = transposition. The percentage shows the sequence identity among the members in the same class.

Genetic dissection of the *Xa21*-mediated resistance pathway

To identify genes essential for the function of resistance against multiple diseases, a mutational approach was used. Both a chemical mutagen (diepoxybutane or DEB) and radiation (fast neutrons, FN) were used in the experiment. DEB has been reported to cause deletion mutations at high frequency in *Drosophila* (Reardon et al 1987) and has been used to generate mutants in several plant species (Salmeron et al 1994). DEB primarily causes deletions of less than 250 base pairs although deletions of up to 8 kb have also been observed (Reardon et al 1987). Fast neutrons induce both base pair changes and deletions and translocation. Salmeron et al (1994, 1996) isolated several *Pto* mutants using DEB and FN as mutagens and cloned the *Prf* gene, which is a common component of the transduction of signals for *Pto*-mediated resistance to *Pseudomonas syringae* pv. *tomato* as well as for sensitivity to the insecticide fenthion. A population of about 4,500 M2 families from IRBB21 (*Xa21* donor) with DEB (3,500 families) and FN (1,000 families) was used to screen for changes in disease resistance response to both bacterial blight and rice blast. We recovered 31 mutants that have changed from resistant to fully susceptible (10) or partially susceptible (21) to nine races of the bacterial blight pathogen in the Philippines (Fig. 2).

Polymerase chain reaction (PCR) analyses were used to detect deletions and rearrangements in the *Xa21* coding region. Two pairs of specific primers in the LRR region and two pairs of specific primers in the kinase region of the *Xa21* gene were developed. We found that some mutants such as 3453, 3469, 3954, and N20-247 lost both the LRR and kinase DNA fragments. Some lines such as N18-116-1 and N18-238-1 have rearrangements in the kinase region (Fig. 3).

For Southern analysis, DNA of selected mutants was digested with the restriction enzyme *Hin*dIII. Amplified fragments from the *Xa21* LRR and kinase regions were used as probes. We found that six FN-induced mutants had hybridization patterns similar to those of three missing LRR-hybridizing bands (data not shown). Three DEB-induced lines had deletions in the *Xa21* gene when the LRR fragment was used as a probe. Line 3453 had lost three bands and the size of the remaining bands was different from that of the *Xa21* donor line IRBB21. Lines 3469 and 3954 lost seven bands and had the same hybridization pattern as the recurrent parent IR24. When the kinase fragment was used in Southern hybridization, all 10 mutants with deletions in

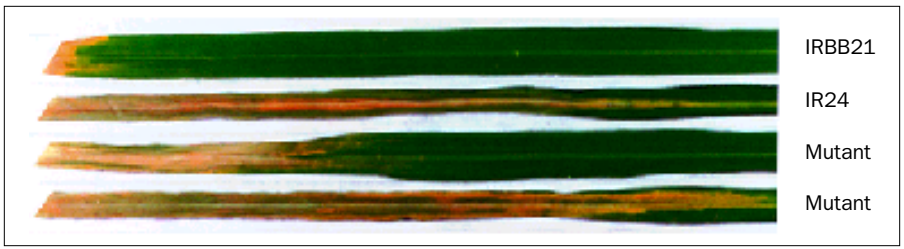


Fig. 2. Bacterial blight reaction of identified mutants. Two-month-old plants were used for inoculation with Philippine isolate POX99A. Disease reaction was recorded 2 wk after inoculation.

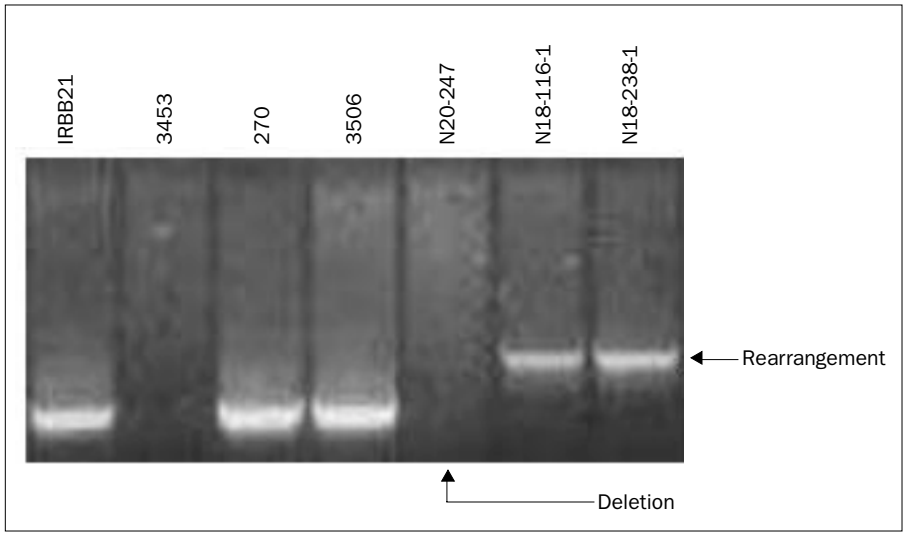


Fig. 3. Polymerase chain reaction analysis of putative mutants using the *Xa21* leucine-rich repeat-specific primers. DEB mutants are 3453, 270, and 3506; N20-247, N18-116-1, and N18-238-1 are fast-neutron mutants.

the LRR region showed only one hybridizing band. Interestingly, all these 10 lines that had deletions in the *Xa21* gene cluster were highly susceptible to *Xoo*. Mutants with partial resistance to race 6 had no apparent deletion in the *Xa21* locus, which was detected based upon both PCR and Southern analyses, suggesting that these mutations occur at other loci controlling the *Xa21*-mediated defense pathway.

To investigate whether the resistance pathway mediated by the *Xa21* gene is shared with the pathways required by the genes with resistance to blast, all mutants were inoculated with four Philippine blast isolates belonging to four different genetic lineages. Most of the lines showed no difference in their resistance specificity to all the isolates except lines 3453, 3469, and N20-247, which were susceptible to two out of seven blast isolates that are incompatible to IRBB21. Interestingly, all three lines have deletions in the *Xa21* locus.

Crosses were made between these three mutant lines and the wild-type IRBB21. The F_2 families were used to determine whether the *Xa21* gene confers a dual-resistance function (to bacterial blight and blast). After inoculation with blast isolate PO6-6, DNA was extracted from at least 50 individual plants for genotype analysis. Both PCR and Southern blot methods are used in the analysis. In all three F_2 families, no correlation was observed between the absence of the *Xa21* gene and blast susceptibility of individual plants. However, phenotypic and molecular analyses of the F_2 families show that susceptibility to two bacterial blight strains (PXO79 and PXO99) cosegregates perfectly with the *Xa21* deletion. These results indicate that bacterial blight and blast susceptibility in these three mutants are completely independent.

References

Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan J, Ellis JG. 1997. Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell 9:641-651.

Baker B, Zambryski P, Staskwicz B, Dinesh-Kumar SP. 1997. Signaling in plant-microbe interactions. Science 276:726-732.

Dixon M, Jones D, Keddie J, Thomas C, Harrison K, Jones J. 1996. The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84:451-459.

Ellis J, Dodds P, Pryor T. 2000. Structure, function and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. 3:278-284.

Ellis JG, Lawrence GJ, Luck JE, Dodds PN. 1999. Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. Plant Cell 11:495-506.

He ZH, Dong HT, Dong JX, Li DB, Ronald PC. 2000. The rice *Rim2* transcript accumulates in response to *Magnaporthe grisea* and its predicted protein product shares similarity with TNP2-like proteins encoded by *CACTA* transposons. Mol. Gen. Genet. 264(1-2):2-10.

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170.

Ikeda R, Khush GS, Tabien RE, Ogawa T. 1990. A new resistance gene to bacterial blight from *Oryza longistaminata.* Jpn. J. Breed. 40(Suppl. 1):280-281.

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JDG. 1994. Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266:789-793.

Khush GS, Bacalangco E, Ogawa T. 1990. A new gene for resistance to bacterial blight from *O. longistaminata*. Rice Genet. Newsl. 7:121-122.

Kinoshita T. 1995. Report of committee on gene symbolization, nomenclature and linkage groups. Rice Genet. Newsl. 12:9-115.

Lin X, Wang C, Wen G, Zhang D, Xie Y, Zhang Q. 1998. Progress in map-based cloning of *Xa22(t)*, a new gene for bacterial blight resistance in rice. Paper presented at the Plant and Animal Genome Conference, 18-22 Jan. 1998, San Diego, Calif., USA.

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432-1436.

Mew TW. 1987. Current status and future prospects of research on bacterial blight of rice. Annu. Rev. Phytopathol. 25:359-382.

Reardon JT, Liljiestrand-Golden CA, Dusenberg RL, Smith PD. 1987. Molecular analysis of diepoxybutane induced mutations at the rosy locus of *Drosophila melanogaster*. Genetics 115:323-331.

Ronald PC, Albano B, Tabien R, Abenes L, Wu K, McCouch SR, Tanksley SD. 1992. Genetic and physical analysis of the rice bacterial blight disease resistance locus, *Xa21*. Mol. Gen. Genet. 236:113-120.

Ronald PC. 1997. The molecular basis of disease resistance in rice. Plant Mol. Biol. 35:179-186.

Salmeron JM, Barker SJ, Carland FM, Mehta AY, Staskawicz BJ. 1994. Tomato mutants altered in bacterial disease resistance provide evidence for a new locus controlling pathogen recognition. Plant Cell 6:511-520.

Salmeron JM, Oldroyd GED, Rommens CMT, Scofield SR, Kim HS, Lavelle, DT, Dahlbeck D, Staskawicz BJ. 1996. Tomato *Prf* is a member of the leucine-rich repeat class of disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86:123-133.

Scofield SR, Tobias CM, Rathjen JP, Chang JH, Lavelle DT, Michelmore RW, Staskawicz BJ. 1996. Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. Science 274:2063-2065.

Song W-Y, Wang G-L, Chen L, Kim H-S, Pi Li-Ya, Gardner J, Wang B, Holsten T, Zhai W-X, Zhu L-H, Fauquet C, Ronald P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. Science 270:1804-1806.

Song W-Y, Pi L-Y, Wang G-L, Gardner J, Holsten T, Ronald PC. 1997. Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9:1279-1287.

Staskawicz BJ, Ausubel FM, Baker B, Ellis JG, Jones JDG. 1995. Molecular genetics of plant disease resistance. Science 268:661-667.

Tang X, Frederick RD, Zhou J, Halterman DA, Jia Y, Martin GB. 1996. Initiation of plant disease resistance by physical interaction of *avrPto* and *Pto* kinase. Science 274:2060-2063.

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurt PJ, Hatzixanthis K, Jones JDG. 1997. Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell 9:2209-2224.

Wang GL, Song WY, Ruan DL, Sideris S, Ronald PC. 1996. The cloned gene, *Xa21*, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic plants. Mol. Plant-Microbe Interact. 9:850-855.

Wang GL, Ruan DL, Song WY, Sideris S, Chen L, Pi L, Whalen MC, Zhang S, Fauquet C, Ronald PC. 1998. The rice disease resistance gene, *Xa21D*, encodes a receptor-like molecule with a LRR domain that determines race specific recognition and is subject to adaptive evolution. Plant Cell 10:765-780.

Yoshimura S, Yamanouchi U, Katayose Y, Tokii S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T. 1998. Expression of *Xa1*, a novel bacterial blight resistance gene in rice, is induced by bacterial inoculation. Proc. Natl. Acad. Sci. USA 95:1663-1668.

Notes

Authors' addresses: Guo-Liang Wang, Department of Plant Pathology, Ohio State University, Columbus, Ohio 43210, USA; Hei Leung, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines; P. Ronald, Department of Plant Pathology, University of California-Davis, Davis, Calif. 95616, USA.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Isolation of candidate genes for tolerance of abiotic stresses

H.J. Bohnert, S. Kawasaki, C.B. Michalowski, Hong Wang,
N. Ozturk, M. Deyholos, and D. Galbraith

High-throughput analysis of expressed genes, achieved by cataloguing expressed sequence tags (ESTs) and monitoring hybridization patterns by microarrays, has recently become possible in rice. As the first results become available, the value of these technologies can be gauged. Through ESTs and microarrays, we can obtain a more complete view than in the past of plant gene complexity, tissue specificity, and developmental or externally affected expression patterns. In particular, EST and microarray analyses can have tremendous impact in plant breeding, based on accelerated identification of complex traits such as those controlling plant responses to abiotic stresses. Owing to the novelty and lack of refinement in the use of microarray technology, we discuss advantages and limitations. We demonstrate responses to salt stress in rice (*Oryza sativa*) monitored by microarray analysis.

The results of plant breeding over millennia provide all food, feed, and fiber for human use. Following the domestication of a few hundred species, about ten crops provide the majority of all food. Selection, increasingly based on scientific criteria, has allowed us to keep pace with population growth (Khush 1999, Evans 1998). Population growth, changes in lifestyle, competition for fresh water between farmers and cities, and possible global environmental changes have led to projections that advocate additional strategies by which food supply can be either established or guaranteed over the long term (Miflin 2000). Areas where food production is problematic coincide largely with places where water is a precious commodity and prolonged droughts are frequent, saline and marginal soils are found, and temperature extremes compromise production. Moreover, the increased productivity achieved by irrigation constitutes a double-edged sword, because salinization following prolonged irrigation is unavoidable (Flowers and Yeo 1995, Postel 1999). These considerations have galvanized a strong interest in studying plant abiotic stress responses and understanding the meaning of stress tolerance as a biological phenomenon.

Understanding abiotic stress tolerance proved difficult because of the trait's multigenicity. As a consequence, no traditional crop lines exist that combine tolerance for high salinity or drought with the equally complex high-yield trait. Many studies have addressed the problem and tolerance mechanisms have been defined (Levitt 1980). The last decade has seen the move from physiology and the genetics of breeding to the study of gene isolation and function and the beginning of transgenic and mutant generation and analyses (Jain and Selvaraj 1997).

Recently, progress in molecular genetics has come from new technologies, which, focusing on plant breeding and productivity, encompass the use of molecular markers, large-scale mutagenesis, genome sequences, and functional analysis of gene and protein expression characters. As of 2000, the sequence of the *Arabidopsis thaliana* genome is publicly available, and the catalog of plant genes exceeds a million transcripts (http://www.ncbi.nlm.nih.gov/dbEST/index.html) (Lin et al 1999, Mayer et al 1999, Somerville and Somerville 1999). The molecular analysis of these transcripts, collected as ESTs (expressed sequence tags), is providing the frequency and type of transcribed genes for organs, tissues, and cells during development and under various experimental conditions. Such expression profiles will increasingly be augmented by the analysis of EST-derived microarrays or oligonucleotide-based DNA chips (Schena et al 1995, Wang 2000, Kawasaki et al 2001). Together, the two strategies provide information about changes in expression levels over time or during experimental perturbations of the plants (Richmond and Somerville 2000, van Hal et al 2000). Accompanying these activities has been yet another breakthrough, whose impact on crop breeding will be even more significant. In *Arabidopsis thaliana*, rice, and possibly maize, saturation mutagenesis has been achieved (Jeon et al 2000, McCallum et al 2000, Barakat et al 2000, Parinov and Sundaresan 2000). These mutants allow for proof of concept experiments on a large scale, for example by characterizing mutants or transgene overexpressing lines for stress-relevant signal transduction and response pathways.

Here we discuss an initial set of data from an ongoing project that is focusing on defining the gene sets associated with tolerance for salinity stress and drought (www.stress-genomics.org). In the grass crops rice, maize, and barley, and the dicot species *Arabidopsis thaliana* and *Mesembryanthemum crystallinum*, we have begun to compare salinity, drought, and oxidative stress responses. In rice, we analyze lines that are distinguished, to some degree, by their tolerance of salt stress using the indica varieties Pokkali, IR29, and IR64, which are distinguished by their degree of salinity tolerance (Kawasaki et al 2001). From salt- and drought-stressed plants, root, shoot, and leaf cDNA libraries were generated and several thousand ESTs have been sequenced. A selection of those cDNAs, arrayed on glass slides, is employed in hybridizations using transcripts taken from unstressed and stressed plants. Here, we will focus on the diversity of expression profiles in different tissues and on changes in these expression profiles that distinguish stressed and unstressed tissues. From microarray analyses, some ideas about candidate genes of salinity stress tolerance seem to have emerged. We discuss what we consider to be the drawbacks of microarray analyses and try to outline requirements for this technique to provide adequate information.

Materials and methods

Plant materials and growth conditions

Seeds of two rice varieties (Pokkali and IR29) were obtained from the International Rice Research Institute, Los Baños, Philippines. Maize (*Zea mays* L., B73) was provided by Dr. Vicky Chandler (University of Arizona) and barley (*Hordeum vulgare* L., Tokak) by Dr. Nermin Gozukirmizi (Marmara Research Center, Turkey). *Mesembryanthemum crystallinum* L. was grown as described (Adams et al 1998). Maize plants were grown on a potting mixture for 4 wk and then treated with 150 mM NaCl using water as a control.

After imbibition, rice seeds were transplanted into Hoagland's solution, with the amount of iron doubled, in hydroponic tanks. Plants were grown at 28/25 °C (50% humidity, 12 h light/dark at 700 μmol photons m^{-2} s^{-1}). Plants were used when the roots and shoot measured approximately 7 and 10 cm, respectively. Salt stress began 3 h after the start of the light period by adding 150 mM NaCl in diluted (1/4) Hoagland's solution, which provided external calcium at 1 mM. Roots and leaves were harvested, frozen in liquid nitrogen, and kept at –80 °C.

Barley seeds were germinated in sand supplied with 0.3X Hoagland's nutrient solution. The plants were grown at 28/23 °C (80–85% relative humidity, 12 h dark/light at 500 μmol photons m^{-2} s^{-1}). At age 3 wk, plants were removed from the sand and placed on a bench for drying at 80–85% relative humidity and light intensity as above. Leaf and root samples were collected after 6 and 10 h of drying. Control plants grown in sand were harvested at the same times.

Physiological analysis

Net CO_2 assimilation, stomatal conductance, and transpiration rates were measured with attached leaves at saturating light intensity (1,000 μmol photons m^{-2} s^{-1}) at 28 °C using an infrared gas analyzing system (Li-6400, Li-cor). Data were collected twice for each time-course experiment. Physiological parameters were measured before and after salt additions.

RNA isolation and cDNA library construction

cDNA libraries were constructed from the plant species and tissues listed in Table 1. Total root RNA and poly(A^+) RNA were isolated and cDNA libraries were generated (Stratagene) with *Escherichia coli* XL1-Blue MRF as the host. Inserts cloned into Bluescript SK+ were sequenced from the 5′ ends. Sequences were annotated accepting rice ESTs included in public databases for transcripts that showed more than 95% identity to the Pokkali ESTs.

Preparation of DNA microarrays

PCR amplification (40 cycles, annealing at 55 °C) was performed in a 96-well format with individual colonies or 1 μL of plasmid DNA as templates, using T3 and T7 primers with amino-modified ends. PCR products were combined with 100 μL of binding solution (150 mM potassium acetate, pH 4.8, 7 M guanidine hydrochloride)

Table 1. Selected cDNA libraries from salt-stressed plants.

Name	Organism	Tissue	Stress condition[a]	Plant age
HB	*H. vulgare* (Tokak)	Leaves	Drought; 6 and 10 h	~3 wk
HC	*H. vulgare* (Tokak)	Roots	Drought; 6 and 10 h	~3 wk
MC	*M. crystallinum*	Roots	No stress	5–6 wk
ME	*M. crystallinum*	Roots	400; 6 h	5–6 wk
MF	*M. crystallinum*	Roots	400; 12 h	5–6 wk
MG	*M. crystallinum*	Roots	400; 30 h	5–6 wk
MH	*M. crystallinum*	Roots	400; 78 h	5–6 wk
MI	*M. crystallinum*	Seedlings	250; 3 d	14 d
ML	*M. crystallinum*	Flowers and seedpods	500; 6 wk	>12 wk
MM	*M. crystallinum*	Epidermis	500; 6 wk	>12 wk
MN	*M. crystallinum*	Side shoots	500; 3 d	6 wk
MO	*M. crystallinum*	Meristems	No stress	5 wk
MP	*M. crystallinum*	Meristems	500; 3 d	6 wk
OA	*O. sativa* (Nipponbare)	Roots	200; 19 h	3–4 wk
OB	*O. sativa* (Nipponbare)	Leaves	200; 19 h	3–4 wk
OC	*O. sativa* (Pokkali)	Roots	No stress	1 wk
OD	*O. sativa* (Pokkali)	Roots	150; 1 d	1 wk
OE	*O. sativa* (Pokkali)	Roots	150; 2, 3 d	1 wk
OF	*O. sativa* (Pokkali)	Roots	150; 1 wk	2 wk
OG	*O. sativa* (Pokkali)	Leaves	150; 1, 2, 3 d, 1 wk	1 wk
OH	*O. sativa* (Pokkali)	Leaves	No stress	1–2 wk
ZA	*Z. mays* (B73)	Roots	150; 24 h	2 wk
ZB	*Z. mays* (B73)	Leaves and shoots	150; 24 h	2 wk

[a]Stress is expressed as mM NaCl and length of the treatment. NaCl is applied as a shock treatment.

and filtered. PCR products were eluted with 10 mM Tris-EDTA. About 1.2–2.4 μg of PCR products were dried and the pellets dissolved in 6 μL of 1X SSC for printing. Microarrays were produced by using the Omnigrid spotter (GeneMachines, San Carlos, Calif.). Slides, coated with either polylysine or aminoalkylsilane, contained cDNAs spotted in duplicate or triplicate. Only amplicons longer than 400 bp were printed.

Preparation of labeled probes

Fluorescence-labeled probes were prepared from RNAs by incorporation of fluorescent nucleotide analogs during first-strand reverse transcription. Each reaction (50 μL) consisted of 1 μg mRNA, 200 ng of in vitro transcripts as human control mixture, 2 μg of oligo(T) primers, 0.5 mM each of dATP, dCTP, and dGTP, 0.2 mM dTTP, and 0.5 unit of reverse transcriptase (Superscript II, GIBCO Inc.) in 1x reaction buffer and 2 nmol of either Cy3-dUTP or Cy5-dUTP (Amersham, Pharmacia). RNA and primers were heated to 65 °C (10 min) and quenched on ice before adding the remaining reaction components. The RT reaction proceeded for 10 min at 42 °C preincubation followed by 90 min at 42 °C. Buffer exchange, purification, and concentration of cDNA products were by microfiltration (Qiagen). Labeled targets were collected by centrifugation after the addition of 0.1 volume of 3M potassium acetate and 1 volume of isopropanol. The dried pellets were reconstituted in 20 μL of 5X SSC/0.1% SDS/50% formamide and denatured (95 °C) prior to use in hybridization.

Microarray hybridization and data analysis

Hybridizations were performed overnight at 42 °C in humidified chambers. The slides were then washed in 1X SSC/0.2% SDS (10 min) and in 0.1X SSC (10 min). Slides were rinsed for 1 min in 0.01X SSC and dried by centrifugation. The fluorescent signatures were captured using a ScanArray 3000 (GSI Luminomics) and analyzed using ImaGene III software (BioDiscovery). Local background was subtracted from the value of each spot on the array. Spots covered by dust particles, missing spots, spots with low signal intensity, and spots in high background areas were flagged as candidates for exclusion after further analysis. Normalization between the Cy3 and Cy5 emission channels was achieved by adjusting the signal intensity of exogenously added nonplant control genes. Transcript regulation is expressed as the ratio of intensities between the stress and control (log ratio, termed LR).

Results and discussion

The data selected and outlined here are designed to make several points, mainly on the variation of transcript profiles in different tissues, the changes that are observed under stress conditions, and the preliminary microarray analyses destined to provide information about likely candidate genes for stress tolerance characteristics.

The generation of cDNA libraries in our projects focused almost exclusively on tissues from salinity- and drought-stressed plants, but we included several cDNA libraries for the unstressed state to obtain EST sequences for comparing expression profiles (Table 1). These cDNA libraries were used for EST sequencing. Typically, we collect approximately 1,000–2,000 EST sequences without subtraction. These ESTs are then subtracted from the next set of approximately 2,000 clones and this subtraction process can be repeated. However, we have rarely found time to go "deeper" into these libraries, that is, truly rare messages are presently not represented in our expression profiles. This statement is based on the classical DNA/RNA renaturation (cot-curves) experiments (Goldberg et al 1978, Kiper et al 1979, Buffard et al 1982, Kamalay and Goldberg 1984). These analyses had indicated that the complexity in, for example, roots could include 10,000 different expressed genes. At present, our way of proceeding produced from 2,000 to 4,000 unique transcripts for three grass species, rice, maize, and barley.

From the initial nonsubtracted 1,000–2,000 EST sequences, expression profiles can be generated that represent the set of expressed genes in functional categories (Table 2) and also provide an overview of highly transcribed transcripts (Table 3). The category profile for salt-stressed maize roots indicates a large number of unknown transcripts (42%) and similar percentages have been found for rice (Kawasaki et al 2001). Among the most highly transcribed genes in salt-stressed maize roots are isoforms for a metallothionein-like protein, a glutathione S-transferase, glutamine synthetase, several different (putative) water channel proteins, and several open reading frames for which no other sequence has been found before ("no hit" category).

Table 2. Abundance profile of expressed sequence tags (ESTs) in a root cDNA library of salt-stressed maize.

Gene category	No. of ESTs	% of total	Category description
No hit[a]	493	26.2	Blast search in nonredundant databases finds no homologous sequences
Unknown[b]	296	15.7	Blast search in databases identifies unknown, hypothetical, and putative protein, or proteins with unknown functions
Metabolism	203	10.8	Amino acid, nucleotide, C-compound, carbohydrate, lipid, fatty-acid, isoprenoid, nitrogen, sulfur, and secondary metabolisms
Cell rescue and defense	165	8.8	Environmental stimuli responses (stress-, wounding-, phytohormone-regulated, ion deficiency-related), radical oxygen scavenging, detoxification, DNA repair, and cell death
Protein synthesis	123	6.5	Ribosomal proteins, translation (initiation, elongation, and termination), translational control, and tRNA synthetases
Energy	111	5.9	Glycolysis and gluconeogenesis, pentose-phosphate pathway, TCA pathway, respiration, metabolism of energy reserves (e.g., glycogen, trehalose), and fatty-acid oxidation
Transport facilitation	76	4.0	Ion channels, water channels, ion, sugar, carbohydrate, amino acid, lipid, purine, pyrimidine transporters, transport ATPases
Protein destination	76	4.0	Protein folding and stabilization, protein targeting, sorting, and translocation, protein modification (e.g., myristylation, farnesylation, palmitoylation, glycosylation), and proteolysis
Signal transduction	75	4.0	(Putative) receptor proteins, second messenger such as calmodulins, key kinases, key phosphatases, phospholipases, and G-proteins
Cell growth, division	74	3.9	Cell growth and development, cell cycle control and mitosis, cytokinesis, and DNA synthesis and replication
Transcription	71	3.8	mRNA synthesis (including general transcription activities, transcriptional control such as transcriptional factors, and chromatin modification), mRNA splicing, and mRNA stabilization and degradation
Photosynthesis	63	3.3	Chlorophyll a/b-binding protein, ferredoxin-related genes, photosystem I & II reaction center proteins
Metal ion homeostasis	39	2.1	Homeostasis of metal ions and other ions
Intracellular transport	17	0.9	Vesicular transport (Golgi network), vacuolar transport, cytoskeleton-dependent transport
Total	1,882	100	

[a]No significant homology with ESTs, cDNAs, or genes deposited in GenBank. [b]Significant homology with entries in GenBank for which no clear function is available.

Table 3. Most abundant transcripts in cDNA library from salt-stressed maize root tissue.

	Putative identification	Functional category	Gene accession no.[a]	Score	Organism	Abundance (%)
1	Metallothionein-like protein 1	Metal ion homeostasis	P30571	198	*Zea mays*	1.63
2	Glutathione S-transferase I	Cell rescue and defense	P12653	291	*Z. mays*	0.64
3	Water channel (tonoplast)	Transport facilitation	AAC09245	245	*Z. mays*	0.54
4	Glutamine synthetase	Metabolism	D14579	270	*Z. mays*	0.50
5	RCc3 protein	Unknown	AAA65512	147	*Oryza sativa*	0.45
6	PS II type II chlorophyll a/b-binding protein	Photosynthesis	CAA48641	227	*Z. mays*	0.40
7	Caffeoyl CoA O-methyltransferase	Metabolism	CAB45149	219	*Z. mays*	0.35
8	Glycine-rich RNA binding protein	Transcription	AF034945	112	*Z. mays*	0.35
9	Water channel protein	Transport facilitation	S60455	147	*Z. mays*	0.35
10	Water channel protein, MipH	Transport facilitation	U87981	157	*Sorghum bicolor*	0.35
11	Submergence-induced protein 2A	Unknown	AAC19375	265	*O. sativa*	0.30
12	Zn-induced protein	Unknown	BAA20532	134	*Daucus carota*	0.30
13	Nicotianamine synthase 1	Metabolism	AB021746	204	*O. sativa*	0.30
14	Novel	No hit	–	–	–	0.30
15	PS I reaction center subunit X (PSI-K)	Photosynthesis	P36886	179	*Hordeum vulgare*	0.30
16	Ribosomal protein l35A	Protein synthesis	P49180	124	*Caenorhabditis elegans*	0.30
17	Protein translation factor	Protein synthesis	P56330	156	*Z. mays*	0.30
18	Histone H3.2	Cell growth and division	AAD23951	199	*Tortula ruralis*	0.25
19	Ascorbate peroxidase	Cell rescue and defense	CAA06996	250	*H. vulgare*	0.25
20	ABA inducible-like protein	Unknown	AAA21866	156	*Z. mays*	0.25
21	Glutathione S-transferase IV	Cell rescue and defense	P46420	239	*Z. mays*	0.25
22	Symbiosis-related protein	Unknown	AAD24645	98	*Arabidopsis thaliana*	0.25
23	Calmodulin	Signal transduction	P01235	134	*Carausius carpio*	0.25
24	Glyceraldehyde-3-P dehydrogenase	Energy	PQ0178	186	*Z. mays*	0.25
25	N-OH-cinnamoyl/benzoyltransferase	Metabolism	BAA87043	142	*Ipomoea batatas*	0.25
26	Carbonic anhydrase	Metabolism	U08401	143	*Z. mays*	0.25
27	Methionine synthase	Metabolism	S57636	172	*Vinca* sp.	0.25
28	S-adenosylmethionine synthetase	Metabolism	P50299	164	*H. vulgare*	0.25
29	Novel	No hit	–	–	–	0.25
30	PS I type III chlorophyll a/b-binding protein	Photosynthesis	AAD25555	196	*A. thaliana*	0.25

continued on next page

Table 3 continued.

Putative identification	Functional category	Gene accession no.[a]	Score	Organism	Abundance (%)
31 Ferredoxin-nitrite reductase	Metabolism	JA0172	176	*Z. mays*	0.25
32 Photosystem II	Photosynthesis	U86018	177	*O. sativa*	0.25
33 Ribosomal protein l27	Protein synthesis	P51419	145	*A. thaliana*	0.25
34 GF14-c protein	Transcription	P49106	141	*Z. mays*	0.25
35 Water channel protein	Transport facilitation	AAA69490	184	*Glycine max*	0.25
36 Unknown	Unknown	AF051205	152	*Picea mariana*	0.25
37 Beta-5 tubulin	Cell growth and division	CAA52719	148	*Z. mays*	0.20
38 Histone H2B	Cell growth and division	P54348	165	*Z. mays*	0.20
39 Histone H3.2	Cell growth and division	AAB36496	206	*Medicago sativa*	0.20
40 Peroxidase 1	Cell rescue and defense	AAD43561	95	*Gossypium hirsutum*	0.20
41 Ras-related protein rab7	Signal transduction	Q40787	292	*Pennisetum ciliare*	0.20
42 Enolase-phosphatase E-1	Energy	AAF14866	129	*Homo sapiens*	0.20
43 Cytochrome C reductase complex 14 kDa protein	Energy	P48502	164	*Solanum tuberosum*	0.20
44 Fumarylacetoacetate hydrolase	Metabolism	AAC17611	118	*A. thaliana*	0.20
45 Lipase	Metabolism	BAA83367	102	*O. sativa*	0.20
46 Nucleoside diphosphate kinase I	Metabolism	Q07661	122	*O. sativa*	0.20
47 Cinnamyl alcohol dehydrogenase	Metabolism	CAA06687	169	*Z. mays*	0.20
48 Novel	No hit	–	–	–	0.20
49 Novel	No hit	–	–	–	0.20
50 Novel	No hit	–	–	–	0.20

[a]Accession number refers to the most highly conserved sequence found by alignment of protein sequences. Sample size: 1,882 clones. Stress conditions: 150 mM NaCl, 24 h.

Microarray technology

The rapid accumulation of genomic sequences has been paralleled by the growth in collections of ESTs, each of which contains the partial sequence of an expressed gene (Adams et al 1993, Marra et al 1998). The public dbEST includes approximately 900,000 plant EST sequences among the more than 6.6 million ESTs deposited in total. For these sequences to be informative, the genes that they identify must be assigned biological functions. This will ultimately be a biochemical or structural function, but many sequences are now annotated in descriptive terms that must not be mistaken for a functional identity. Realistically, we know the precise function of maybe a quarter of all genes and we know some characteristics for about another quarter. At least one-half of all genes, however, remain functionally unknown. Expression profiling using DNA microarrays provides a powerful tool for correlating some aspects of gene functions, such as tissue-, developmental-, or external manipulation-specific expression, with sequences. DNA microarrays are also useful for identifying differentially expressed genes in organisms for which genomic DNA sequences are limited.

The use of microarrays for expression profiling is based on two fundamental principles. The first is that, for many genes, a predominant factor underlying changes in expression is an alteration in the abundance of the cognate mRNA. It is clear that post-transcriptional factors also affect gene expression but the analysis of these factors is not amenable through microarray analysis. The second principle is that only DNA strands possessing complementary sequences can hybridize to each other to form a stable, double-stranded molecule. Microarrays exploit this property through the immobilization of single-strand copies of a gene as individual array elements on surfaces such as glass slides coated with polylysine. These elements are termed the "probe" (Phimister 1999). After incubation with a mixture of (radioactively or fluorescently) labeled DNA molecules (termed "target"), representing a proportional representation of all genes present in a given (tissue) sample, the labeled molecules that represent the same gene as the immobilized DNA elements can form heteroduplexes. By measuring the amount of label that is bound to each array element at the end of the hybridization reaction, the relative transcript abundance of each gene can be determined. In this way, RNA abundance levels for thousands of genes can be measured in a single experiment. Patterns of gene expression can be correlated with particular tissues or experimental conditions by comparing abundance levels from different experiments.

Two types of microarrays, which differ primarily in the length of the probes that make up the array elements and in the production of these elements, are being used for expression profiling. For DNA microarrays, printed on membranes or glass, the elements are large gene fragments (approximately 400 to 2,000 bp or more), typically produced through polymerase chain reaction (PCR) amplification. In contrast, oligonucleotide-based arrays (termed "chips") (Lipshutz et al 1999) contain elements made up of short synthetic DNA molecules. These are synthesized such that a set of 10 or more oligonucleotide fragments report the presence of one gene and they include controls, which have one base mismatch for each of the selected fragments. The pro-

duction of chips is done commercially, requiring specialized equipment and procedures and chip array-specific software. Chips, at least at present, are not within the means of most academic research laboratories, but the generation and use of DNA microarrays are.

The production of DNA microarrays includes the selection and preparation of the probes and their robotic deposition on suitable surfaces (see Deyholos and Galbraith 2001, for an in-depth discussion of the technology). Pursuing this technology requires a long-term commitment in resources and training. Table 4 presents an over-

Table 4. Microarray use and problems.

Equipment and procedures[a]	Remarks
The following are minimum equipment necessities, associated with considerable cost PCR machines Spotting robot Microarray reader (scanner)	General equipment category, which is essential. Best used when funds and equipment are shared through a center. Permanent supervision required.
Amplification of ESTs or cDNAs	Commercial kits; robots optional; PCR products may be amplified from bacterial colonies or from plasmids; production and quality control are time-consuming. Assembly of plates (96/384) is a source for error. Bar-coding and random checks by resequencing are recommended.
DNA purification and quantification	Commercial kits; robots optional but required for large-scale work. Rearrangement of PCR products may be necessary to eliminate inefficiently amplified products or those producing multiple bands.
Spotting	Training required; various treatments of slide surfaces for DNA attachment; controlled environmental conditions are essential during spotting/printing.
Control genes	Inclusion of positive and negative control genes. Negative controls are best taken from widely divergent species (e.g., mammalian genes).
Quality control	Slides must be visually checked for irregularities.
Hybridization	Background control; hybridization chambers must be tested; ion precipitates must be avoided.
Analysis of fluorescence images	Aberrant spots (off-center, dust particles, poor DNA retention) must be flagged and eliminated; time-consuming step.
Statistical analysis and clustering	Require repeat hybridizations and multiple time points.

[a]PCR = polymerase chain reaction, EST = expressed sequence tag.

view of procedures, pitfalls, and necessary controls in microarray generation, hybridization, and analysis.

Tissue-specific expression profiles

One useful application of microarray analysis is the determination of tissue- and cell-specificity. Figure 1 shows the intensity distribution for 2,400 ESTs (open circles) from *Mesembryanthemum* in stems of this plant and in epidermal bladder cells (EBC), which are modified trichomes that accumulate NaCl in this halophyte. Plotted are intensity data. Added to this array were 96 probes derived from a cDNA library generated from isolated EBC (black squares). It is obvious that most EBC-derived ESTs are more highly represented in the target RNA derived from EBC. An inspection of ESTs that are more highly represented in EBC RNA indicated that these sequences encode functions that can be considered epidermis-specific, for example, pathogenesis-induced (antifungal, antibacterial, wound-inducible) or lipid transfer proteins.

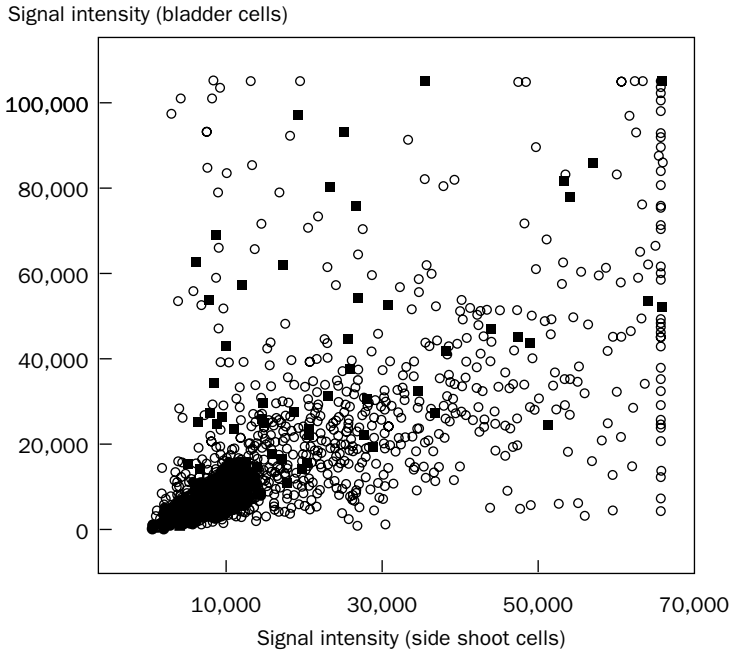


Fig. 1. Tissue- and cell-specific transcript abundance detected by microarray analysis. A microarray including in excess of 2,000 ESTs from various cDNA libraries (○) from the ice plant (*Mesembryanthemum crystallinum*) is shown. Additionally included were 96 ESTs derived from a cDNA library made from isolated epidermal bladder cells (EBC) (■). After hybridization with target DNA (Cy3) from stems and EBC (Cy5), most EBC-derived ESTs are more highly expressed in the EBC.

Microarray determination of salt stress-dependent transcript changes

We focus on the responses of rice to salinity stress and drought. A first set of experiments targeted the response of young plants to salinity stress (150 mM), applied as a salt shock, in the moderately salt-tolerant line Pokkali in comparison with responses exhibited by the sensitive line IR29 (Kawasaki et al 2001). Under identical conditions, Pokkali survives long-term stress and resumes slow growth after 1 wk, while IR29 is destroyed within 1 to 2 d. The most important lesson from these experiments was that Pokkali showed early responses in the upregulation of transcripts, within minutes following stress imposition, which were absent in IR29. After several hours, both lines showed upregulation of transcripts that have previously been associated with abscisic acid- or ethylene-induced responses. Over the long term, the relative abundance of transcripts, both up- and downregulated, in Pokkali returned to pre-stress levels, while in IR29 a general decline in most transcripts preceded death. The transcript categories that showed higher expression levels throughout the time-course experiment in Pokkali were associated with stress defense reactions, such as radical oxygen-scavenging enzymes. At least under such shock treatments, which do not model the progression of stress in a field, the early responses seem to be important in determining the successful adaptation of the plants.

The results of these experiments, covering a time scale from minutes to days, indicated a surprisingly complex progression over time of fundamentally different categories of stress responses. Changes in the expression of regulatory function, such as calcium-dependent protein kinase, were detected early (15 min) after salt treatment. Pokkali, immediately and coincident with the decrease in photosynthesis, responded to salt stress at the level of transcription and/or transcript half-life, with the most dramatic changes occurring over the initial 3-h salt stress period. After 1 h, ribosomal proteins and translation elongation factors (EF1) were synchronously upregulated. During the time period between 3 and 6 h, most of the upregulated transcripts belonged to the categories of hormonal regulation and changes in metabolism. After 24 h, and continuing for up to 1 wk, the prevalent transcripts were those in the categories of cell defense to oxidative stress and further adaptations to the stress environment. The expression profiles of Pokkali and the salt-sensitive IR29 line were not fundamentally different although IR29 was distinguished from Pokkali by the lack of the initial response and a general decline in transcripts after and beyond 3 h of stress (Kawasaki et al 2001).

Comparisons between experiments

Microarray technology has matured such that conclusions can be drawn from individual arrays containing control sequences whose behavior is known. It is then possible to determine relative regulatory differences for sets of genes by a factor of approximately 2. RNA blot analyses using, for example, radioactively labeled probes confirm such differences. Comparing the results from different microarray hybridizations is more problematic because of unavoidable differences in target labeling (Cy5/Cy3) and uneven background levels that distinguish individual slides. Figure 2 provides an example indicating that comparisons are possible. By determining the over-

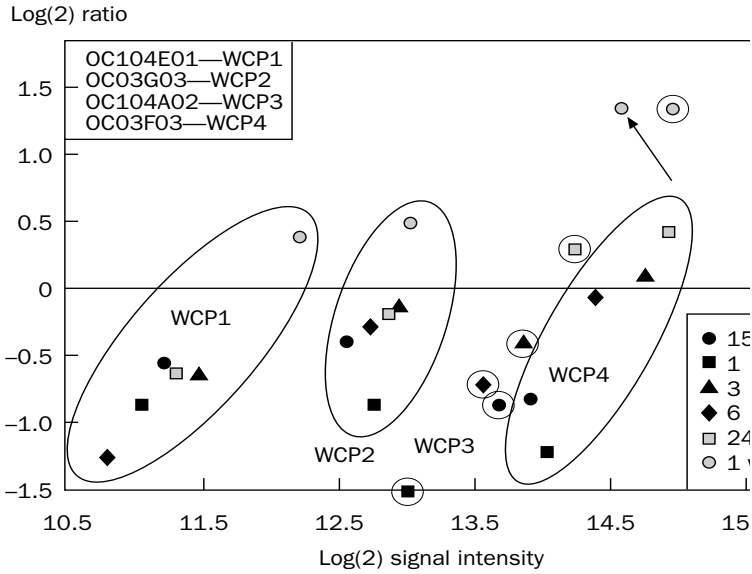


Fig. 2. Regulation and intensity of selected (putative) water channel proteins during a salt stress time course from rice cultivar Pokkali. Compared are four putative water channel protein transcripts. WCP1 shows high similarity to plasma membrane-located WCP and WCP2–4 are similar to tonoplast-specific WCP. Intensity and ratio of regulation relative to no-stress controls are shown for six time points during salt treatment. WCP1, 2, and 4 are surrounded as a group (and arrow for the 7-d time point for WCP4); time points for WCP3 are individually enclosed by circles.

all hybridization intensity and background levels of different slides, normalization can be achieved, assuming an overall similar or identical intensity even though individual transcripts may be variable. We compare the log(2) ratio (up- or downregulation) and the log(2) absolute signal intensity for four different putative water channel proteins. Plotted are the intensities and ratios for six time points during the stress experiment of rice cultivar Pokkali. For all putative water channel protein transcripts, intensity declined initially during salt stress. Lower expression levels for water channel proteins following salinity stress have been observed in other models (e.g., Yamada et al 1995). After only 1 wk (and after 24 h for WCP3 and WCP4), all water channels are more highly expressed than before stress and increases in abundance are obvious. Most significant, however, is the intensity distribution: the four water channel protein transcripts occupy intensity ranges that are similar in all comparisons, which were derived from six different microarray hybridizations. Also, WCP1 is a less abundant transcript than the others, and it encodes a plasma membrane-specific WCP. WCP2–4 encode tonoplast-specific transcripts that are typically more highly expressed than those located in the plasma membrane. From these experiments, it seems justified to state that comparisons are possible and that they generate data that can be incorporated into hypotheses on salt stress responses.

Conclusions

These analyses must be considered preliminary because of the limited number of transcripts that are included. Another limitation is the experimental setup that relies on salt shock, which does not simulate natural conditions. With the shock treatments, however, we hope to define the model systems and to derive a set of transcripts with microarrays of selected ESTs to which ESTs for functionally unknown yet stress-regulated transcripts are added, which should then be used under more natural stress conditions. Such experiments have not yet been reported with plants. Irrespective of the limited set of transcripts included in the experiments, a list of regulated transcripts is included (Table 5) because the nature of the regulated ESTs is informative. Included in the list are many transcripts that have previously been associated with salinity stress responses (Bohnert et al 1999, Hasegawa et al 2000). Table 5 also includes several functionally unknown or novel transcripts. The latter category seems to include transcripts that are only expressed in (salt) stressed plants and they are not represented in the large collection of rice ESTs deposited in GenBank. Among the approximately 4,000 rice ESTs that we have obtained so far, mainly from stressed plants, more than 300 are in the category "novel" after a comparison of our sequences against the 67,000 rice ESTs (October 2000; www.ncbi.nih.gov/dbEST/dbEST_summary.html).

The rapid development of microarray technologies has brought about a better understanding of handling, experimental setup, and statistical data analysis (Eisen et al 1998, Kehoe et al 1999, Deyholos and Galbraith 2001). The inclusion of larger sets of genes for comparisons is possible but equally valuable data will come from small arrays for selected pathways (e.g., nitrogen or one-carbon metabolisms), groups of functionally related genes (mitogen-activated protein kinases, P450), or conditionally or developmentally related genes (drought-induced unknown transcripts, cell cycle). We have not yet included other abiotic stress factors such as drought or low temperature in the analysis of rice ESTs, but such experiments are in progress (drought) or in the planning stage. The time-dependent progression of upregulation in high salinity in various categories of genes seems to indicate the temporal operation of different programs and several regulated transcripts. Based on these results, a challenging next step will be the analysis of the many functionally unknown genes that are regulated when plants experience abiotic stress.

Table 5. Strongly regulated transcripts during salinity stress in rice.[a]

Open reading frame no.	Annotation
0C01A03	Ripening-associated protein
0C01B06	Glycine-serine-rich protein (2)
0C01B11	Glutathione S-transferase
0C01C05	Unknown (TPR-containing protein)
0C01C11	Subtilisin-chymotrypsin inhibitor 2
0C01D02	40S ribosomal protein, S10
0C01D05	Cyclophilin 2
0C01D06	Low-temperature and salt-responsive protein, LTI6B
0C01E06	40S ribosomal protein, S6
0C01G03	40S ribosomal protein, S9
0C02C09	Sucrose synthase-2, identical to *sus2*
0C02E02	Putative integral membrane protein
0C02F03	No hit; rice EST exists
0C02F04	Cationic peroxidase (*Oryza sativa*)
0C02G09	Adenine phosphoribosyltransferase form 3
0C02H02	Fibrillarin
0C02H03	No hit; partial EST exists
0C02H04	No hit; no EST
0C03A06	60S ribosomal protein, L6 (YL16-like)
0C03B08	No hit; rice EST exists
0C03B10	Phenylalanine ammonia-lyase
0C03C06	No hit; rice EST exists
0C03C09	Unknown, similarity to streptomycin adenylase (*Lactococcus lactis*)
0C03C10	High-mobility group protein HMGd1 (*Zea mays*)
0C03D06	H^+-translocating ATPase (*Nicotiana plumbaginifolia*)
0C03E09	Putative protein similar to ADP-ribosylation factor 1 (*O. sativa*)
0C03F02	Sugar transporter (*Arabidopsis thaliana*)
0C03F03	Water channel protein (WCP3)
0C03F06	ABA- and stress-induced protein (Asr1) (*O. sativa*)
0C03H02	Glycine-serine-rich protein (1)
0C04B01	GDP-L-fucose synthetase
0C04B03	Histone H3
0C04B04	Putative alanine acetyl transferase (*A. thaliana*)
0C04C10	Unknown protein, contains chromosome condensation motifs (*A. thaliana*)
0C04D05	Peroxidase-1
0C04D08	Histone H3.3
0C04E01	S-adenosylmethionine decarboxylase
0C04E02	Hypothetical protein (*A. thaliana*) T6K21.50
0C04E04	Unknown, putative membrane protein (*A. thaliana*)
0C04F11	Alpha-tubulin (*O. sativa*)
0C04G06	Putative galactinol-raffinose galactosyltransferase (*Vigna angularis*)
0C04G08	Unknown, similar to Kunitz/bovine pancreatic trypsin inhibitor
0C04G09	Tonoplast intrinsic protein, γ-Tip
0C04H08	Ketol-acid reductoisomerase (KAR-1)
0C04H10	Unknown (contains zinc finger (C3HC4) ring finger motif) (*A. thaliana*)
0C05F01	Unknown (cysteine string protein)
0C05F11	Protoporphyrinogen IX oxidase, mitochondrial
0C05G11	60S acidic ribosomal protein, P0
0C101B09	60S ribosomal protein, L18A
0C101D06	40S ribosomal protein, S7
0C101D09	Trypsin inhibitor-(1) (Bowman-Birk type)
0E04H05	Unknown, putative DNA adenine methylase
0C01C03	Putative serine/threonine protein kinase

continued on next page

Table 5 continued.

Open reading frame no.	Annotation
OC02A01	Similarity to protein phosphatase-2C (*A. thaliana*)
OC04B07	Glycine hydroxymethyltransferase (SHM-2)
OE04F01	Unknown protein (*A. thaliana*) (AC009853)
OE04H12	Putative glucansynthase component
OC104A03	60S ribosomal protein, L44
OE201C03	Calcium-dependent protein kinase
OC104B10	dnaK-type molecular chaperone; hsp70 (*O. sativa*)
OC104F04	60S ribosomal protein, L18
OC104F12	40S ribosomal protein, S8
OC104G03	Putative transcription factor ATB2 (*A. thaliana*)
OC104G08	Nucleoside diphosphate kinase (NDPK-1)
OE04F03	S-adenosylmethionine decarboxylase 2, identical to SAMDC2 (*O. sativa*)
OC104H02	Late embryogenesis abundant-like protein (*Prunus armeniaca*)
OC104H06	GTP-binding protein (*Zea mays*)
OC03F07	Unknown (similar to CG14223 gene product [*Drosophila melanogaster*])
OC104A02	Water channel protein (WCP1)
OCP-06D12	Hin1 protein—common tobacco
OCP-06F12	Similar to DCL protein precursor, chloroplast (*Lycopersicon esculentum*)
OD102A07	Hypothetical protein K09C6.4 (*Caenorhabditis elegans*)
OC05B01	No hit; rice EST exists
OD102G08	60S ribosomal protein, L5
OD103A07	No hit; rice EST exists
OD103F11	Putative seed imbibition protein (*Persea americana*)
OE05F06	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase
OE04E07	GF14; 14-3-3 protein
OE06C09	Functionally unknown membrane protein
OE06E06	40S ribosomal protein, S30
OE06H12	Unknown protein (*A. thaliana*) (AC006418)
OE202F11	Water channel protein (WCP4)
OE04C08	Unknown (similar to LEF-3; AcMNPV orf67 [*Bombyx mori*])
OE04D07	Branched-chain amino acid aminotransferase
OE04D09	Similar to ARE1 (*Rattus norvegicus*)
OE04F10	Metallothionein-like protein, OsMT-1
OE05F02	Putative glycosylation enzyme
OE05G03	60S ribosomal protein, L30
OE05G09	α-galactosidase
OE05G10	Elongation factor-1α; EF-1α
OE06A08	Alanine transaminase
OE06A09	Putative protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13)—rabbit
OE06C12	Gigantea protein (*A. thaliana*)
OE06F04	S-adenosyl-L-methionine synthetase; pOS-SAMS2
OE06G06	Putative protein (*A. thaliana*) (AL035524)

[a]Regulation of selected ESTs according to microarray analysis (Kawasaki et al 2001). All ESTs have been deposited in GenBank and can be located by open reading frame number.

References

Adams MD, Soares MB, Kerlavage AR, Fields C, Venter JC. 1994. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nat. Genet. 4:373-380.

Adams P, Nelson DE, Yamada S, Chmara W, Jensen RG, Bohnert HJ, Griffiths H. 1998. Growth and development of *Mesembryanthemum crystallinum* (Aizoaceae). New Phytol. 138:171-190.

Barakat A, Gallois P, Raynal M, Mestre-Ortega D, Sallaud C, Guiderdoni E, Delseny M, Bernardi G. 2000. The distribution of T-DNA in the genomes of transgenic *Arabidopsis* and rice. FEBS Lett. 471:161-164.

Bohnert HJ, Su H, Shen B. 1999. Molecular mechanisms of salinity tolerance. In: Shinozaki K, editor. Cold, drought, heat, and salt stress: molecular responses in higher plants. Austin, Tex. (USA): R.G. Landes. p 29-60.

Buffard D, Valliant V, Esnault R. 1982. Complexity of polysomal polyadenylated RNAs in *Vicia faba* meristematic root cells. Eur. J. Biochem. 126:129-134.

Deyholos M, Galbraith D. 2001. High-density microarrays for gene expression analysis. Cytometry. (In press.)

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95:14863-14868.

Evans LT. 1998. Feeding the ten billion. Cambridge (UK): Cambridge University Press.

Flowers TJ, Yeo AR. 1995. Breeding for salinity tolerance in crop plants: Where next? Aust. J. Plant Physiol. 22:875-884.

Goldberg RB, Hoschek G, Kamalay JC, Timberlake WE. 1978. Sequence complexity of nuclear and polysomal RNA in leaves of the tobacco plant. Cell 14:123-131.

Hasegawa PM, Bressan RA, Zhu J-K, Bohnert HJ. 2000. Plant cellular and molecular responses to high salinity. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:463-499.

Jain RK, Selvaraj G. 1997. Molecular genetic improvement of salt tolerance in plants. Biotechnol. Annu. Rev. 3:245-267.

Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, Kim C, Jang S, Yang K, Nam J, An K, Han MJ, Sung RJ, Choi HS, Yu JH, Choi JH, Cho SY, Cha SS, Kim SI, An G. 2000. T-DNA insertional mutagenesis for functional genomics in rice. Plant J. 22:561-570.

Kamalay JC, Goldberg RB. 1984. Organ-specific nuclear RNAs in tobacco. Proc. Natl. Acad. Sci. USA 81:2801-2805.

Kawasaki S, Deyholos M, Borchert C, Brazille S, Kawai K, Galbraith DW, Bohnert HJ. 2001. Temporal succession of salt stress responses in rice by microarray analysis. Plant Cell 13: (in press.)

Kehoe DM, Villand P, Somerville S. 1999. DNA microarrays for studies of higher plants and other photosynthetic organisms. Trends Plant Sci. 4:38-41.

Khush GS. 1999. Green revolution: preparing for the 21st century. Genome 42:646-655.

Kiper M, Bartels D, Herzfeld F, Richter G. 1979. The expression of a plant genome in hnRNA and mRNA. Nucleic Acids Res. 6:1961-1978.

Levitt J. 1980. Responses of plant to environmental stress chilling, freezing, and high temperature stresses. 2nd edition. New York, N.Y. (USA): Academic Press.

Lin X, Kaul S, Rounsley S, Shea TP, Benito M-I, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldblyum TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Shea M, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Nierman WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter C. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761-769.

Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. 1999. High density synthetic oligonucleotide arrays. Nat. Genet. 21(Suppl.):20-24.

Marra MA, Hillier L, Waterston RH. 1998. Expressed sequence-tags: ESTablishing bridges between genomes. Trends Genet. 14:4-7.

Mayer K and 228 others. 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana.* Nature 402:769-777.

McCallum CM, Comai L, Greene EA, Henikoff S. 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol. 123:439-442.

Miflin B. 2000. Crop improvement in the 21st century. J. Exp. Bot. 51:1-8.

Parinov S, Sundaresan V. 2000. Functional genomics in *Arabidopsis:* large-scale insertional mutagenesis complements the genome sequencing project. Curr. Opin. Biotechnol. 11:157-161.

Phimister B. 1999. A note on nomenclature. Nat. Genet. 21(Suppl.):1.

Postel S. 1999. Redesigning irrigated agriculture. In: Brown LR, Flavin C, French H, editors. State of the world 2000. New York, N.Y. (USA): W.W. Norton & Co. p 39-58.

Richmond T, Somerville S. 2000. Chasing the dream: plant EST microarrays. Curr. Opin. Plant Biol. 3:108-116.

Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-470.

Somerville C, Somerville S. 1999. Plant functional genomics. Science 28:380-383.

van Hal NL, Vorst O, van Houwelingen AM, Kok EJ, Peijnenburg A, Aharoni A, van Tunen AJ, Keijer J. 2000. The application of DNA microarrays in gene expression analysis. J. Biotechnol. 78:271-280.

Wang J. 2000. From DNA biosensors to gene chips. Nucleic Acids Res. 15:3011-3016.

Yamada S, Katsuhara M, Kelly W, Michalowski CB, Bohnert HJ. 1995. A family of transcripts encoding water channel proteins: tissue-specific expression in the common ice plant. Plant Cell 7:1129-1142.

Notes

Authors' addresses: H.J. Bohnert, Department of Biochemistry and Molecular Biophysics, and Department of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA, phone: 520-621-7961, fax: 520-621-1697, E-mail: bohnerth@email.arizona.edu; S. Kawasaki, C.B. Michalowski, Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721, USA; Hong Wang, M. Deyholos, D. Galbraith, Department of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA; N. Ozturk, Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721, USA, permanent address: Tubitak Marmara Research Center, Research Institute for Genetic Engineering and Biotechnology, Plant Biotechnology Laboratory, PK. 21, 41470, Gebze-Kocaeli, Turkey.

Acknowledgments: We thank Chris Borchert, Susan Brazille, Caitlin Murphy, Christiane Normand, Chris Palacio, and Stacia Sant for help with DNA sequencing and annotations. We thank Drs. Renee Lafitte and Len Wade (IRRI, Philippines) and Scott Chapman (University of Queensland, Australia) for advice and discussions. Our work is supported by a grant from the National Science Foundation (98-13360; "Functional genomics of plant stress tolerance").

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Molecular dissection of cell death in rice

H. Uchimiya, M. Kawai, M. Yamaguchi, and M. Umeda

Programmed cell death has been observed in various developmental processes in higher plants. The presence of air-filled spaces known as aerenchyma in rice is considered to be an important anatomical adaptive feature necessary for plant survival under flooded conditions. Ample evidence has been presented to show that this system provides a diffusion path for the transportation of oxygen from aerial plant parts to roots. In addition, the molecular dissection of cell death is discussed.

Cyclin-dependent protein kinases (CDKs) are serine/threonine protein kinases that are involved in the regulation of the eukaryotic cell cycle. The growing list of CDKs in plant cells suggests that, during development, each CDK may play a specific role at a specific time in the cell cycle. Rice plants also express different kinds of CDKs. We present the outcome of our work indicating that the division of rice cells might be regulated by different sets of CDK/cyclin, leading to cell proliferation.

The destination of plant cell life is under strict control with regard to adaptive features of biological mechanisms. Topics related to this control will be presented in terms of cellular and molecular understanding of cell death and cell cycle regulation in rice plants.

Cell death in rice: adaptive cellular pathways leading to water tolerance

The presence of gas-filled spaces, known as aerenchyma, in numerous plant species is considered to be an important anatomical adaptive feature necessary for survival under flooded conditions (Kawase 1981, Kawai and Uchimiya 2000). Ample evidence has been presented to show that aerenchyma provide a diffusion path of low resistance for the transport of oxygen from aerial plant parts to roots or rhizomes in a waterlogged, O_2-deficient environment (Armstrong 1971). The formation of aerenchyma is thought to occur by either lysigeny or schizogeny (Smirnoff and Crawford 1983). Lysigenous gas spaces form via cell lysis, whereas schizogenous spaces form

by cells separating during tissue development. Our study focuses on the lysigenous gas spaces produced in the primary roots of rice. During the normal course of development, many cells in the cortex collapse to form aerenchyma.

In maize, ethylene has been implicated in signaling cell death during the lysigenous formation of aerenchyma in adventitious roots subjected to hypoxia (Drew et al 1979). According to Jackson (1994), a limited amount of oxygen outside the root may favor the development of an anoxic stele core, where the absence of oxygen can be expected to increase 1-aminocyclopropane-1-carboxylic acid (ACC) synthase levels, leading to the synthesis of additional ACC. This ACC would then enter the encircling hypoxic cortex, where the small amount of oxygen present (Armstrong and Armstrong 1994) is sufficient to allow oxidative conversion of ACC to ethylene by ACC oxidase.

In rice, however, Jackson et al (1985) reported that aerenchyma formation in adventitious roots was not controlled by ethylene or a low partial pressure of oxygen. On the other hand, Das and Jat (1977) reported that flooded conditions enhanced aerenchyma formation in the root cortex in some rice cultivars. Furthermore, Justin and Armstrong (1991) demonstrated that ethylene can promote aerenchyma formation in adventitious roots of rice, and showed that cultivar-specific differences explain the discrepancies.

Gas-filled spaces in rice arise from the separation of cell walls from adjacent cells so that the radial walls from the collapsing cells aggregate together, forming "forks," leaving a large gas-filled space or lacuna between them (Clark and Harris 1981). In an electron-microscope study, Webb and Jackson (1986) proposed that the processes of aerenchyma formation differ in rice and maize for the order in which events occur. In maize roots, cellular collapse was preceded by the loss of tonoplast integrity (Campbell and Drew 1983). In rice, however, the middle lamella degenerated, followed by cell wall disintegration and the loss of tonoplast integrity (Webb and Jackson 1986). Although there has been some controversy about the process of aerenchyma formation in rice and maize, it is now generally agreed that the aerenchyma of rice roots is rapidly formed as an integral part of ordinary root development (Webb and Armstrong 1983, Jackson et al 1985). The detailed mechanisms of aerenchyma formation are not well understood.

Our investigation (Kawai et al 1998, Samarajeewa et al 1999) contributes to a mechanistic understanding of the events leading to cortical cell death, which eventually results in the formation of aerenchyma in the primary roots of rice. The most important finding is that cell death began at a specific cell position. The first cells to collapse were located at the center of the cortical tissues and cell collapse never began at the peripheral cortical cells. The locations of cells undergoing lysis appeared precise. This indicates the existence of a targeting mechanism for initiating the first cell death. Cells in this position were characterized by being shorter but larger in radial diameter than other cortical cells. These features were generated by the nonuniform expansion of cortical cells in each position. The cells of position 5 tended to expand radially more than tangentially, unlike peripheral cortical cells.

The development of aerenchyma is often associated with the enlargement of cortical cells. Schussler and Longstreth (1996) also reported that the cells undergoing

lysis appeared to be larger in diameter than the other cortical cells in *Sagittaria lancifolia* roots. Kawase (1974) demonstrated that cortical cells of sunflower stem treated with cellulase were enlarged radially, and some disintegrated, leading to an intercellular space. Cell expansion depends on change in cell walls, and radial cell enlargement was explained by Fan and Maclachlan (1996) as an induction of cellulase acting on microfibrils, which are predominantly oriented transversely. Cellulase activity is increased by ethylene in plants (Horton and Osborne 1967). It has already been established that aerenchyma development in flooded maize roots is associated with increased activity of cellulase (Horton and Osborne 1967), and that this process is promoted by ethylene (Drew et al 1979). Saab and Sachs (1996) cloned a flooding-inducible xyloglucan endo-transglycosylase (*XET*) gene, a cell-wall-loosening enzyme, and showed that ethylene treatment under aerobic conditions induced *XET* gene expression.

Based on a microscopic analysis, it has been reported that the first clearly observed event in aerenchyma formation in maize roots is tonoplast disruption (Campbell and Drew 1983). In our study, the first cells to be stained with neutral red were at position 5 in the cortex. This result suggests that vacuoles may mature (i.e., acidosis) more rapidly in position 5 than in other positions. After this process, cytoplasmic acidification resulting in the loss of tonoplast integrity may occur by the diffusion of H^+ from vacuoles. The signal that provokes vacuole disruption is unknown.

Evans blue clearly stained cortical cells in position 5 first in section III. These cells comprise the upper range of the cells that were stained first with neutral red (section II) and comprise the lower range of the cells in which collapse began (section IV). Furthermore, Evans blue did not stain concave cells (i.e., collapsing cells). These results indicate that cells in position 5 develop vacuoles, followed by a loss of membrane integrity, and collapse. Several studies show that the ability of plant cells to take up Evans blue signifies an irreversible change that leads to cell death (Gaff and Okong'O-Ogola 1971).

Webb and Jackson (1986) observed the process of aerenchyma formation in rice roots by transmission and cryo-scanning electron microscopy and reported that breakdown of the cell wall preceded the loss of tonoplast integrity. This result seems to contradict our finding. In maize, cell wall degradation is preceded by tonoplast disruption and the dispersal of cytoplasm. Until this stage is reached, cell walls remain unaltered (Campbell and Drew 1983). Based on our data, it is difficult to determine the sequence of the processes that occur inside of cells. However, we have shown in rice that the ability to be stained by Evans blue, which may be related to plasma membrane disruption, precedes cell collapse.

Once cell collapse occurred, the cells lost contact with neighboring (tangential) cells and were sequentially destroyed in a radial direction in cortical parenchyma tissues. Using electron microscopy, van der Weele et al (1996) observed an earlier stage of cell death, which began with the loss of vacuolar solutes and the collapse of cells in maize, and hypothesized that dying cells may release a message that initiates the process of cell death in neighboring cells. The movement of small molecules between plant cells occurs through a cytoplasmic syncytium termed the symplast

(Erwee and Goodwin 1983). The symplast is known to be important in ion and metabolite transport, especially where apoplastic barriers are found, as in the root cortex (Clarkson and Robards 1975). Morphological investigations have provided evidence that symplasmic transport is mediated by specialized trans-cell wall structures called plasmodesmata.

The highly selective death in the specific cortical cells is reminiscent of programmed cell death in animal cells. In animals, apoptosis has been known as an active control mechanism that is important in developmental and pathological processes (Yuan and Horvits 1990). Our knowledge of programmed cell death in plants lags behind that of animal systems. Programmed cell death is thought to occur during the autolysis of xylem vessels (Fukuda and Komamine 1982), during sex determination (Delong et al 1993), and in other processes. Mittler and Lam (1995) demonstrated in pea that the nuclei of vessel elements undergoing programmed cell death contained fragmented nuclear DNA. This finding may be evidence of the activation of a DNA-degradation mechanism prior to the final disruption of the nucleolus that occurs during autolysis in this differentiation process. The *dad1* gene, known as a suppressor of programmed cell death in *Caenorhabditis elegans* and mammals, has been isolated during petal senescence of pea (Orzáez and Granell 1997), and Tanaka et al (1997) showed that the rice *dad* gene functions as a suppressor of cell death in animal cells. We also isolated a rice gene similar to the *Bax* inhibitor in yeast (Kawai et al 1999). These reports show that at least some characteristics may be shared between animal and plant programmed cell death. Even in collapsing cells, we observed that some cells contained nuclei. Thus, further studies of DNA degradation are needed.

In relation to the mechanism of plant cell death, it is interesting to speculate that the sequential spread of cell death may be due to H_2O_2 produced as an oxidative burst, since a high dose of H_2O_2 induces cell death in higher plants (Levine et al 1994). It would be interesting to investigate further the interaction of these signals in conjunction with ethylene, which has been shown to stimulate aerenchyma formation.

Figure 1 presents a proposed model of the processes leading to cortical cell death in rice roots. After germination, cell division regulates the number of cells in the region where the root meristem is located. Longitudinal cell elongation as well as horizontal cell expansion follow. After these events, cell acidosis and tonoplast breakdown occur in the mid cortex, followed by cell death, which coincides with movement of injected probes through plasmodesmata in a radial direction. The specific step involved in the first cell death in aerenchyma formation of rice roots cannot be identified unambiguously based on the present evidence, but the program leading to tonoplast disruption and first cell death might begin in response to developmental signals, which may include endogenous ethylene. Berg et al (1995) demonstrated that the fate of cells in the root meristem is under positional control. It is interesting that the fates of both meristematic cells and those undergoing successive cell death in plant roots are under strict spatial-related regulation. Further investigation is necessary to clarify the process of aerenchyma formation in the root cortex.

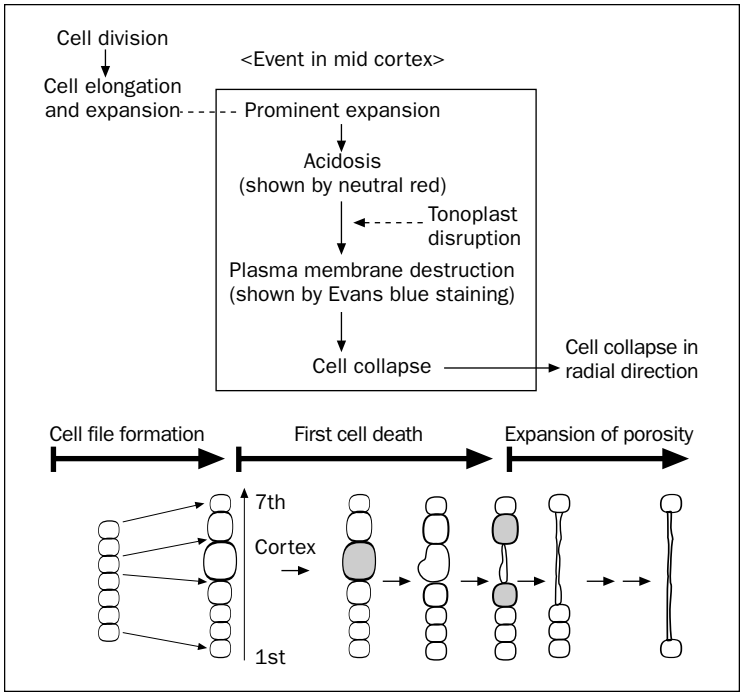


Fig. 1. A proposed model of the processes leading to cortical cell death in rice roots.

Cell cycle genes in rice: essential components for sustainable cell proliferation

Cyclin-dependent protein kinases (CDKs) are serine/threonine protein kinases that are involved in the regulation of the eukaryotic cell cycle. A single major CDK has been identified in the fission yeast *Schizosaccharomyces pombe* (CDC2) and in the budding yeast *Saccharomyces cerevisiae* (CDC28). However, the growing list of CDKs in human cells suggests that, during development, each CDK in metazoans plays a specific role at a specific time in the cell cycle. CDKs are activated by the binding of cyclins and phosphorylation (Morgan 1995). Each CDK interacts with a specific subset of cyclins, and the size of this subset varies among CDKs. For example, CDC28 can associate with many different cyclins, whereas human CDC2 interacts with relatively few (Nigg 1995). A short conserved amino acid sequence, PSTAIRE, in CDKs is responsible for the binding of cyclins, which activate CDKs by changing the conformation at the catalytic site (Jeffrey et al 1995, Morgan 1996). These cyclins also function in the targeting of CDKs to specific substrates or subcellular locations (Hoffmann et al 1993).

Plants also express different kinds of CDK and multiple genes for CDKs have been found in *Arabidopsis*, alfalfa, rice, soybean, maize, and *Antirrhinum*. A correla-

tion between the abundance of CDK transcripts and the proliferative state of cells was demonstrated in *Arabidopsis*, maize, and alfalfa. However, it has also been shown that, in *Arabidopsis*, transcripts of *cdc2aAt* are localized not only in dividing cells but also in differentiated tissues, such as the parenchyma of the vascular cylinder and the pericycle, which contains cells responsible for formation of lateral roots. Moreover, expression of *cdc2aAt* could be induced without cell division in suspension cultures. These results suggest that at least some CDK transcripts might be correlated with the acquisition of the competence to divide rather than with the actual division of cells.

Genes for four different CDKs have been isolated from rice: *cdc2Os1*, *cdc2Os2* (Hashimoto et al 1992), *cdc2Os3* (Kidou et al 1994, Umeda et al 1998), and *R2* (Hata 1991), which is similar to the gene for a CDK-activating kinase (CAK) that is required for activation of CDK by phosphorylation of a conserved threonine residue in the so-called T loop (Morgan 1995). In contrast, a new type of *CAK* gene was isolated in *Arabidopsis* (Umeda et al 1998). The deduced amino acid sequence of Cdc2Os3 showed that it is distinct from proteins in the CDC2/CDC28 family (Kidou et al 1994), whereas Cdc2Os1 and Cdc2Os2 are closely related to the homologues of CDC2 that have been isolated from various organisms. The *cdc2Os1* gene was able to partially complement a temperature-sensitive mutation in the *cdc28* gene in yeast, but *cdc2Os2* and *R2* were unable to complement the same mutation (Hashimoto et al 1992).

We have analyzed levels of transcript of genes for CDKs in rice plants by *in situ* hybridization (Umeda et al 1999b). We found that *cdc2Os3*, which has an altered PSTAIRE sequence, was expressed in a cell cycle-dependent manner, whereas the other two CDKs with the conserved PSTAIRE motif were expressed throughout the cell cycle. The transcript and protein product of *cdc2Os3* were abundant from the G2 to the M phase, an indication that the Cdc2Os3 protein might function in mitosis.

We found that *cdc2Os1* and *cdc2Os2* were also expressed in the sclerenchyma, the pericycle, and the parenchyma of the central cylinder in the differentiated zone of roots. In parts of *Arabidopsis* roots beyond the apical meristem, expression of *cdc2aAt* is restricted to the parenchyma of the vascular cylinder and to the pericycle cells (Martinez et al 1997). Both Cdc2Os1 and Cdc2Os2 of rice are closely related to *Cdc2aAt* of *Arabidopsis* at the amino acid level, and may also be correlated with the competence of these cells to divide. The pericycle is a differentiated tissue but it retains the potential to divide and is responsible for lateral root formation. The expression of *cdc2Os1* and *cdc2Os2* in the parenchyma of the central cylinder suggests that this cell layer might engage in some mitotic activity that contributes to the thickening of primary roots in rice plants. The rice-specific expression of these two *cdc2* genes in the sclerenchyma remains to be investigated.

In contrast to transcripts of *cdc2Os1*, *cdc2Os2*, and *R2*, transcripts of *cdc2Os3* were distributed with a patchy pattern in the dividing region of the root apex. Such a pattern was also observed in the region near the shoot meristem and in the primordia for lateral root formation. Counterstaining of sections with DAPI (4,6-diamidino-2-phenylindole) indicated that almost all cells with mitotic nuclei contained *cdc2Os3* transcripts, whereas cells forming cell plates had trace levels of transcripts. In double-labeling experiments with probes specific for transcripts of a gene for histone H4 and *cdc2Os3*, signals did not overlap, an indication that expression of *cdc2Os3* did not

extend to the S phase. Furthermore, treatment of seedlings with hydroxyurea, which blocks cells in the early S phase, inhibited the patchy expression of *cdc2Os3* at the root apex, whereas transcripts were still detectable in roots treated with colchicine, which blocks cells in mitosis. The patchy pattern on colchicine-treated sections might reflect the partial synchrony of cell division in this case. Thus, transcripts of *cdc2Os3* appeared to be abundant from the G2 to the M phase but almost disappeared when cells had completed mitosis at telophase.

Both Cdc2Os1 and Cdc2Os2 include the characteristic PSTAIRE domain (Hashimoto et al 1992) and are classified as PSTAIRE CDKs on the phylogenetic tree. PSTAIRE CDKs, such as products of *cdc2aAt* in *Arabidopsis*, *Amcdc2a* and Amcdc2b in *Antirrhinum*, and *cdc2MsA* and *cdc2MsB* in alfalfa, are expressed throughout the cell cycle (Magyar et al 1997). Thus, the products of *cdc2Os1* and *cdc2Os2* can also be classified as PSTAIRE CDKs in terms of their pattern of expression. Several plant CDKs with altered PSTAIRE motifs have been reported, namely, the products of Amcdc2c and Amcdc2d in *Antirrhinum*; *cdc2MsC*, *cdc2MsD*, *cdc2MsE*, and *cdc2MsF* in alfalfa; and cdc2bAt in *Arabidopsis*. Rice *cdc2Os3* encodes a PPTALRE sequence, which is the same as those of AmCdc2c, Cdc2MsD, and Cdc2bAt. However, when Cdc2Os3 was compared with the other non-PSTAIRE CDKs in the whole region, it is located close to Amcdc2d and cdc2MsF rather than to AmCdc2c, Cdc2MsD, or Cdc2bAt. Transcripts of *cdc2s* belonging to the group including *cdc2Os3* are abundant from the G2 to the M phase (Magyar et al 1997). On the other hand, Amcdc2c is expressed from the mid-S phase to early M phase, and *cdc2bAt* is preferentially expressed during the S and G2 phases. Transcripts of *cdc2MsD* are abundant at the G2/M phase and are also detected just after alfalfa cells are released from arrest by aphidicolin (Magyar et al 1997). Therefore, we propose that the products of *cdc2Os3*, *Amcdc2d*, and *cdc2MsF* form a distinct subclass of non-PSTAIRE CDKs that are preferentially expressed from the G2 to the M phase. Our results also indicated that domains other than the PSTAIRE region are important for the function of non-PSTAIRE CDKs in distinct cell cycle phases. The alfalfa genes cdc2MsC and *cdc2MsE* as well as rice *R2*, which also encode divergent PSTAIRE sequences, are expressed throughout the cell cycle, and the encoded amino acid sequences are distinct from those of other CDKs (Magyar et al 1997).

The level of Cdc2Os3 protein was reduced by treatment of cultured cells with hydroxyurea but not with colchicine. The histone H1-kinase activity associated with Cdc2Os3 was correlated with the level of the protein. Expression of *cdc2Os3* is probably controlled at the transcriptional level and Cdc2Os3 protein probably accumulates from the G2 to the M phase.

When rice *R2* was overexpressed in a CAK-deficient mutant of budding yeast, it suppressed the temperature sensitivity of the mutation (Yamaguchi et al 1998). Immunoprecipitates of rice proteins with the anti-R2 antibody phosphorylated human CDK2, one of the rice CDKs (Cdc2Os1), and the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II of *Arabidopsis*. Mutational analysis indicated that R2 phosphorylated the threonine residue within the T loop of CDK2 and Cdc2Os1. R2 was found mainly in two protein complexes that had molecular masses of 190 kDa and 70 kDa, whereas the CDK- and CTD-kinase activities associated with R2 were identified

in a complex of 105 kDa. These results indicate that R2 is closely related to CAKs of animals and fission yeast in terms of its phosphorylation activity.

Plant cyclins are classified into either mitotic cyclins (A- and B-type) or D-type cyclins (Mironov et al 1999, Renaudin et al 1996). Figure 2 shows the systematic classification of plant cyclin genes. We isolated a third class of cyclins from poplar and rice plants (Yamaguchi et al 2000). Their deduced amino acid sequences had high similarity to that of vertebrate cyclin H, especially in the cyclin box region. Poplar and rice cyclin H were expressed in all tissues although their transcripts were abundant in suspension cells and tissues with cell division activity. Moreover, transcripts of *Os;cycH;1* were most abundant in the meristematic region of the growing rice internode, but were also found at lower levels in elongating and differentiated cells in deepwater rice. This is in contrast to mitotic cyclin genes whose expression is restricted to dividing cells (Umeda et al 1999a). In the growing internode, *Os;cycH;1* transcripts clearly increased at 6 h after submergence and then moderately decreased. Lorbiecke and Sauter (1998) demonstrated that, at 4 h after submergence, more cells in the meristem began to replicate

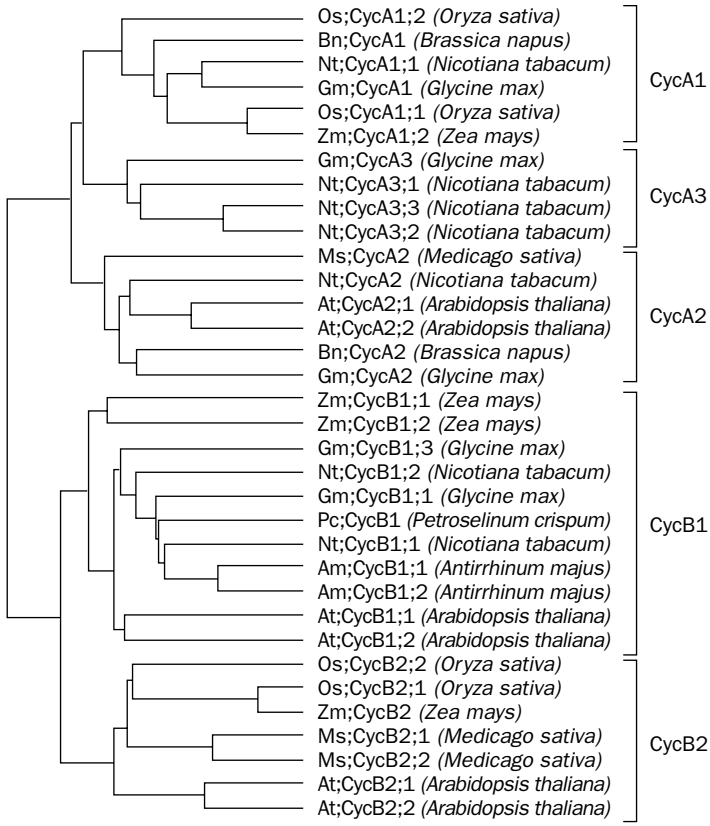


Fig. 2. Systematic classification of plant cyclin genes.

and, at 6 h, the population of S phase cells was at its peak with a threefold increase over uninduced plants. Between 6 h and 18 h of submergence, the number of cells in the S phase leveled out to approximately twice the number found at 0 h. Therefore, the kinetic pattern of *Os;cycH;1* expression was similar to the kinetic pattern of S phase cells in the meristem. Moreover, in partially synchronized suspension cells of *Os;cycH;1*, transcripts were abundant in the S phase. These data suggested that *Os;cycH;1* expression was induced when cells entered the S phase at an elevated level and that CAK activity was required for S phase progression.

In plants, only little information is available about which cyclin binds to which CDK to form an active kinase complex. Two *Arabidopsis* D-type cyclins, At;CycD1;1 and At;CycD4;1, were capable of interacting with Cdc2aAt and Cdc2bAt in the yeast two-hybrid system (De Veylder et al 1997, 1999). Immunoprecipitates of tobacco extract with anti-Nt;CycD3;1 antibody contained a protein that cross-reacted with the anti-PSTAIRE antibody, and that Cdc2Nt1-CycD3 complex was able to phosphorylate Rb-related protein in vitro, whereas Cdc2Nt1 alone did not exhibit a phosphorylation activity (Nakagami et al 1999). However, whether Nt;CycD3;1 specifically activates Cdc2Nt1 in tobacco remains to be determined. Therefore, rice cyclin H would be the first plant cyclin whose specific partner has been identified (Yamaguchi et al 2001).

References

Armstrong W. 1971. Radial oxygen losses from rice roots as affected by distance from the apex, respiration and water logging. Physiol. Plant 25:192-197.

Armstrong J, Armstrong W. 1994. Chlorophyll development in mature lysigenous and schizogenous root aerenchyma provides evidence of continuing cortical cell viability. New Phytol. 126:493-497.

Berg C, Willemsen V, Hage W, Weisbeek P, Scheres B. 1995. Cell fate in the *Arabidopsis* root meristem determined by directional signaling. Nature 378:62-65.

Campbell R, Drew MC. 1983. Electron microscopy of gas space (aerenchyma) formation in adventitious roots of *Zea may*s L. subjected to oxygen shortage. Planta 157:350-357.

Clark LH, Harris WH. 1981. Observation on the root anatomy of rice (*Oryza sativa* L.). Am. J. Bot. 68:154-161.

Clarkson DT, Robards AW. 1975. The endodermis, its structural development and physiological role. In: The development and function of roots. London (UK): Academic Press. p 415-436.

Das DK, Jat RL. 1977. Influence of three soil water regimes on root porosity and growth of four rice varieties. Agron. J. 69:197-210.

De Veylder L, Segers G, Glab N, Van Montagu M, Inzé D. 1997. Identification of proteins interacting with the *Arabidopsis* Cdc2aAt protein. J. Exp. Bot. 48:2113-2114.

De Veylder L, Engler JA, Burssens S, Manevski A, Lescure B, Van Montagu M, Engler G, Inzé D. 1999. A new D-type cyclin of *Arabidopsis thaliana* expressed during lateral root primordia formation. Planta 208:453-462.

Delong A, Calderon-Urrea A, Dellaporta SL. 1993. Sex determination gene *TASSELSEED 2* of maize encodes a short-chain alcohol dehydrogenase required for stage-specific floral organ abortion. Cell 74:757-768.

Drew MC, Jackson MB, Giffard S. 1979. Ethylene-promoted adventitious rooting and development of cortical air spaces (aerenchyma) in roots may be adaptive responses to flooding in *Zea mays* L. Planta 147:83-88.

Erwee MG, Goodwin PB. 1983. Characterization of the *Egeria densa* Planch. leaf symplast. Planta 158:320-328.

Fan DF, Maclachlan GA. 1966. Control of cellulase activity by indoleacetic acid. Can. J. Bot. 44:1025-1034.

Fukuda H, Komamine A. 1982. Lignin synthesis and its related enzymes as markers of tracheary-element differentiation in single cells isolated from the mesophyll of *Zinnia elegans*. Planta 155:423-430.

Gaff DF, Okong'O-Ogola O. 1971. The use of non-permeating pigments for testing the survival of cells. J. Exp. Bot. 22:756-758.

Hashimoto J, Hirabayashi T, Hayano Y, Hata S, Ohashi Y, Suzuka I, Utsugi T, Toh-E A, Kikuchi Y. 1992. Isolation and characterization of cDNA clones encoding *cdc2* homologues from *Oryza sativa*: a functional homologue and cognate variants. Mol. Gen. Genet. 233:10-16.

Hata S. 1991. cDNA cloning of a novel *cdc2*$^{+}$*/CDC28*-related protein kinase from rice. FEBS Lett. 279:149-152.

Hoffmann I, Clarke PR, Marcote MJ, Karsenti E, Draetta G. 1993. Phosphorylation and activation of human cdc25-C by cdc2-cyclin B and its involvement in the self-amplification of MPF at mitosis. EMBO J. 12:53-63.

Horton RF, Osborne DJ. 1967. Senescence, abscission and cellulase activity in *Phaseolus vulgaris*. Nature 214:1086-1088.

Jackson MB, Fenning TM, Jenkins W. 1985. Aerenchyma (gas-space) formation in adventitious roots of rice (*Oryza sativa* L.) is not controlled by ethylene or small partial pressures of oxygen. J. Exp. Bot. 36:1566-1572.

Jackson MB. 1994. Root-to-shoot communication in flooded plants: involvement of abscisic acid, ethylene, and 1-aminocyclopropane-1-carboxylic acid. Agron. J. 86:775-782.

Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massagu J, Pavletich NP. 1995. Mechanism of CDK activation revealed by the structure of a cyclin A-CDK2 complex. Nature 376:313-320.

Justin SHFW, Armstrong W. 1991. Evidence for the involvement of ethylene in aerenchyma formation in adventitious roots of rice (*Oryza sativa* L.). New Phytol. 118:49-62.

Kawai M, Umeda M, Uchimiya H. 1998. Stimulation of adenylate kinase in rice seedlings under submergence stress. J. Plant Physiol. 152:533-539.

Kawai M, Samarajeewa PK, Berrero RA, Nishiguchi M, Uchimiya H. 1998. Cellular dissection of the degradation pattern of cortical cell death during aerenchyma formation of rice roots. Planta 204:277-287.

Kawai M, Uchimiya H. 2000. Coleoptile senescence in rice (*Oryza sativa* L.). Ann. Bot. 86:405-414.

Kawai M, Pan L, Reed JC, Uchimiya H. 1999. Evolutionally conserved plant homologue of the *Bax Inhibitor-1(BI-1)* gene capable of suppressing *Bax*-induced cell death in yeast. FEBS Lett. 464:143-147.

Kawase M. 1974. Aerenchyma development is often associated with a radial enlargement of cortical cells. Physiol. Plant 31:29-38.

Kawase M. 1981. Anatomical and morphological adaptation of plants to waterlogging. Hortic. Sci. 16:30-34.

Kidou S, Umeda M, Uchimiya H. 1994. Nucleotide sequence of rice (*Oryza sativa* L.) cDNA homologous to *cdc2* gene. DNA Sequence 5:125-129.

Levine A, Tenhaken R, Dixon R, Lamb C. 1994. H_2O_2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. Cell 79:583-593.

Lorbiecke R, Sauter M. 1998. Induction of cell growth and cell division in the intercalary meristem of submerged deepwater rice (*Oryza sativa* L.). Planta 204:140-145.

Magyar Z, Mészáros T, Miskolczi P, Deák M, Fehér A, Brown S, Kondorosi É, Athanasiadis A, Pongor S, Bilgin M, Bakó L, Koncz C, Dudits D. 1997. Cell cycle phase specificity of putative cyclin-dependent kinase variants in synchronized alfalfa cells. Plant Cell 9:223-235.

Martinez AM, Afshar M, Martin F, Cavadore JC, Labbé JC, Dorée M. 1997. Dual phosphorylation of the T-loop in cdk7: its role in controlling cyclin H binding and CAK activity. EMBO J. 16:343-354.

Mironov V, De Veylder L, Van Montagu M, Inzé D. 1999. Cyclin-dependent kinases and cell division in plants: the nexus. Plant Cell 11:509-521.

Mittler R, Lam E. 1995. In situ detection of nDNA fragmentation during the differentiation of tracheary elements in higher plants. Plant Physiol. 108:489-493.

Morgan DO. 1995. Principles of CDK regulation. Nature 374:131-134.

Morgan DO. 1996. The dynamics of cyclin-dependent kinase structure. Curr. Opin. Cell Biol. 8:767-772.

Nakagami H, Sekine M, Murakami H, Shinmyo A. 1999. Tobacco retinoblastoma-related protein phosphorylated by a distinct cyclin-dependent kinase complex with Cdc2/cyclin D in vitro. Plant J. 18:243-252.

Nigg EA. 1995. Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. BioEssays 17:471-480.

Orzáez D, Granell A. 1997. The plant homologue of the defender against apoptotic death gene is down-regulated during senescence of flower petals. FEBS 404:275-278.

Renaudin J-P, Doonan JH, Freeman D, Hashimoto J, Hirt H, Inzé D, Jacobs T, Kouchi H, Rouzé P, Sauter M, Savourá A, Sorrell DA, Sundaresan V, Murray JA. 1996. Plant cyclins: a unified nomenclature for plant A-, B- and D-type cyclins based on sequence organization. Plant Mol. Biol. 32:1003-1018.

Saab IN, Sachs MM. 1996. A flooding-induced xyloglucan endo-transglycosylase homolog in maize is responsive to ethylene and associated with aerenchyma. Plant Physiol. 112:385-391.

Samarajeewa PK, Barrero RA, Umeda-Hara C, Kawai M, Uchimiya H. 1999. Cortical cell death, cell proliferation, macromolecular movements and *rTipl* expression pattern in roots of rice (*Oryza sativa* L.) under NaCl stress. Planta 207:354-361.

Schussler E, Longstreth DJ. 1996. Aerenchyma develops by cell lysis in roots and cell separation in leaf petioles in *Sagittaria lancifolia* (Alismataceae). Am. J. Bot. 83:1266-1273.

Smirnoff N, Crawford RMM. 1983. Variation in the structure and response to flooding of root aerenchyma in some wetland plants. Ann. Bot. 51:237-249.

Tanaka Y, Makishima T, Sasabe M, Ichinose Y, Shiraishi T, Nishimoto T, Yamada T. 1997. *dad-1*, a putative programmed cell death suppressor gene in rice. Plant Cell Physiol. 38:379-383.

Umeda M, Bhalerao RP, Schell J, Uchimiya H, Koncz C. 1998. A distinct cyclin-dependent kinase-activating kinase of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 95:5021-5026.

Umeda M, Iwamoto N, Umeda-Hara C, Yamaguchi M, Hashimoto J, Uchimiya H. 1999a. Molecular characterization of mitotic cyclins in rice plants. Mol. Gen. Genet. 262:230-238.

Umeda M, Umeda-Hara C, Yamaguchi M, Hashimoto J, Uchimiya H. 1999b. Differential expression of genes for cyclin-dependent protein kinases in rice plants. Plant Physiol. 119:31-40.

van der Weele CM, Canny MJ, McCully ME. 1996. Water in aerenchyma spaces in roots: a first diffusion path for solutes. Plant Soil 184:131-141.

Webb J, Armstrong W. 1983. The effects of anoxia and carbohydrates on the growth and viability of rice, pea and pumpkin roots. J. Exp. Bot. 34:579-603.

Webb J, Jackson MB. 1986. A transmission and cryoscanning electron microscopy study of formation of aerenchyma (cortical gas-filled space) in adventitious roots of rice (*Oryza sativa*). J. Exp. Bot. 37:832-841.

Yamaguchi M, Umeda M, Uchimiya H. 1998. A rice homolog of Cdk7/MO15 phosphorylates both cyclin-dependent protein kinases and the carboxy-terminal domain of RNA polymerase II. Plant J. 16:613-619.

Yamaguchi M, Umeda M, Fabian T, Sauter M, Bhalerao RP, Schrader J, Sandberg G, Uchimiya H. 2001. Activation of CDK-activating kinase is dependent on interaction with H-type cyclins in plants. Plant J. 24:11-20.

Yuan J, Horvits HR. 1990. The *Caenorhabditis elegans* genes *ced-3* and *ced-4* act cell-autonomously to cause programmed cell death. Dev. Biol. 138:33-41.

Notes

Authors' address: Institute of Molecular and Cellular Biosciences, Yayoi 1-1-1, Bunkyo ku, Tokyo, Japan, E-mail: uchimiya@imcbns.iam.u-tokyo.ac.jp.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Molecular tools for achieving synthetic apomixis in hybrid rice

J. Bennett, X.-Z. Bi, A. Kathiresan, and G.S. Khush

Apomixis is asexual reproduction through seed. It occurs in more than 300 plant species but appears to be absent from the genus *Oryza*. One of the most important applications of apomixis in rice would be to develop true-breeding hybrid genotypes that would allow poor farmers to benefit from hybrid yield advantage. Although considerable progress has been made in mapping loci that control apomixis in other members of the family Poaceae, no gene capable of inducing apomixis has yet been isolated for introduction into rice. This chapter reports on the development of molecular tools to achieve a synthetic apomixis in rice, in which the sexual embryo would be replaced by an asexual embryo induced in the nucellus. We cover two objectives of the research: induction of the nucellar embryo and ablation of the zygotic embryo. We emphasize the tasks of identifying embryo-inducing genes, placing them under the control of nucellus-specific promoters, and assaying their efficacy in rice.

Hybrid rice accounts for about 65% of rice production in China and is increasingly common in Korea, India, the United States, and several other countries (Li and Yuan 1999). The main incentive for planting hybrid rice is the yield advantage over the best inbred lines (Virmani 1994). This yield advantage is now 15–20% for indica-indica and japonica-japonica hybrids, but the recent development of indica-tropical japonica hybrids could further increase yield heterosis (Khush et al 1998).

The main disincentive to planting hybrid rice is the cost of buying expensive hybrid seed each season. Although the cost of hybrid seed is declining in China (Li and Yuan 1999), it remains high in tropical countries. Many rice farmers in developing countries are accustomed to reducing the cost of inbred seed by setting aside a portion of their production each year for sowing in the following season. If this practice were adopted for hybrid rice, farmers would notice a decline in yield advantage and plant uniformity, as the heterozygosity of the F_1 generation declined in the F_2 and later generations. However, if the heterozygosity of hybrid rice could be fixed by

asexual reproduction through apomixis, the disincentive of buying expensive hybrid seed annually would be removed. Apomixis would render hybrid seed production more flexible and less costly, and farmers could replant their seed indefinitely (Koltunow et al 1995, Calzada et al 1996, Savidan 2000). Even poor farmers could enjoy the greater productivity of hybrid rice.

An overview of apomixis

Sexual reproduction in the ovary

Figure 1A summarizes the normal pathway of sexual reproduction in the ovary of cereals. The product of the pathway is a developing seed consisting of a diploid embryo, a triploid endosperm, and a seed coat. The double fertilization event leading to formation of the embryo and the endosperm occurs in the embryo sac after the delivery of two sperm nuclei by a pollen tube. The embryo sac and the pollen are both products of meiosis. If the flower in question belongs to an F_1 hybrid plant derived from divergent inbred parents, recombination between the parental chromosomes will occur during both male and female meiosis, producing an endosperm, an embryo, and hence an F_2 plant that are genetically distinct from the hybrid and the two parents.

Types of apomixis

Apomixis has been found in more than 300 species of plants, including close relatives of cereals such as maize, wheat, *Pennisetum*, and sorghum (Asker and Jerling 1992, Sharma and Thorpe 1995, Peel et al 1997, Huang and Chen 1999). The asexual embryo of apomictic seed is formed without passage through a normal meiosis (Koltunow 1993). This is illustrated in Figure 1A for the form of apomixis known as adventitious embryony; the embryo is genetically identical to maternal tissue. The other major forms of apomixis are known as diplospory and apospory.

In all three forms of apomixis, male sporogenesis and gametogenesis are generally normal, but female sporogenesis and gametogenesis are altered (diplospory) or suffer competition from an asexual embryo (apospory and adventitious embryony). The diplosporous embryo arises from an unreduced cell of the sexual embryo sac, whereas the aposporous embryo arises from an embryo sac formed from a single differentiated nucellar cell adjacent to the sexual embryo sac. By contrast, adventitious embryos arise directly from cells in maternal tissue, most frequently from the nucellus (Fig. 1A).

In the family Poaceae, which includes the cereals, apomixis is predominantly either diplosporous or aposporous, with the endosperm formed by pseudogamy, i.e., the fertilization of the polar nucleus or nuclei by a reduced male gamete. We are attempting to introduce into hybrid rice a synthetic form of apomixis that is unusual in its combination of adventitious embryony in the nucellus with pseudogamous endosperm formation (Fig. 1B). If successful, this approach will generate a seed containing an embryo that has bypassed meiosis completely. The embryo will be genetically identical to the F_1 plant on which it forms; the accompanying endosperm, by contrast, will be formed by the normal sexual pathway and will remain triploid. Be-

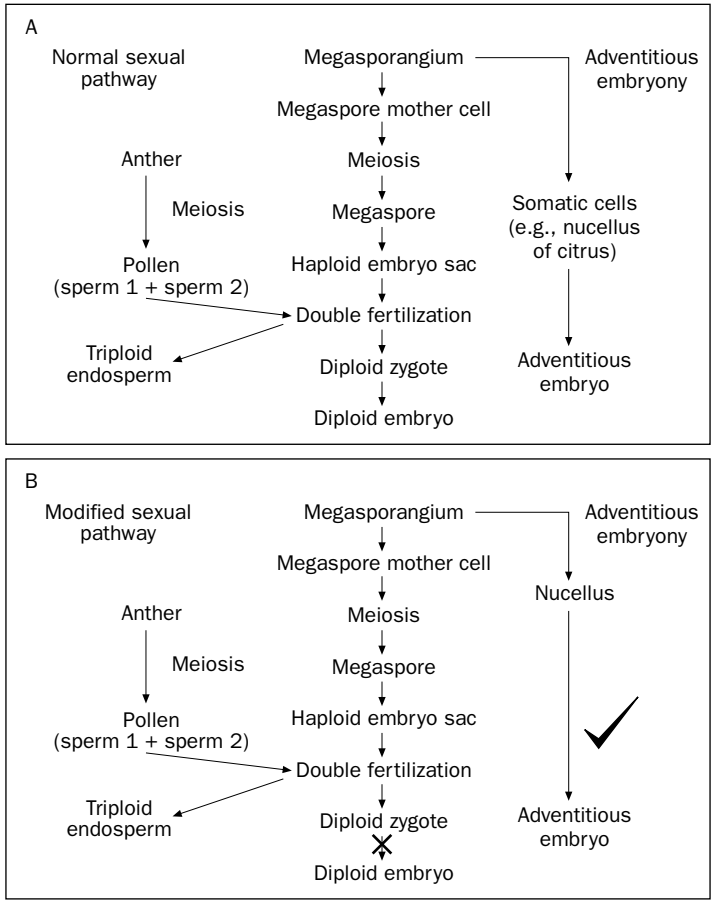


Fig. 1. Outline of three reproductive pathways in the ovary of plants. (A) Pathways of sexual reproduction and adventitious embryony. (B) Proposed pathway of synthetic apomixis in rice, featuring the induction of an adventitious embryo in the nucellus and ablation of the sexual embryo.

cause pollination is required for pseudogamy, the sexual embryo will also be formed but will be ablated to avoid competition with the adventitious embryo. Genes are therefore required for induction of the adventitious embryo, ablation of the sexual embryo, and the correct timing of the two events.

Figure 2 shows where the switch to apomixis fits into the activities of hybrid rice breeders and farmers. Breeders use a male parent (MP) and a female parent (FP) that have been genetically engineered to contain the several transgenes required for synthetic apomixis. These genes are discussed in detail later in "Achieving synthetic apomixis," "Induction of an adventitious embryo in the rice nucellus," and "Ablation of the sexual embryo." The MP and FP are not apomictic because each contains only a subset of the genes needed for apomixis; they are maintained on a small scale in the

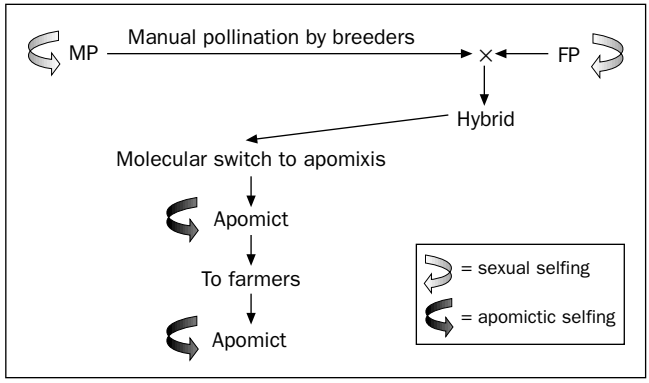


Fig. 2. Outline of the breeding of apomictic hybrid rice and its reproduction by farmers. FP = female parent, MP = male parent.

breeder's field by selfing (Fig. 2). After the breeder pollinates manually, the resulting hybrid contains the full set of the transgenes for switching to apomixis. When the hybrid plant flowers, it produces apomictic seeds. These seeds may be multiplied in the breeder's field to generate adequate amounts of seed for distribution to seed companies and farmers, both of whom may subsequently reproduce the apomictic seed. Costly field-scale hybridization is not required at any stage.

Mapping genes for control of apomixis

Which genes control the naturally occurring forms of apomixis? Little is known about the genetic control of adventitious embryony except that in citrus a single gene controls the switch between polyembryonic and monoembryonic states and is dominant for polyembryony (Iwamasa et al 1967). By contrast, genetic control of apospory and diplospory is well studied, particularly in members of the Poaceae. Apospory and diplosory may both be analyzed as a single trait or as a set of component traits. Apomeiosis, parthenogenesis, and facultativeness are some of the component traits analyzed genetically in diplosporous apomicts (Savidan 2000). Each component trait often appears to be under single-locus control and DNA markers have been linked to some of these loci.

Apospory is controlled by a single locus in *Brachiaria* and in *Pennisetum*. The fact that these loci share common DNA markers (Pessino et al 1997) strongly suggests that apospory has a common molecular basis in these genera. Pessino et al (1998) developed a fine map of the apospory region in *Brachiaria*. They identified markers located 1.2 and 5.7 cM on either side of the target locus. The markers showed synteny to regions of maize chromosome 5 and rice chromosome 2. Parthenogenesis in aposporous Kentucky bluegrass (*Poa pratensis*) and apospory in buffelgrass (*Cenchrus ciliaris* = *Poa ciliare*) are also controlled by single genetic loci linked to DNA markers (Barcaccia et al 1998, Gustine et al 1997).

Tripsacum, a close relative of maize, carries a locus conferring diplospory. Kindiger et al (1996) evaluated several apomictic and sexual maize-*Tripsacum* hybrids for the presence or absence of *Tripsacum* chromosomes. Using cytological and molecular analysis, they associated the control of diplospory with the long arm of chromosome 16 of *Tripsacum*, corresponding to the long arm of chromosome 6 in maize. Blakey et al (2001) exploited a relationship between the expression of apomixis in natural polyploids of *Tripsacum dactyloides* and fertility as measured by percent seed set. They found fertility linked to six DNA markers on *Tripsacum* genetic linkage groups A, D, F, I, H, and L. Linkage group D has a syntenic relationship to the short arm of maize chromosome 5, where markers for apospory in *Brachiaria* and in *Pennisetum* map (Pessino et al 1997). Also included in this marked interval is the gene *Ameiotic1* (*Am1*), whose function is required for the initiation of meiosis in both micro- and megaspore mother cells. The absence of expression of *Am1* in the female is, in all likelihood, a prerequisite for the expression of apomixis (Blakey et al 2001).

Restricted recombination around apomixis loci

DNA marker analysis has shown that recombination around some apomixis loci is restricted (Ozias-Akins et al 1998, Roche et al 1999, Grimanelli et al 1998a). When pearl millet (*Pennisetum glaucum*), which reproduces sexually, was pollinated with an aposporous wild relative (*P. squamulatum*), 12 marker loci segregated with apospory and showed no recombination (Ozias-Akins et al 1998). As recombination is usually restricted in interspecific crosses compared with intraspecific crosses, the restricted recombination seen for the apomixis locus and the 12 markers could be coincidental. Alternatively, restricted recombination may have evolved as a mechanism to allow several genes needed for apomixis to function together as a single genetic unit. When Roche et al (1999) examined the behavior of the same markers in two intraspecific crosses between aposporous and sexual forms of *Cenchrus ciliaris* (= *P. ciliare* [L.]Link), they found that 9 out of the 12 markers were linked to apospory and 8 still showed no recombination. These data indicate that the restriction of recombination is unlikely to be coincidental but they do not yet constitute proof that the region of restricted recombination contains more than one gene for apomixis.

Diplospory in *Tripsacum* involves both the failure of meiosis and the parthenogenetic development of the unreduced gametes. One single dominant allele is believed to be responsible for the whole developmental process. Construction of a linkage map for the chromosome controlling diplosporous apomixis in *T. dactyloides* was carried out in both tetraploid-apomictic and diploid-sexual *Tripsacum* species using maize restriction fragment length polymorphism (RFLP) probes (Grimanelli et al 1998a). A high level of colinearity was observed between the *Tripsacum* chromosome carrying the control of apomixis and a duplicated segment in the maize genome (chromosomes 6L and 8L). In the apomictic tetraploid, a marked restriction of recombination occurred compared with the corresponding genomic segment in sexual *Tripsacum* and maize. This result is again consistent with the idea that apomixis, although inherited as a single Mendelian allele, might really be controlled by a cluster of linked loci, but the idea needs further substantiation.

Search for apomixis in rice

There are many reports of apomixis in rice (Guo 1991, Chen 1992, Tan et al 1994, Ye et al 1994, 1995, Mu et al 1996, Yao et al 1997) but they remain to be validated. The crucial criterion for declaring apomixis is the demonstration, preferably through the use of molecular markers, that progeny are identical in genotype to the apomictic parent. Rigorous progeny analysis has not been performed for the presumed apomictic lines of rice; evidence for apomixis has been based on microscope observations of various abnormalities in the sexual reproductive pathway. Most intensively studied is cultivar 84-15. Chen (1992) gathered together 39 papers describing aspects of the histology, genetics, and agronomy of 84-15, including the report that nucellar embryos compete with the zygotic embryo for dominance. Rigorous progeny analysis is needed. Shi et al (1996) examined 84-15 for the abnormalities reported by Chen (1992). They found megasporogenesis and embryo sac development to be normal in 97.3% of 1,380 ovules. Fertilization and embryo and endosperm development were also normal. They concluded that 84-15 is not an apomict.

All validated, naturally occurring apomictic species of the Poaceae are polyploid (Huang and Chen 1999). For this reason, when IRRI's germplasm collection was examined for evidence of apomixis, polyploid accessions were included along with diploid accessions. The cytological search failed to reveal any abnormal structures suggestive of apomixis in diploid or polyploid wild species of rice (Brar et al 1995).

Synthetic apomixis for hybrid rice

Strategies for achieving apomixis in rice

In the past decade, several laboratories embarked on programs to introduce some form of apomixis into rice. Han et al (1999) attempted to transfer apomixis to rice from *Panicum maximum* by protoplast fusion. Protoplasts were derived from suspension cells of japonica line 02428, inactivated with iodoacetamide, and fused with protoplasts of *P. maximum* that had been treated with soft X-rays. Southern hybridization confirmed the somatic hybrid character of regenerated plants, but an analysis of apomictic character was not reported. Most other laboratories have focused on the transfer of isolated genes through genetic engineering (Peacock 1992, Jefferson 1994). Peacock (1992) proposed a genetic engineering and mutagenesis approach based on the assumption that there is one major gene (a putative "embryo sac induction" gene) with two allelic forms controlling apomixis. The sexual allele of this gene would be switched on in the megaspore after meiosis, but the apomeiosis allele would be expressed early, prior to meiosis, in nucellar cells. Peacock (1992) hypothesized that a mutation equivalent to the apomeiosis allele could be identified if an appropriate high-throughput genetic screen could be devised. He proposed that *Arabidopsis thaliana* would offer major advantages in the search for the relevant gene. Subsequent research on *Arabidopsis* led to the isolation of mutants displaying fertilization-independent seed (fis) production (Peacock et al 1995) and fertilization-independent endosperm (fie) production (Ohad et al 1996) and isolation of the corresponding genes (Luo et al 1999, 2000, Ohad et al 1999, Vinkenoog et al 2000). Howden et al (1998)

used T-DNA-tagging in *Arabidopsis* to isolate eight male and female gametophytic mutants detected by segregation distortion. This approach led to the isolation of MEDEA, a polycomb group gene that is involved in maternal control of embryogenesis (Grossniklaus et al 1998). Maternal-effect phenotypes can result from regulation by genomic imprinting, a phenomenon of critical importance for both sexual and apomictic seed development (Grossniklaus et al 2001).

IRRI adopted a three-fold strategy for inducing apomixis in rice (Khush et al 1994). The three approaches were (1) mutagenesis of diploid germplasm that had been genetically tagged to facilitate detection of putative apomicts, (2) transfer of cloned apomixis genes from other grasses into rice as they became available, and (3) development of molecular tools for achieving synthetic apomixis in hybrid rice. The first approach has not yet yielded mutants with the desired genotype. The second approach has stalled through lack of cloned apomixis genes (see the above section "An overview of apomixis"). The third approach has been pursued collaboratively by IRRI and the Commonwealth Scientific and Industrial Research Organization (CSIRO) of Australia and is the focus of the remainder of this chapter.

Collaboration on apomixis

IRRI's project on molecular tools for achieving synthetic apomixis in hybrid rice is a collaboration with the CSIRO Division of Plant Industry in Australia and it comprises three different approaches (Bennett at al 1998). One approach being pursued by CSIRO is based on the proposal of Peacock (1992) and envisions the induction of autonomous embryo and endosperm development for fis production. The isolation of *FIS* and *FIE* genes from *Arabidopsis* (Luo et al 1999, Ohad et al 1999, Luo et al 2000, Vinkenoog et al 2000) has permitted the search for rice homologues. A second approach at CSIRO is the isolation of genes for apomixis from the genus *Hieracium* (Koltunow et al 1995). An understanding of how such genes function in sexual and apomictic species of *Hieracium* may permit the isolation and appropriate manipulation of rice homologues.

A third approach, pursued by IRRI, envisions a combination of autonomous embryogenesis and normal endosperm formation. Autonomous embryogenesis in the nucellus of hybrid rice would deliver an apomict that remains heterozygous generation after generation. Normal endosperm formation would deliver the triploid tissue (2m:1p) found in all cereals. By contrast, autonomous endosperm formation would produce a diploid tissue (2m:0p) that may have qualitative and quantitative characteristics not acceptable to farmers or consumers. IRRI plans to maintain triploidy in the apomict by not interfering with pollination, but we must therefore address the question of ablating the sexual (zygotic) embryo to avoid competition with the apomictic embryo. For these reasons, IRRI's research focuses on two objectives: induction of a nucellar embryo and ablation of the sexual embryo. However, before discussing these objectives in detail (see the section on "IRRI's general approach to achieving synthetic apomixis"), we wish to address several important issues related to the project as a whole.

Issues related to achieving synthetic apomixis

Is apomixis compatible with diploidy?

Our goal is to introduce apomixis into diploid hybrid rice, and yet the overwhelming majority of apomicts are polyploid. Is our project therefore doomed to failure because apomixis requires polyploidy? Or are apomixis and polyploidy associated for other reasons: perhaps both mechanisms help plants to survive and reproduce in adverse environments.

Two reports on maize-*Tripsacum* hybrids suggest that apomixis may indeed be compatible with diploidy (Leblanc et al 1996, Grimanelli et al 1998b). In *Tripsacum*, all polyploids reproduce through diplospory and all diploids are sexual. In an effort to transfer apomixis from diplosporous tetraploid *Tripsacum* (2n = 4x = 72) into maize (2n = 20) through conventional backcrosses, Leblanc et al (1996) produced polyhaploid plants combining one complete set of chromosomes from each genus. These polyhaploid plants were totally male sterile but viable seeds were produced apomictically when they were pollinated using maize. Apomictic reproduction in such polyhaploids, which show a diploid-like chromosomal complement, suggested that diplosporous apomixis and polyploidy are not totally linked and that apomixis might be compatible with diploidy.

Grimanelli et al (1998b) used RFLP markers linked to diplospory to analyze various generations of maize-*Tripsacum* hybrids and backcross derivatives and to derive a model for the inheritance of diplosporous reproduction. The results suggested that the gene or genes controlling apomixis in *Tripsacum* are linked with a segregation distorter-type system promoting the elimination of the apomixis alleles when transmitted through haploid gametes. This model offers an explanation of the relationship between apomixis and polyploidy and suggests that the evolutionary importance of the segregation distortion system is to protect the diploid level from being invaded by apomixis.

Until recently, all of the apomictic biotypes studied so far in the genus *Hieracium* subgenus Pilosella had been described as polyploids. However, Bicknell (1997) identified a diploid apomictic form of *H. aurantiacum* reproducing through apospory. Seed set was low, apparently because of the presence of competing megagametophytes within each ovule. This report not only provides encouragement for efforts to introduce apomixis into diploid rice but confirms the importance of ablating the sexual embryo to prevent competition with the apomictic embryo. Savidan (2000) reviews other reports of diploid or dihaploid apomicts and concludes that, "the absence of apomixis at the diploid level, as generally observed in the wild, is not a problem of expression but rather of transmission constraints."

Should the triploidy of the endosperm be preserved?

One way of preventing competition between the apomictic and sexual embryos is to eliminate the sexual process entirely. Such an approach would, of course, have the important consequence of preventing normal endosperm formation and necessitate the triggering of an autonomous (fertilization-independent) process. In autonomous

aposporous apomicts such as *H. aurantiacum* and *H. piloselloides*, both embryo and endosperm form spontaneously inside an unreduced embryo sac (Koltunow et al 1998). Both fis production (Peacock et al 1995) and fie production (Ohad et al 1996) have been achieved by mutagenesis in *Arabidopsis*. Should we attempt to produce some equivalent form of fie formation in a synthetic apomictic of rice? Or should we retain the sexual process to preserve the triploidy of the endosperm? Savidan (2000) presented an argument in favor of retaining triploidy. We examine both sides of the argument below.

Triploidy results when the maternal binucleate central cell of the embryo sac is fertilized by a paternal sperm nucleus. The genotype of the endosperm is represented as 2m:1p. In maize, any departure from this genotype has deleterious effects on the quality and quantity of the endosperm (Birchler 1993). One explanation for this effect invokes imprinting, the phenomenon in which expression of alleles differs depending on whether they originate from the male or female parent. Imprinting in the endosperm of angiosperms has been shown to explain most failures of interploidy or interspecific crosses in plants (Haig and Westoby 1991, Birchler 1993). Because of imprinting, seeds develop normally only if the 2m:1p dosage is represented in the endosperm.

Not all species appear to be strongly imprinted for the 2m:1p genotype. In *Tripsacum*, all polyploids reproduce through diplospory. Meiotic failure in megasporocytes leads to the development of eight-nucleate unreduced megagametophytes. Microgametophytes remain unaffected. Grimanelli et al (1997) used flow cytometry to determine ploidy levels in the endosperm of both apomictic and sexual *Tripsacum* accessions. In both cases, fertilization appeared to involve only one sperm nucleus. Endosperm of apomictic *Tripsacum* developed normally even though the ratio of genomic contributions deviated from the normal 2m:1p ratio. Ratios of 2:1, 4:1, 4:2, 8:1, and 8:2 were observed, depending on both the ploidy level of the parents and the mode of reproduction. Thus, specific dosage effects are seemingly not required for endosperm development in *Tripsacum*. Grimanelli et al (1997) suggest that evolution of diplosporous apomixis might have been restricted to species with few or no imprinting requirements.

In work on the transfer of apomixis from *Tripsacum* to maize, Sokolov and Khatypova (1999) encountered problems with small seed size and high sterility (associated with ovary death in the early stages of development in intergeneric hybrids). The material used comprised F_1 and backcross hybrids from the crossing of tetraploid maize with *Tripsacum* (2n = 56). Hybrids were obtained with different ratios of complete parental genomes or of complete maize genomes and *Tripsacum* subgenomes or chromosomes. Sokolov and Khatypova (1999) concluded that the availability of *Tripsacum* chromosomes in some way inhibited imprinting expression of the maize parental genome in the hybrids. For total suppression, a critical number of nine *Tripsacum* chromosomes was needed.

Imprinting in *Arabidopsis* is less marked than in maize (Scott et al 1998). This may reflect the fact that *Arabidopsis* is inbreeding while maize is outcrossing. The parental conflict model of imprinting (Haig and Westoby 1991) predicts that the parent investing the most in reproduction should have the major influence on the out-

come and may use imprinting to achieve that influence. As inbreeding involves only a single parent, the absence of parental conflict should abolish the need for imprinting. Any imprinting that remains in inbreeding species may reflect either an evolutionary vestige from an outcrossing ancestor (Scott et al 1998) or an additional role for imprinting.

If rice is strongly imprinted for endosperm production, maintenance of the 2m:1p genotype will be essential and can be achieved by retaining fertilization of the central cell by one sperm nucleus from pollen. If rice is not strongly imprinted for endosperm production, autonomous endosperm formation (2m:0p) may be acceptable. The fact that rice is an inbreeder would suggest that imprinting may be weaker than in maize. However, our focus on hybrid rice and heterosis may restore the importance of imprinting. DNA methylation may be the molecular link between heterosis (Tsaftaris and Polidoros 2000) and imprinting (Luo et al 2000, Grossniklaus et al 2001). Methylation of maternal and paternal alleles controls the expression of zein and tubulin genes in the maize endosperm (Lund et al 1995a,b).

Is apomixis safe nonsex?

According to evolutionary biological theories, a dominant apomixis gene will rapidly become fixed in an outcrossing sexual population (Van Dijk and Van Damme 2000). Therefore, in theory, apomixis transgenes could have unconditional advantages that might result in their uncontrollable spread. A synthetic apomict that displays both autonomous endosperm formation and autonomous embryo formation should be able to dispense with the apparatus of pollination entirely and would not pose a threat to biodiversity. However, a synthetic apomict that retained self-pollination for the sake of triploid endosperm formation might release pollen to the environment and would have to be designed to be safe (Savidan 2000).

Our approach to achieving synthetic apomixis should satisfy this requirement because it includes a mechanism to ablate the sexual embryo (Fig. 1B). This ablation mechanism was devised principally to avoid competition between the adventitious embryo and the sexual embryo but it serves the additional purpose of preventing gene flow by ablating the zygotic embryo in any plant to which pollen of the apomict might spread from a farmer's field.

Our approach is also safe when viewed from breeders' fields (Fig. 2). Breeders will need to grow very small numbers of the male and female parents of the hybrid. The feasibility and safety of our approach lies in the fact that we distribute the transgenes required for synthetic apomixis between the two parents of the hybrid in such a way that neither parent carries the apomixis trait and neither parent ablates the sexual embryo (see Fig. 3). Only when breeders manually cross the two parents is the full complement of transgenes assembled together. Thus, the two parents will not ablate their own embryos and pollen escaping from them will not carry the apomixis trait to neighboring fields. Pollen on apomictic hybrids in breeders' fields will be just as safe as pollen on the apomict grown by farmers: the zygotic embryo of the recipient plant will be ablated and gene flow will be immediately arrested.

Achieving synthetic apomixis

The minimum number of transgenes required for IRRI's approach to apomixis is three (Fig. 3). One transgene is required to induce an adventitious embryo in the nucellus; it will take the form of a chimera of a nucellus-specific promoter and the coding region of an embryo-inducing gene. A second transgene is required to ablate the sexual embryo. It will be a chimera of an embryo-specific promoter and a gene encoding an ablating protein. However, it will also contain a green fluorescent protein (GFP) gene located between the promoter and the ablating gene to block transcription. Two *lox* sites in tandem (arrows in Fig. 3) will flank the GFP gene to allow its removal by *cre* recombinase. GFP will be driven by a suitable promoter and GFP function can be monitored to check on the success of excision. The third transgene consists of the *cre* recombinase controlled by a meiosis-specific promoter. In a related approach, Zuo et al (2001) placed GFP downstream from a pair of *lox* sites to demonstrate *cre* recombinase-mediated gene activation.

The first and second transgenes will be in one parent (male parent in Fig. 3), whereas the third transgene will be located in the other parent (female). Prior to pollination by the breeder, the *cre* recombinase will be expressed during meiosis in the male parent but will have no DNA substrate on which to work. After pollination, the *cre* recombinase will be expressed during meiosis and will immediately remove the GFP gene from the second construct via recombination at the *lox* sites. The ablating gene will be expressed in all subsequent zygotic embryos, thus killing them. It will not be expressed in adventitious nucellar embryos because the lineage leading to each successive adventitious embryo does not pass through meiosis. Figure 4 shows the location of the three transgenes before and after manual hybridization. Because of the action of the *cre* recombinase, the ablating gene becomes activated but only in postmeiotic cell lineages. To avoid ablating the endosperm, it is essential that the embryo-specific promoter is not expressed in the pollen, the central cell, or the endosperm.

From the foregoing outline, it will be evident that our two objectives (induction of the adventitious embryo and ablation of the zygotic embryo) consist of several dis-

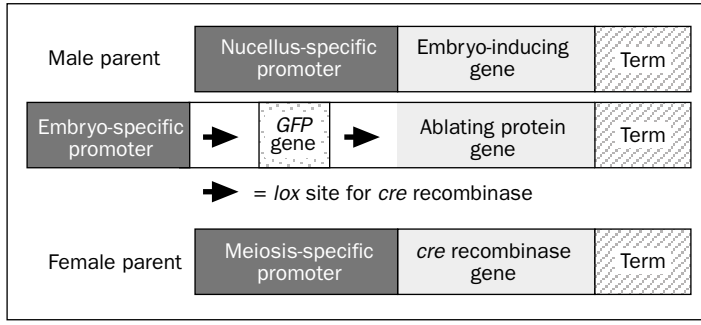


Fig. 3. At a minimum, three constructs will be required to generate the proposed synthetic form of apomixis for hybrid rice. See text for details.

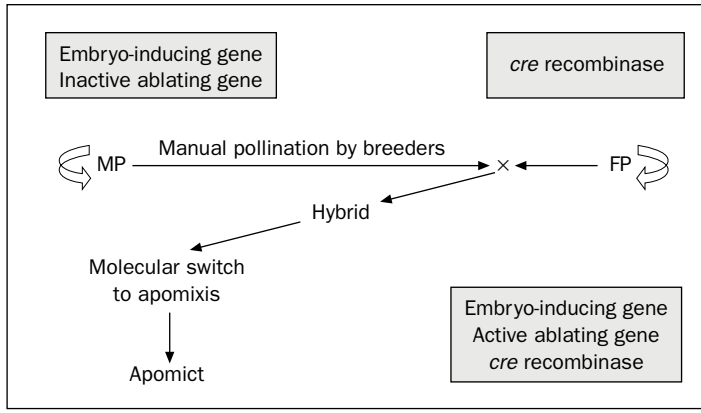


Fig. 4. The allocation of genes between the two parental lines delays ablation of the sexual embryo until after hybridization. FP = female parent, MP = male parent.

tinct tasks. The induction of the adventitious embryo includes three tasks: (1) isolation of genes capable of inducing ectopic embryogenesis, (2) isolation of promoters expressed specifically in the nucellus, and (3) development of assays to detect successful function of the embryo-inducing transgene in the nucellus. The ablation of the zygotic embryo involves another three tasks: (1) isolation of genes capable of ablating the sexual embryo, (2) isolation of promoters expressed specifically in the zygotic embryo but not in other postmeiotic tissues, and (3) isolation of meiosis-specific promoters. In the next section and in "Ablation of the sexual embryo," we report progress to date.

Induction of an adventitious embryo in the rice nucellus

Isolation of nucellus-specific promoters

In addition to being nucellus-specific, the desired promoters should be expressed before the nucellus begins the programmed cell death that supplies nutrients to the developing embryo and endosperm. Several genes are known to be specifically or preferentially expressed in the nucellus. The first such gene to be identified encoded an aspartate proteinase known as nucellin (Chen and Foolad 1997). Transcriptional activity of the nucellin gene in barley begins 1 d before flowering and reaches a peak a few days after flowering. The function of the putative proteinase activity of nucellin is not known but the protein contains a C-terminal tripeptide serine-arginine-leucine (SRL) that may target the enzyme to microbodies. Chen and Foolad (1997) suggest that nucellin is required for the programmed cell death that occurs in the nucellus, thus helping to provide the growing endosperm and embryo with amino acids and peptides.

We isolated a rice homologue of the barley nucellin gene. Southern blotting established that cv. IR64 contains only one copy of the gene. The gene contained seven introns, so that it was possible to design polymerase chain reaction (PCR) primers that discriminated between nucellin mRNA and nucellin genomic DNA (Fig. 5). As judged by reverse transcriptase-PCR (RT-PCR), the gene was not expressed in roots, calli, coleoptiles, anthers, or leaves of IR64, but it was expressed in spikelets from –3 to 6 d after flowering (DAF) (Fig. 5). It was most abundantly expressed 1–4 DAF. Its expression at 2 DAF in a cytoplasmic male-sterile (CMS) A-line and an emasculated CMS B-line was low but detectable, confirming that expression of nucellin in rice, as in barley (Chen and Foolad 1997), is not entirely pollination-dependent. A major difference between barley and rice appeared with *in situ* hybridization of nucellus. Whereas Chen and Foolad (1997) found in barley that nucellin expression was nucellus-specific, in rice the homologue of nucellin is also strongly expressed in embryos. This discovery reduces the likelihood that rice nucellin promoter is suitable for expressing embryo-inducing genes in the nucellus and requires a change of name: we suggest OsAsp1, for rice aspartyl proteinase. OsAsp1 also possesses the SRL motif for targeting to microbodies.

Several other proteinase genes are exclusively or preferentially expressed in the nucellus, including nucellain (Linnestad et al 1998), a barley homologue of the vacuolar-processing protease of the castor bean. In dicots, the vacuolar-processing enzyme is believed to be involved in the processing of vacuolar storage proteins. RNA-blot and *in situ* hybridization analyses detected nucellain transcripts in autolyzing nucellus parenchyma cells, in the nucellar projection, and in the nucellar epidermis. No nucellain transcripts were detected in the highly vacuolate endosperm or in the other maternal tissues of developing grains such as the testa or the pericarp. Using an antibody raised against castor bean vacuolar-processing protease, a single polypeptide was recognized in protein extracts from barley grain. Immunogold-labeling experiments with this antibody localized the nucellain epitope not in the vacuoles but in the cell walls of all nucellar cell types. We propose that nucellain plays a role in processing and/or turnover of cell wall proteins in developing cereal grains.

Xu and Chye (1999) identified a cysteine proteinase of brinjal (*Solanum melongena*) that was highly expressed throughout flower development. *In situ* hybridization studies on flower sections using an antisense RNA probe localized the mRNA to the xylem, the epidermis, and the endothecium of the anther and the nucellar cells, suggesting mRNA involvement in programmed cell death during xylogenesis, anther senescence, and ovule development, respectively. Expression of mRNA during nucellar cell degeneration suggests that protein reserves of the nucellus are released to the developing embryo. Polarity in its pattern of expression in the nucellus of the developing seed further implies a directional flow of these nutrients.

Dominguez and Cejudo (1998), in a study of grain development in wheat, found that genes encoding carboxypeptidase III and a thiol protease were most abundantly expressed in the nucellus. Other nucellar genes preferentially expressed in the nucellus encode a histone H2B (apple, Dong et al 1998), a lipid transfer protein (barley, Chen and Foolad 1999), Nuc1, a protein of unknown function (barley, Doan et al

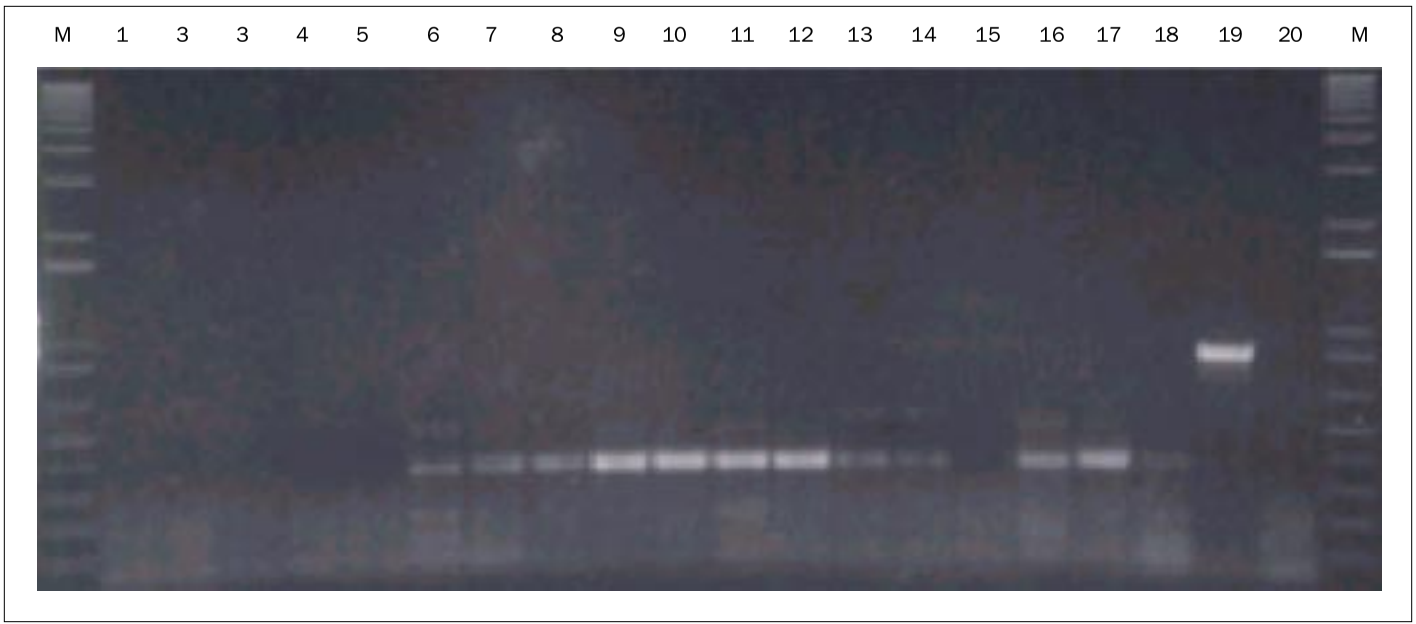


Fig. 5. Reverse transcriptase-PCR shows that a rice homologue of barley nucellin is expressed most strongly 1–4 days after flowering (DAF) in IR64. Sources of RNA for RT-PCR: roots (1), embryogenic calli (2), coleoptile (3), anthers (4), leaves (5), spikelets at –3, –1, 0, 1, 2, 3, 4, 5, 6, and 10 DAF (6–15), emasculated spikelet at 2 DAF (18). IR64 genomic DNA (19) shows a larger PCR product (800 bp) than RT-PCR products (400 bp) because of presence of two introns (total length 400 bp). RT-PCR of RNA from CMS A line V20A (16) and restorer line V20B (17) at 2 DAF.

1996), and a putative hydroxyproline-rich protein of the extensin family (barley, Sturaro et al 1998).

All of the above proteins are expressed abundantly 0–10 d after flowering, when the nucellus is undergoing programmed cell death to supply nutrients and space for the developing grain and embryo. We are more interested in identifying genes that are expressed earlier in nucellar formation, prior to the initiation of programmed cell death. Schiefthaler et al (1999) have identified a gene, *NOZZLE* (*NZZ*), that is involved in pattern formation and early sporogenesis in *Arabidopsis thaliana*. In several *nzz* mutants, the nucellus and the pollen sac fail to form, indicating that *NZZ* plays an early and central role in the development of both types of sporangia. *NZZ* may also have an early function during male and female sporogenesis. It is speculated that *NZZ* is a nuclear protein and possibly a transcription factor. If rice contains a homologue of *NZZ*, its expression would mark the earliest limit for nucellar gene expression. Genes that operate downstream from *NZZ* might be useful candidates for early induction of nucellar embryogenesis.

Another gene that is likely to be expressed in the nucellus is *ASG-1* (apomixis-specific gene), discovered by Chen et al 1999. They sampled different developmental stages of embryo sac formation in obligate sexual and facultatively apomictic genotypes of guinea grass (*Panicum maximum*). From a cDNA library, they isolated a stage-specific cDNA clone of a gene that they designated *ASG-1*. The gene was expressed in flower buds of the apomictic but not of the sexual accession. The gene was expressed only during a phase characterized by the appearance of aposporous initial cells in the nucellus. The appearance of these cells is strictly limited to apomictic genotypes. The gene codes for a 34.2-kDa protein related in sequence to various known proteins, including *RD22*, a seed-specific and drought-induced gene of *Arabidopsis thaliana*, and *Polyg1*, a polygalacturonase 1 beta chain precursor of *Lycopersicon esculentum*. A rice homologue of *RD22* has already been reported by the Rice Genome Project of Japan.

Genes for induction of apomictic embryos

Two mutations (*lec1* and *pkl*) in *Arabidopsis* have implicated the corresponding proteins in the control of ectopic embryogenesis. We have isolated rice homologues of these genes to test their embryogenic potential in rice.

Leafy cotyledon 1 (*lec1*) is an embryo-defective mutation that affects cotyledon identity in *Arabidopsis* (West et al 1994). Mutant cotyledons possess a cellular organization that is intermediate between that of cotyledons and leaves from wild-type plants; for example, they possess trichomes, which are normally restricted to leaves. Several lines of evidence suggest that the control of late embryogenesis is compromised by the mutation. First, although mutant embryos were desiccation-intolerant, they could be rescued before they dried and yielded homozygous recessive plants that produced defective embryos exclusively. Second, although many genes normally expressed during embryonic development were active in the mutant, at least one maturation phase-specific gene (12S storage proteins) was not activated. Third, the shoot apical meristem was activated precociously in mutant embryos. Fourth, in mutant

embryos, several genes characteristic of postgerminative development were expressed at levels typical of wild-type seedlings rather than embryos. West et al (1994) concluded that postgerminative development begins prematurely and that embryonic and postgerminative programs operate simultaneously in mutant embryos. The pleiotropic effects of the mutation indicate that the *LEC1* gene plays a fundamental role in regulating late embryogenesis.

The *LEC1* gene encodes a homologue of the HAP3 subunit of the transcription factor that binds to CCAAT boxes (Lotan et al 1998). *LEC1* RNA accumulates only during seed development in embryo cell types and in endosperm tissue. Ectopic postembryonic expression of the *LEC1* gene in vegetative cells induces the expression of embryo-specific genes and initiates formation of embryo-like structures. The results suggested that *LEC1* is an important regulator of embryo development and activates the transcription of genes required for both embryo morphogenesis and cellular differentiation.

LEC1 is not expressed exclusively in embryos. Tsukaya et al (2000), in a study of heteroblasty in *Arabidopsis thaliana*, found that phenotypes associated with foliage leaf-specific mutations were also induced ectopically in cotyledons in the presence of the *lec1* mutation. This implies that *LEC1* is also expressed at some stage of cotyledon formation.

In *Arabidopsis* plants carrying the mutation designated as *pickle* (*pkl*), the primary root meristem retains characteristics of embryonic tissue (Ogas et al 1997). GA suppressed the expression of this aberrant differentiation state. Analysis of double mutants of *pkl* with the loci *ga* or *gai* (encoding a GA biosynthesis gene and GA signalling gene, respectively) suggested that *PKL* plays a signaling role. Root tissue from plants carrying the *pkl* mutation spontaneously regenerated new embryos and plants without the need for hormones, but root explants from the wild type required the auxin 2,4-D to produce embryogenic calli (Fig. 6). The root formed in the *pkl* mutant expressed *LEC1* and other embryonic genes.

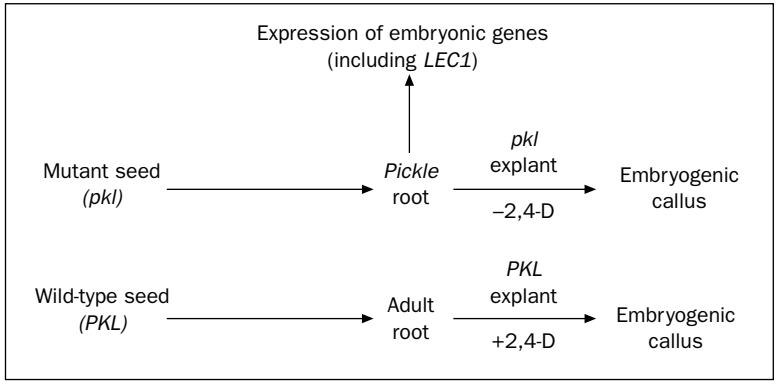


Fig. 6. Regulation of gene expression and embryogenic callus formation in wild-type *Arabidopsis thaliana* and the *pkl* mutant.

The life cycle of angiosperms is punctuated by a dormant phase that separates embryonic and postembryonic development of the sporophyte. In the pickle (*pkl*) mutant of *Arabidopsis*, embryonic traits are expressed after germination (Ogas et al 1999). The penetrance of the *pkl* phenotype is strongly enhanced by inhibitors of gibberellin biosynthesis. Map-based cloning of the *PKL* locus revealed that it encodes a CHD3 protein. CHD3 proteins have been implicated as chromatin-remodeling factors involved in repression of transcription. *PKL* is necessary for repression of *LEC1*. Ogas et al (1999) proposed that *PKL* is a component of a gibberellin-modulated developmental switch that functions during germination to prevent reexpression of the embryonic developmental state.

We have isolated a *LEC1* homologue from rice. It is specifically expressed in response to fertilization and is most abundantly expressed 2–5 DAF (Fig. 7). It is not expressed in panicles before flowering or in leaves, roots, or embryogenic calli. The absence of transcripts of this *LEC1*-like gene in calli is intriguing, because it suggests that the gene is not active in somatic embryos and is therefore involved in events unique to zygotic embryogenesis. *In situ* hybridization shows that the *LEC1*-like gene is highly expressed in zygotic embryos and in endosperm. We have also isolated a second *LEC1* homologue that is expressed in both zygotic and somatic embryos. We shall compare the effectiveness of both *LEC1* homologues to induce ectopic embryogenesis.

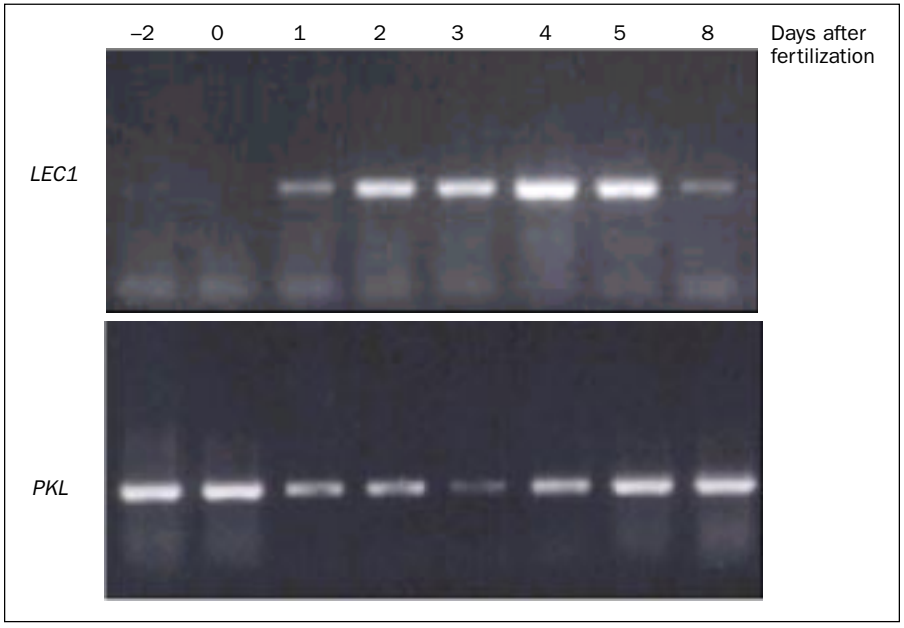


Fig. 7. RT-PCR shows the inverse relationship between transcript levels for *LEC1* and *PKL* genes in rice spikelets after flowering. *LEC1* transcripts are most abundant and *PKL* transcripts are least abundant at about 3 days after fertilization.

Our search for a rice homologue of *PKL* was also successful. This gene is expressed in panicles prior to flowering and 6 DAF, but its abundance declines sharply at about the time when the transcripts of the *LEC1*-like gene appear (Fig. 7). This result is compatible with the control of *LEC1* by *PKL* in rice. It will be important to determine whether *PKL* is expressed in nucellar cells at the time when we hope to induce embryos. If it is so expressed, it may prevent ectopic *LEC1* expression. We are therefore prepared to express the antisense form of *PKL* to allow *LEC1* expression. We shall also check whether the down-regulation of *PKL* in the nucellus is by itself enough to induce transcription of endogenous *LEC1*.

Assessing the first steps in synthetic embryo induction

It is unclear whether ectopic expression of a single gene in the rice nucellus will induce adventitious embryogenesis there. A more likely scenario is that several genes will have to be expressed in a coordinated manner to achieve full embryogenesis. We should therefore plan to judge the effects of individual genes such as *LEC1* and antisense *PKL* by molecular analysis rather than by microscopic examination of nucellar tissue for recognizable embryos. In the case of antisense *PKL*, we would use *in situ* hybridization to assay for the accumulation of *LEC1* transcripts in the nucellus. In the case of *LEC1*, we would use the same technique to assay for expression of the cascade of *LEC1*-dependent embryonic genes of rice. At present, the identities of the *LEC1*-dependent genes are unknown in rice but in *Arabidopsis* they include MADS-domain protein AGL15 (Perry et al 1996), 12S storage proteins and ABI3 (Parcy et al 1997), and six transcription factors (Kirik et al 1998, Kurup et al 2000). However, as several of these proteins are also *FUSCA3*-dependent genes, it might be necessary to express both *LEC1* and *FUS3* in the nucellus to obtain significant expression of the genes dependent on them.

Maize *VIVIPAROUS1* (*VP1*) gene is a functional homologue of *ABI3* (McCarty 1995). *VP1* is a transcriptional activator of the *Em* and *C1* genes of maize. Rice genes acting downstream of *LEC1* may therefore include rice homologues of *VP1*, *Em*, and *C1*. At least one of the abovementioned genes (encoding seed storage protein) is also induced ectopically in *pkl* mutant and *LEC1*-overexpressing plants (Lotan et al 1998).

It is interesting to note that naturally occurring polyembryony in the nucellus of citrus is accompanied by the expression of citrus seed storage proteins (Koltunow et al 1996). The marker used was a gene encoding a 33-kDa subunit of citrin, a salt-soluble, globulin fraction of Valencia seeds. The gene was expressed in polyembryonic seeds when the majority of the embryos were at the early globular stage of the embryo development. Somatic citrus embryos cultured *in vivo* were observed to initiate 33-kDa polypeptide accumulation later in embryo development but they accumulated these peptides at only 10–20% of the level observed in polyembryonic seeds. Therefore, factors within the seed environment must influence the higher quantitative levels of citrin accumulation in nucellar embryos developing *in vivo*, even though nucellar embryos, like somatic embryos, are not derived from fertilization events.

Ablation of the sexual embryo

Isolation of genes capable of ablating the sexual embryo

The gene most commonly used for cellular ablation experiments encodes barnase, a bacterial ribonuclease (Goldman et al 1994). We would prefer to enhance the specificity of ablation by using a gene that is more specific for the embryo than an RNAase. The responsibility for the specificity of ablation would thus be shared between the promoter and the gene. One possibility is to overexpress *PKL* in the zygotic embryo at the time when *PKL* transcripts are declining to allow *LEC1* to be expressed (Fig. 7). Repression of *LEC1* would be expected to block zygotic embryogenesis and result in a gentle and specific form of ablation. Ablation of the sexual embryo requires two promoters: one driving the meiosis-specific expression of *cre* recombinase and the other allowing transcription of the gene for ablation. The progress toward these two goals was described by Kathiresan et al (2001) and therefore is only briefly mentioned here.

Isolation of promoters expressed in the zygotic embryo

Following a report that *REE5*, a partial cDNA clone, is expressed only in the embryo and not in the endosperm (Kikuchi et al 1998), we isolated the full-length *REE5* cDNA and the gene. Our expression studies show that *REE5* is expressed in a wide range of tissues including root tips and various developmental stages of panicle and spikelet. Although we would prefer that the promoter allow expression of the gene for ablation only in the zygotic embryo, it is not desirable for the promoter to be active in other vital vegetative and reproductive tissues. Therefore, the promoter of *REE5* may not be suitable for driving the expression of the gene for ablation. The first *LEC1* homologue that we isolated (Fig. 7) appears to be expressed exclusively in the ovule 1–4 DAF. To judge by RT-PCR, it is not expressed in calli, roots, coleoptiles, leaves, or anthers. However, *in situ* hybridization indicates that it is expressed not only in the embryo but also in some neighboring tissues. These additional sites of expression may eliminate the *LEC1* promoter as a potential promoter for expressing the ablation gene exclusively in the zygotic embryo.

Isolation of meiosis-specific promoters

Several genes of yeast and animal cells are expressed only during meiosis. One of the yeast genes, disrupted meiosis cDNA 1 (*DMC1*), encodes a DNA strand-exchange protein that participates in homologous recombination (Bishop et al 1992). Both yeast and animal *DMC1* are meiosis-specific. An initial report on *DMC1* in *Arabidopsis* indicated that the gene is also expressed in a meiosis-specific manner in plants (Klimyuk and Jones 1997). However, subsequent studies on *Arabidopsis* showed clearly that *DMC1* is expressed not only in meiotic tissue but also in certain mitotic tissues such as root tips (Doutriaux et al 1998). In rice, we have identified two copies of the *DMC1* gene (*DMC1A* and *DMC1B*) by Southern hybridization. Barley, another diploid cereal, contains only one copy of the gene, suggesting that the rice genes diverged after the separation of the ancestors of rice from the ancestors of barley. We isolated and

sequenced the two rice *DMC1* genes and compared their exons, introns, and promoters for conservation of sequence. The exons were highly conserved whereas the introns were divergent. This result establishes that both copies of the gene are functional and under selection for protein structure. Both *DMC1* genes are expressed in panicles at the time of meiosis and in roots and calli. In addition, *DMC1B* is expressed in the pollen during mitosis; it is also expressed in the embryo. Consistent with such differential expression, the promoter regions of the two genes and that of barley *DMC1* show only partial conservation. We are focusing on the promoter of *DMC1A* because of its more limited pattern of expression. We have made promoter-*gus* fusions and promoter mutations to determine whether any of the conserved promoter elements are capable of conferring meiosis-specific expression. We shall also use these constructs to determine whether the promoter of *DMC1A* is expressed in any tissues in the developmental cycle leading from nucellus to nucellus in the planned apomict. If root tips are the only sites where *DMC1A* is expressed mitotically, the promoter should be usable directly to promote correct activation of *cre* recombinase.

Conclusion: integrating apomixis into rice breeding programs

Naturally occurring apomicts are either obligate or facultative. Agriculturally successful apomicts of rice, by contrast, will have to be inducible, with breeders being able to control the switch between the sexual and asexual pathways of reproduction. Apomixis is useful because it fixes the genotype of elite cultivars but sexuality is useful for making further improvements as changing circumstances may require. Even when naturally occurring apomixis genes become available, this sort of control will be desirable. For this reason, it is highly likely that all commercial forms of apomixis will involve some degree of genetic engineering.

If apomixis were available in hybrid rice, breeders could conduct hybridization manually and then multiply the apomictic hybrid seed cheaply in the field. At present, they must conduct inefficient field-scale hybridization. Farmers would be able to reproduce apomictic hybrid seed in their own fields and so would not need to purchase fresh seed every season. The advantages of hybrid rice would thus become available to poor farmers.

References

Asker S, Jerling L. 1992. Apomixis in plants. Boca Raton, Fla. (USA): CRC Press.

Barcaccia G, Mazzucato A, Albertini E, Zethof J, Gerats A, Pezzotti M, Falcinelli M. 1998. Inheritance of parthenogenesis in *Poa pratensis* L.: auxin test and AFLP linkage analyses support monogenic control. Theor. Appl. Genet. 97:74-82.

Bennett J, Ladha JK, Schmit V, Sheehy J. 1998. New Frontier Projects: beyond the pipeline. In: Dowling NG, Greenfield SM, Fischer KS, editors. Sustainability of rice in the global food system. Manila (Philippines): International Rice Research Institute. p 225-243.

Bicknell RA. 1997. Isolation of a diploid, apomictic plant of *Hieracium aurantiacum*. Sexual Plant Reprod. 10:168-172.

Birchler JA. 1993. Dosage analysis of maize endosperm development. Annu. Rev. Genet. 27:191-204.
Bishop DK, Park D, Xu LZ, Kleckner N. 1992. *DMC1*: a meiosis-specific yeast homolog of *E. coli RecA* required for recombination, synaptonemal complex formation and cell cycle progression. Cell 69:439-456.
Blakey CA, Goldman SL, Dewald CL. 2001. Apomixis in *Tripsacum*: comparative mapping of a multigene phenomenon. Genome 44:222-230.
Brar DS, Elloran RM, Lanuang MC, Tolentino VS, Khush GS. 1995. Screening wild species of rice for apomixis. Philipp. J. Crop Sci. 20:6.
Calzada-JV, Crane-CF, Stelly-DM. 1996. Apomixis: the asexual revolution. Science 274:1322-1323.
Chen JS. 1992. A collection of papers on breeding of apomictic rice. Beijing (China): China Science and Technology Press. 118 p.
Chen F, Foolad MR. 1997. Molecular organization of a gene in barley which encodes a protein similar to aspartic protease and its specific expression in nucellar cells during degeneration. Plant Mol. Biol. 35:821-831.
Chen F, Foolad MR. 1999. Nucellar-cell-specific expression of a lipid transfer protein gene in barley (*Hordeum vulgare* L.). Plant Cell Rep. 18:445-450.
Chen LZ, Miyazaki C, Kojima A, Saito A, Adachi T. 1999. Isolation and characterization of a gene expressed during early embryo sac development in apomictic guinea grass (*Panicum maximum*). J. Plant Physiol. 154:55-62.
Doan DNP, Linnestad C, Olsen OA. 1996. Isolation of molecular markers from the barley endosperm coenocyte and the surrounding nucellus cell layers. Plant Mol. Biol. 31:877-886.
Dominguez F, Cejudo FJ. 1998. Germination-related genes encoding proteolytic enzymes are expressed in the nucellus of developing wheat grains. Plant J. 15:569-574.
Dong YH, Kvarnheden A, Yao JL, Sutherland PW, Atkinson RG, Morris BA, Gardner RC. 1998. Identification of pollination-induced genes from the ovary of apple (*Malus domestica*). Sexual Plant Reprod. 11:277-283.
Doutriaux MP, Couteau F, Bergounioux C, White C. 1998. Isolation and characterization of the *RAD51* and *DMC1* homologs from *Arabidopsis thaliana*. Mol. Gen. Genet. 257:283-291.
Goldman MH, Goldberg RB, Mariani C. 1994. Female sterile tobacco plants are produced by stigma-specific cell ablation. EMBO J. 13:2976-2984.
Grimanelli D, Hernandez M, Perotti E, Savidan Y. 1997. Dosage effects in the endosperm of diplosporous apomictic *Tripsacum* (Poaceae). Sexual Plant Reprod. 10:279-282.
Grimanelli D, Leblanc O, Espinosa E, Perotti E, Gonzalez de Leon D, Savidan Y. 1998a. Mapping diplosporous apomixis in tetraploid *Tripsacum*: one gene or several genes? Heredity 80:33-39.
Grimanelli D, Leblanc O, Espinosa E, Perotti E, Gonzalez de Leon D, Savidan Y. 1998b. Non-Mendelian transmission of apomixis in maize-*Tripsacum* hybrids caused by a transmission ratio distortion. Heredity 80:40-47.
Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB. 1998. Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. Science 280:446-450.
Grossniklaus U, Spillane C, Page DR, Kohler C. 2001. Genomic imprinting and seed development: endosperm formation with and without sex. Curr. Opin. Plant Biol. 4:21-27.
Guo XX. 1991. Progress of studies on rice apomixis in China. Sichuan (China): Sichuan Publishing House of Science and Technology.

Gustine DL, Sherwood RT, Huff DR. 1997. Apospory-linked molecular markers in buffelgrass. Crop Sci. 37:947-951.
Haig D, Westoby M. 1991. Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. Philos. Trans. R. Soc. B 333:1-13.
Han X, Teng S, Zhang XQ, Yan QS. 1999. Molecular identification of plants regenerated from somatic hybridization between rice and apomictic *Panicum*. Chin. Rice Res. Newsl. 7:3-4.
Howden R, Park SK, Moore JM, Orme J, Grossniklaus U, Twell D. 1998. Selection of T-DNA-tagged male and female gametophytic mutants by segregation distortion in *Arabidopsis*. Genetics 149:621-631.
Huang QC, Chen QF. 1999. Basic tactics of selecting and breeding apomictic germplasm in rice. J. Fujian Agric. Univ. 28:129-134.
Iwamasa M, Ueno I, Nishiura M. 1967. Inheritance of nucellar embryony in citrus. Bull. Hortic. Res. Stn. Jpn. Ser. B 7:1-8.
Jefferson RA. 1994. Apomixis: a social revolution? Biotechnol. Dev. Monit. 19:14-16.
Khush GS, Brar DS, Bennett J, Virmani SS. 1994. Apomixis for rice improvement. In: Khush GS, editor. Apomixis: exploiting hybrid vigor in rice. Manila (Philippines): International Rice Research Institute. p 1-21.
Khush GS, Aquino RC, Virmani SS, Bharaj TS. 1998. Using tropical japonica germplasm to enhance heterosis in rice. In: Virmani SS, Siddiq EA, Muralidharan K, editors. Advances in hybrid rice technology. Manila (Philippines): International Rice Research Institute. p 59-66.
Kikuchi K, Chung CS, Yoshida KT. 1998. Isolation and analysis of new molecular markers for early embryogenesis in rice (*Oryza sativa* L.). Plant Biotechnol. 15:77-81.
Kindiger B, Bai D, Sokolov V. 1996. Assignment of a gene(s) conferring apomixis in *Tripsacum* to a chromosome arm: cytological and molecular evidence. Genome 39:1133-1141.
Kirik V, Kolle K, Misera S, Baumlein H. 1998. Two novel MYB homologues with changed expression in late embryogenesis-defective *Arabidopsis* mutants. Plant Mol. Biol. 37:819-827.
Klimyuk VI, Jones JDG. 1997. *AtDMC1*, the *Arabidopsis* homologue of the yeast *DMC1* gene: characterization, transposon-induced allelic variation and meiosis-associated expression. Plant J. 11:1-14.
Koltunow AM. 1993. Apomixis: embryo sac and embryos formed without meiosis or fertilization in ovules. Plant Cell 5:1425-1437.
Koltunow AM, Bicknell RA, Chaudhury AM. 1995. Apomixis: molecular strategies for the generation of genetically identical seeds without fertilization. Plant Physiol. 108:1345-1352.
Koltunow AM, Hidaka T, Robinson SP. 1996. Polyembryony in citrus: accumulation of seed storage proteins in seeds in embryos cultured in vitro. Plant Physiol. 110:599-609.
Koltunow AM, Johnson SD, Bicknell RA. 1998. Sexual and apomictic development in *Hieracium*. Sexual Plant Reprod. 11:213-230.
Kurup S, Jones HD, Holdsworth MJ. 2000. Interactions of the developmental regulator ABI3 with proteins identified from developing *Arabidopsis* seeds. Plant J. 21:143-155.
Leblanc O, Grimanelli D, Islam-Faridi N, Berthaud J, Savidan Y. 1996. Reproductive behavior in maize-*Tripsacum* polyhaploid plants: implications for the transfer of apomixis into maize. J. Hered. 87:108-111.

Li JM, Yuan LP. 1999. Hybrid rice: genetics breeding and seed production. Plant Breed. Rev. 17:15-156.

Linnestad C, Doan DNP, Brown RC, Lemmon BE, Meyer DJ, Jung R, Olsen OA. 1998. Nucellain, a barley homolog of the dicot vacuolar-processing protease, is localized in nucellar cell walls. Plant Physiol. 118:1169-1180.

Lotan T, Ohto MA, Yee KM, West MAL, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ. 1998. *Arabidopsis LEAFY COTYLEDON1* is sufficient to induce embryo development in vegetative cells. Cell 93:1195-1205.

Lund G, Ciceri P, Viotti A. 1995a. Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of *Zea mays* L. Plant J. 8:571-581.

Lund G, Messing J, Viotti A. 1995b. Endosperm-specific demethylation and activation of specific alleles of alpha-tubulin genes of *Zea mays* L. Mol. Gen. Genet. 246:716-722.

Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM. 1999. Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 96:296-301.

Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A. 2000. Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. Proc. Natl. Acad. Sci. USA 97:10637-10642.

McCarty DR. 1995. Genetic control and integration of maturation and germination pathways in seed development. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46:71-93.

Mu XJ, Shi GC, Zhu ZQ, Cai X. 1996. Embryological observations on the apomictic rice Ap III (Shuang 13). Acta Bot. Sin. 38:969-971.

Ogas J, Cheng J, Sung ZR, Somerville C. 1997. Cellular differentiation regulated by gibberellin in the *Arabidopsis thaliana* pickle mutant. Science 277:91-94.

Ogas J, Kaufmann S, Henderson J, Somerville C. 1999. *PICKLE* is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 96:13839-13844.

Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL. 1996. A mutation that allows endosperm development without fertilization. Proc. Natl. Acad. Sci. USA 93:5319-5324.

Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL. 1999. Mutations in *FIE*, a *WD* polycomb group gene, allow endosperm development without fertilization. Plant Cell 11:407-416.

Ozias-Akins P, Roche D, Hanna WW. 1998. Tight clustering and hemizygosity of apomixis-linked molecular markers in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes. Proc. Natl. Acad. Sci. USA 95:5127-5132.

Parcy F, Valon C, Kohara A, Misera S, Giraudat J. 1997. The *ABSCISIC ACID-INSENSITIVE3*, *FUSCA3*, and *LEAFY COTYLEDON1* loci act in concert to control multiple aspects of *Arabidopsis* seed development. Plant Cell 9:1265-1277.

Peacock WJ. 1992. Genetic engineering and mutagenesis for apomixis in rice. Apomixis Newsl. 4:3-7.

Peacock WJ, Ming L, Craig S, Dennis E, Chaudhury A. 1995. A mutagenesis programme for apomixis genes in *Arabidopsis*. In: Induced mutations and molecular techniques for crop improvement. Proceedings, Vienna, Austria, 19-23 June 1995. International Atomic Energy Agency, Vienna. p 117-125.

Peel MD, Carman JG, Leblanc O. 1997. Megasporocyte callose in apomictic buffelgrass, Kentucky bluegrass, *Pennisetum squamulatum* Fresen, *Tripsacum* L., and weeping lovegrass. Crop Sci. 37:724-732.

Perry SE, Nichols KW, Fernandez DE. 1996. The MADS domain protein AGL15 localizes to the nucleus during early stages of seed development. Plant Cell 8:1977-1989.

Pessino SC, Ortiz JPA, Leblanc O, Do Valle CB, Evans C, Hayward MD. 1997. Identification of a maize linkage group related to apomixis in *Brachiaria*. Theor. Appl. Genet. 94:439-444.

Pessino SC, Evans C, Ortiz JPA, Armstead I, Do Valle CB, Hayward MD. 1998. A genetic map of the apospory-region in *Brachiaria* hybrids: identification of two markers closely associated with the trait. Hereditas 128:153-158.

Roche D, Cong P, Chen Z, Hanna WW, Gustine DL, Sherwood RT, Ozias-Akins P. 1999. An apospory-specific genomic region is conserved between Buffelgrass (*Cenchrus ciliaris* L.) and *Pennisetum squamulatum* Fresen. Plant J. 19:203-208.

Savidan Y. 2000. Apomixis: genetics and breeding. Plant Breed. Rev. 18:13-86.

Schiefthaler U, Balasubramanian S, Sieber P, Chevalier D, Wisman E, Schneitz K. 1999. Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 96:11664-11669.

Scott RJ, Spielman M, Bailey J, Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development 125:3329-3341.

Sharma KK, Thorpe TA. 1995. Asexual embryogenesis in vascular plants in nature. In: *In vitro* embryogenesis in plants. In: Thorpe TA, editor. Current plant science and biotechnology in agriculture. Vol. 20. Dordrecht (Netherlands): Kluwer Academic Publishers. p 17-72.

Shi GC, Ni PC, Song JX, Mu XJ, Liang TB. 1996. A cytoembryological study on rice 84-15, a rice line presumed to be apomictic. Acta Agron. Sin. 22:657-660.

Sturaro M, Linnestad C, Kleinhofs A, Olsen OA, Doan DNP. 1998. Characterization of a cDNA encoding a putative extensin from developing barley grains (*Hordeum vulgare* L.). J. Exp. Bot. 49:1935-1944.

Sokolov VA, Khatypova IV. 1999. Inactivation of the imprinting effects in maize-*Tripsacum* hybrids. Maize Genet. Coop. Newsl. 73:74-76.

Tan W, Guo XX, Kong FL. 1994. A study of the influence of genetic background on the frequency of twin seedlings in the apomictic rice genotype C1001. J. Southwest Agric. Univ. 16:369-372.

Tsaftaris AS, Polidoros AN. 2000. DNA methylation and plant breeding. Plant Breed. Rev. 18:87-176.

Tsukaya H, Shoda K, Kim GT, Uchimiya H. 2000. Heteroblasty in *Arabidopsis thaliana* (L.) Heynh. Planta 210:536-542.

Van Dijk P, Van Damme J. 2000. Apomixis technology and the paradox of sex. Trends Plant Sci. 5:81-84.

Vinkenoog R, Spielman M, Adams S, Fischer RL, Dickinson HG, Scott RJ. 2000. Hypomethylation promotes autonomous endosperm development and rescues postfertilization lethality in fie mutants. Plant Cell 12:2271-2282.

Virmani SS. 1994. Heterosis and hybrid rice breeding. New York, N.Y. (USA): Springer-Verlag.

West MAL, Matsudaira-Yee K, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ. 1994. Leafy cotyledon1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. Plant Cell 6:1731-1745.

Xu FX, Chye ML. 1999. Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. Plant J. 17:321-327.
Yao JL, Cai DT, Ma PF, Zhu H. 1997. Embryological studies on apospory in rice HDAR. Chin. J. Rice Sci. 11:113-117.
Ye XL, Chen ZL, Li YQ. 1994. Polyembryogenesis in apomixis parent IV of rice. J. Trop. Subtrop. Bot. 2:67-72.
Ye XL, Chen ZL, Li YQ. 1995. The characteristics of apomictic PDER rice variety and the development of diploid embryo sac. J. Trop. Subtrop. Bot. 3:36-42.
Zuo J, Niu QW, Moller SG, Chua NH. 2001. Chemical-regulated, site-specific DNA excision in transgenic plants. Nat. Biotechnol. 19:157-161.

Notes

Authors' address: Plant Breeding, Genetics, and Biochemistry Division, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines.
Acknowledgments: This research was supported by a grant from the Australian Centre for International Agricultural Research. AK thanks the Natural Science and Engineering Research Council of Canada for a fellowship. We thank the Rice Genome Research Project of Japan for cDNA clones and Zhikang Li and his staff for assistance in isolation of bacterial artificial chromosome clones. We are grateful to Darshan Brar, Abed Chaudhury, Liz Dennis, Anna Koltunow, Jim Peacock, and Sant Virmani for advice and to Beng Albano, Evelyn Liwanag, and Nards Estanor for technical assistance. We thank Darshan Brar and Anna Koltunow for comments on the manuscript.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Transformation

Engineering for virus resistance in rice

N.M. Upadhyaya, Z. Li, M-B. Wang, S. Chen, Z-X. Gong, and P.M. Waterhouse

Plant viruses together with their insect vectors cause considerable losses in rice production. Several viruses attack rice plants. Resistance sources are limited or the available resistance genes are either difficult to transfer into cultivated rice or are under threat of being overcome by evolving viral strains. Pathogen-derived resistance (PDR), the expression of pathogen-derived transgenes in plants to interrupt the virus infection cycle, has been employed as an alternative strategy. This approach has given resistance in rice against rice stripe virus (RSV), rice tungro spherical virus (RTSV), rice tungro bacilliform virus (RTBV), rice ragged stunt virus (RRSV), rice hoja blanca virus (RHBV), and rice yellow mottle virus (RYMV). PDR for viruses was originally achieved using gene constructs designed to express wild-type or dysfunctional viral proteins. In many cases, however, these PDR genes appear to be operating at the RNA level associated with posttranscriptional gene silencing. Our latest research shows that viral immunity can be efficiently obtained using constructs that produce dsRNA. This approach has great promise for producing virus resistance in rice. Our experiences and those of other research groups in engineering virus resistance are also discussed.

Among the 23 different viruses known to infect rice, 15 cause significant yield losses (Brunt et al 1998) with most of them vectored either by the leafhopper or planthopper (see Waterhouse and Upadhyaya 1998). Viruses causing yield losses are the rice tungro virus (RTV) complex, rice tungro spherical virus (RTSV) and rice tungro bacilliform virus (RTBV), in India and Southeast Asia, rice stripe virus (RSV) in China and Japan, and rice ragged stunt virus (RRSV), rice grassy stunt virus (RGSV), rice gall dwarf virus (RGDV), rice transitory yellowing virus (RTYV), rice dwarf virus (RDV), and rice black-streaked dwarf virus (RBSDV) throughout Asia. The important rice-infecting virus on the American continent is rice hoja blanca virus (RHBV) and in Africa it is rice yellow mottle virus (RYMV), which is transmitted by beetles (Bakker 1975). The impact of these viruses on rice production has been assessed on national and international scales. For example, RTV, RRSV, and RGSV in Southeast Asia cause

yield losses per annum of about US$700 million, $40 million, and $10 million, respectively (Herdt 1991, Ramasamy and Jatileksono 1996).

Although viral spread can be controlled by regular insecticide application, the most cost-effective, lowest risk, and environmentally sound method of protection for rice from viruses is by the use of virus resistance genes. Protection of rice from viruses can be mediated by natural resistance genes that inhibit virus replication (or spread) in the plant and by genes that repel virus-vectoring insects. Resistance can range from complete immunity to moderate tolerance and resistance sources have been identified in indica, japonica, and wild species of rice (see Waterhouse and Upadhyaya 1998). Sebastian et al (1996) have mapped the chromosomal location of the green leafhopper (GLH) and RTSV resistance gene(s). Progress has also been made in identifying the RYMV resistance locus (Pressoir et al 1998). However, many of the resistance sources found for rice viruses are polygenic and in most cases their mechanism of action is unknown and they are also difficult to transfer to elite varieties. A few of the single-gene resistance sources are under threat of being overcome by evolving virus strains. Since the first demonstration of virus resistance by a viral coat protein transgene from the tobacco mosaic virus (TMV) in 1986 (Powell-Abel et al 1986), there have been many examples of success stories of engineered resistance (virus-derived transgenes) in dicotyledonous plants. Recent advances in rice transformation and molecular characterization of several rice-infecting viruses have greatly facilitated the application of pathogen-derived resistance (PDR) in rice (see Waterhouse and Upadhyaya 1998). Here we describe various strategies of PDR that have been successful in engineering virus resistance in crops other than rice. We also outline the progress made in understanding the underlying mechanisms of PDR as well as the progress, problems, and prospects of genetically engineered virus-resistant rice.

Engineered virus resistance in plants

Recent advances in tissue culture, molecular biology, and transformation technology have facilitated engineering pathogen (originally termed as parasite)-derived resistance with a wide range of potential PDR genes for crop plants. The concept of PDR is that genes from a pathogen's genome can be expressed inappropriately in its host, thereby interrupting the infection cycle (Sanford and Johnston 1985). To use PDR effectively, it is important to understand the life cycle of the target virus at the molecular level. Three genes common to both DNA and RNA viruses are a replicase, a coat protein, and a movement protein gene. The replicase is responsible for initiating replication of the invading virus to produce many new viral RNA molecules. The coat protein packages the newly produced RNA molecules to form virus particles, whereas the movement protein interacts with the viral RNA and the plant cell's plasmodesmata to facilitate the spread of the viral RNA from cell to cell. Interfering with any one of these genes will interrupt the virus life cycle. These three genes have been the most common targets for PDR against plant viruses. According to the success stories, mostly with dicot plants and to a limited extent with monocot plants, PDR for virus can be divided into three categories: (1) the expression in plants of wild-type virus

proteins, (2) the expression of dysfunctional virus proteins, and (3) the expression of virus RNA or virus-associated RNA.

PDR using wild-type virus proteins

It is logical to assume that deliberately crippled viruses can be complemented by transgenic expression of the wild-type gene as shown by Osbourn et al (1990). In most cases, however, virus-derived protein expression has either inhibited or had no effect on virus multiplication. This could be due to the viral protein being expressed inappropriately in terms of the regulated expression of viral genes during a normal infection.

Coat proteins. Coat protein–mediated resistance has been the most widely used form of genetically engineered resistance against plant virus infection (Fitchden and Beachy 1993). It is achieved by incorporating a gene sequence encoding the viral coat protein into a plant gene expression cassette and transferring the cassette into the plant genome. Expression of the virus coat protein in the plant usually gives resistance against not only the virus from which the coat protein was isolated but also against other related viruses. Since the first report of this type of resistance for TMV (Powell-Abel et al 1986), coat protein transgenes have been used to engineer virus resistance for 23 plant virus groups. Although this approach has been widely effective, the mechanisms by which it operates are not well understood. The resistance seems to operate by interfering with virus particle un-coating (Wilson and Watkins 1986), virus replication (Osbourn et al 1989), and virus spread (Wisniewski et al 1990). There are inconsistencies, however, between the attributes of TMV coat protein–mediated resistance and those of other viruses. Many of the examples of "coat protein–mediated resistance" have attributes more consistent with RNA-mediated mechanisms (discussed later).

Replicase. Replicase-mediated resistance was first reported for TMV (Golemboski et al 1990) and was shown to be protein-mediated by Carr and Zaitlin (1991) using a mutated replicase. Replicase-mediated resistance has been shown to be effective with other viruses (see Waterhouse and Upadhyaya 1998). However, in plants carrying a transgene derived from the replicase genes of potato virus X (Mueller et al 1995), cowpea mosaic virus (Sijen et al 1996), and pepper mild mottle virus (Tenllado et al 1995), there is clear evidence that resistance is mediated at an RNA level.

Movement protein. The suggestion that the expression of a heterologous movement protein, which is not adapted to a particular plant species, can interfere with viral spread comes from TMV resistance in transgenic tobacco (*Nicotiana tabacum*) expressing the brome mosaic virus (BMV) movement protein gene (Malyshenko et al 1993).

PDR using dysfunctional virus proteins

A mutated protein may be able to compete with, and disrupt, the function of a wild-type protein. Mutant replicase and movement protein transgenes have been shown to be effective PDR genes.

Mutant replicase. With replicase genes, deletions or substitutions have been made in conserved domains such as glycine-aspartate-aspartate (see Waterhouse and Upadhyaya 1998). In most of these cases, the resistance has been effective only against the homologous virus or its close relatives.

Mutant movement protein. Mutant movement proteins seem to confer broader resistance. Use of a mutated open reading frame (ORF) of one of the so-called "triple-block" genes (ORF2 encoding a ~13 kDa protein) from the potexvirus white clover mosaic virus (WCMV), having the six conserved hydrophilic amino acids converted to hydrophobic amino acids, conferred resistance to WCMV and other potexviruses and a carlavirus (Beck et al 1991) in tobacco. Similarly, broad resistance has been found using mutant movement proteins from TMV (Cooper et al 1995) and potato leafroll virus (Tacke et al 1996). In both cases, the plants produced resistance to viruses outside their respective families.

PDR using RNA

Recent studies suggest that most of the PDR genes operate at the nucleic acid level and not at the protein level. The three different RNA-mediated mechanisms that have been shown to be successful in transgenic plants are (1) expression of an RNA that competes against viral RNA for viral proteins, (2) expression of a ribozyme designed to cleave viral RNA at a specific sequence, and (3) expression of sense and/or antisense viral RNA in plants to induce specific degradation of viral RNA (i.e., posttranscriptional gene silencing or PTGS).

Competitive RNA. Some viruses replicate and package small single-stranded satellite RNAs that are devoid of replicase or coat protein genes. Transgenic plants expressing cloned copies of satellite RNA of tobacco rattle virus (Gerlach et al 1987) or CMV (Harrison et al 1987) showed reduced levels of helper virus replication and amelioration of symptom development when challenged with respective helper viruses. It is thought that this attenuation may be caused by the transgene satellite RNA competing with viral RNA for replicase and/or coat protein. However, certain virus/satellite/host-plant combinations give enhanced, rather than reduced, viral symptoms (Tien and Wu 1991).

Defective interfering (DI) RNAs (which are small RNAs derived from the helper virus) can also ameliorate symptoms caused by their helper virus and probably by the same competitive mechanism as mentioned before. Transgenic plants expressing a DI RNA from cymbidium ringspot virus (CRSV) showed reduced virus symptoms when challenged with the virus (Koller et al 1993).

Ribozymes. The discovery by Haseloff and Gerlach (1988) of a sequence-dependent self-cleavage of satellite RNA (to produce unit-size satellite RNA from multimers) has led to the development of the "ribozyme" concept. They showed that ribozymes can be made with desired specificity to cleave (in *cis* or *trans*) any target RNA molecules at a GUC sequence (cleavage occurs immediately after the C). Transgenic tobacco expressing a ribozyme targeted against potato spindle tuber viroid showed a very high level of resistance to infection by the viroid (Yang et al 1997).

Antisense and cosuppression RNA. Expression of antisense RNA in transgenic plants has been extensively used to inhibit gene expression (see van den Elzen et al 1989, Finnegan and McElroy 1994). In animal systems, antisense has been reported to operate by either preventing the processing or transport of RNA in the nucleus or hybridizing to the target mRNA in the cytoplasm and preventing translation (Green et al 1988). Huntley and Hall (1996) transformed rice with several PDR genes designed to express (1) an artificial DI RNA derived from RNA-2, (2) a sense tRNA-like structure corresponding to the 3′ end of RNA-2, (3) an antisense sequence corresponding to the intercistronic region of BMV RNA-3, and (4) RNA encoding the viral capsid protein. When inoculated with virion RNA, protoplasts (obtained from transgenic plants or callus lines) showed up to 95% reduction in accumulation of progeny viral RNAs. The addition of 20 times the normal level of inoculum was required to overcome the induced resistance. The observed interference appeared to be mediated through viral RNAs rather than protein products.

There is increasing evidence that the antisense RNA in plants is inducing a post-transcriptional gene-silencing mechanism (PTGS), the same mechanism that is induced by cosuppression RNA (Baulcombe 1996). PTGS (see Finnegan and McElroy 1994, Flavell 1994) was first proposed as a mechanism for induced virus resistance in plants by Lindbo and Dougherty (1992) from their work on tobacco transformed with potyvirus coat protein genes. They observed that the plants showing high levels of resistance had low levels of coat protein mRNA in the cytoplasm yet their transgenes were being highly transcribed in the nucleus. Their conclusions were that very high levels of transcription of the transgene (usually associated with high transgene copy number) induce a mechanism in the plant that specifically degrades the mRNA of that transgene and, as the transgene mRNA has the same sequence as part of the viral genome, the virus genome is also degraded by the induced mechanism (see Baulcombe 1996, Stam et al 1997).

Posttranscriptional gene silencing. PTGS is frequently observed in transgenic plants in which active transcription of a transgene results in silencing of its homologous genes or resistance to the virus from which the transgene sequence is derived. With conventional sense and antisense constructs, gene silencing or virus resistance usually occurs in only a small proportion of a transgenic population. Recent studies have shown that these plants often contain transgene integration as an inverted repeat, in which the direction of transcription is toward the center of the repeats (Stam et al 1997, Wang and Waterhouse 2000). Recently, we have discovered that an inverted repeat transgene, designed to produce RNA with internal self-complementarity, was exceptionally effective at inducing PTGS and conferring virus resistance in plants compared with conventional sense and antisense transgenes (Waterhouse et al 1998, 1999, Wang and Waterhouse 2000). We have also found that co-expression of a sense and an antisense strand of a viral sequence in plants can confer complete immunity to the virus (Waterhouse et al 1998). These findings have demonstrated that double-stranded RNA is an important inducer of PTGS in plants and offers a novel strategy for engineering virus resistance in plants.

Antiviral protein–mediated resistance

Virus resistance has also been achieved in transgenic plants by expressing antiviral proteins from other organisms. These include expression of an antiviral protein from pokeweed (Lodge et al 1993), an interferon-associated protein 2′-5′ oligoadenylate synthetase from rat (Truve et al 1993), a human cysteine-protease inhibitor (Garcia et al 1993), a single-chain antibody against a virus coat protein (Tavladoraki et al 1993), and a double-stranded RNA-specific ribonuclease (Watanabe et al 1995).

Progress with engineering virus resistance in rice

Some of the approaches proven to work in dicotyledonous plants using transgenics may also be useful in engineering virus resistance in rice and other monocotyledonous plants. However, several of the rice viruses differ significantly from the type of viruses shown to be responsive to the approaches described previously. Such differences in the nature of rice viruses may make them intractable to some of these approaches, but at the same time these differences might provide opportunities for alternative resistance strategies. Here we describe the progress made in understanding the genome organization of economically important rice viruses, the PDR strategies adopted, and results obtained for virus resistance (Table 1).

Rice tungro viruses

Rice tungro disease is caused by two viruses, RTSV and RTBV, and genomes of both viruses have been well characterized. RTSV is a waikavirus and has a single-stranded (plus sense) linear RNA genome (Fig. 1A) with a life cycle strategy similar to that of the majority of plant viruses (Shen et al 1993). Its isometric virus particles, however, are composed of three different types of coat protein subunits. RTBV can be vectored by insects only in the presence of RTSV. Providing rice with RTSV resistance may prevent tungro epidemics by limiting the spread of the two-virus complex. Researchers at the Scripps Research Institute have recently introduced three coat protein genes (*cp1*, *cp2*, and *cp3*), each under the control of the maize ubiquitin (*Ubi1*) promoter, individually into rice using the particle bombardment method of transformation (Sivamani et al 1999). The resulting transgenic plants were tested for RTSV resistance as primary transformants and as T_1 and T_2 progeny. Some of the plants containing the *cp1*, *cp2*, or *cp3* genes showed moderate levels of resistance. Interestingly, the plants showing this resistance had detectable levels of mRNA for the transgenes but undetectable levels of the coat proteins (Sivamani et al 1999). Recent results using the RTSV replicase transgene seem to be delivering a high level of resistance (Huett, personal communication).

RTBV is a *Badnavirus* and has a circular double-stranded DNA genome (Qu et al 1991, Fig. 1B). A large part of its replication cycle occurs in the nucleus with its genes expressed from a single RNA transcript containing four ORFs. The functions of the proteins encoded by ORFs 1, 2, and 4 remain unknown. ORF3 is by far the largest of the four and is predicted to encode a polyprotein containing the coat protein, viral proteinase, reverse transcriptase, and ribonuclease H. RTBV causes the majority of

Table 1. Transgenic rice with pathogen-derived resistance genes for virus resistance.

Virus[a]	Cultivar	Transgene	Promoter	Transformation method	Resistance level	Laboratory/reference[b]
RTSV	Taipei 309, TN1	*cp1, cp2, cp3*	*Ubi1(I)*	Biolistics	Moderate, delayed symptoms	Sivamani et al (1999); ILTAB/MARDI
RTSV	Taipei 309	*rep*, truncated *rep* sense or antisense	?	Biolistics	High to immune	ILTAB/MARDI
RTBV	Taipei 309	ORFs 1, 2, 3, and 4 sense ORF4 antisense, wild-type, and mutants	CaMV 35S(I) RTBV(I)	Biolistics	Low or none	ITH Jurich /IRRI
RRSV	Taipei 309, C. Boro II	*S5, S7, S8, S9, S10* sense and antisense	Various	Biolistics/*Agrobacterium*	Low to moderate, few high	Upadhyaya et al (1998b); CSIRO Plant Industry/ SIBS/ZAAS
RRSV	Yokara	*S9*	CaMV 35S	Electroporation	None	Matsumura and Tabayashi (1995)
RDV	Yokara	*S8*	CaMV 35S	Electroporation	None	Matsumura and Tabayashi (1995)
RDV	Zhonghua 8, Zhonghua 10	*S8*	CaMV 35S	Biolistics	na[e]	Zheng et al (1997)
RSV	Kinuhikari/Nipponbare	*N*	CaMV 35S(I)	Electroporation	Moderate to high	Hayakawa et al (1992)
RSV	Annong S-1 (indica)	*N*	CaMV 35S	Biolistics	Moderate to high	Yan et al (1997)
RHBV	Circa-8	*N*	CaMV 35S	Biolistics	High	CIAT
RHBV	2 Costa Rican indica cvs.	*cp, cp* (MAR)[c]	*Act1*	Biolistics/ *Agrobacterium*	na	CIBCM/Cornell University
RYMV	Taipei 309	*cp*	*Ubi1(I)*	Biolistics	Varied levels	Kouassi et al (1997)
RYMV	Bouaké 189, ITA 212, BG9-2	ORF2 (RDRP)[d]	CaMV 35S	Biolistics	Medium to high	Pinto et al (1999)

[a]RTSV = rice tungro spherical virus, RRSV = rice ragged stunt virus, RDV = rice dwarf virus, RSV = rice stripe virus, RHBV = rice hoja blanca virus, RYMV = rice yellow mottle virus. [b]ILTAB = International Laboratory for Tropical Agricultural Biotechnology, MARDI = Malaysian Agricultural Research and Development Institute, CSIRO = Commonwealth Scientific and Industrial Research Organization, SIB = Shanghai Institute for Biological Sciences, ZAAS = Zhejiang Academy of Agricultural Sciences, CIAT = Centro Internacional de Agricultura Tropical, CIBCM = Centro Internacional de Biología Celular y Molecular. [c]MAR = matrix attachment region. [d]RDRP = RNA-dependent RNA polymerase. [e]na = not available.

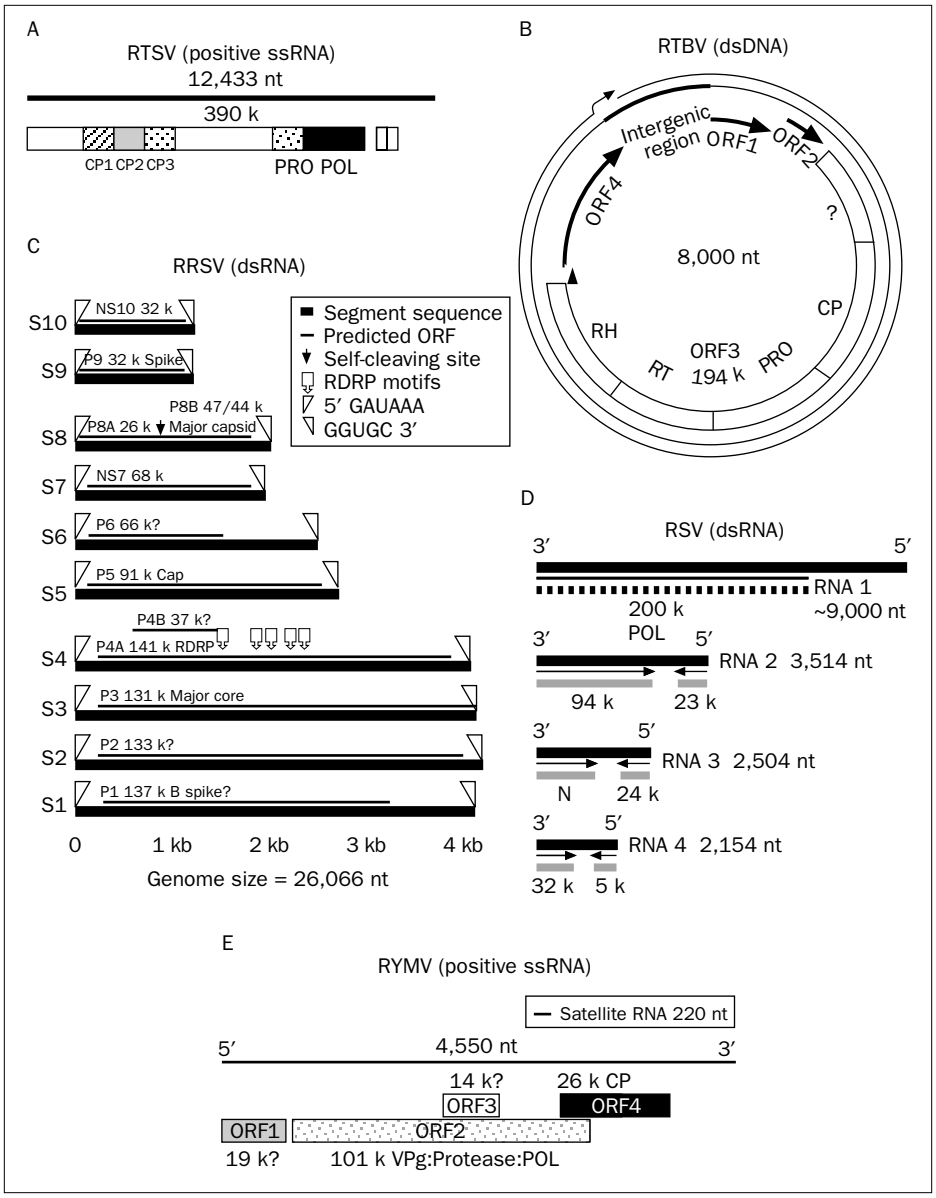


Fig. 1. Genome organization of rice-infecting viruses. (A) RTSV has a single-stranded RNA genome that is translated into one polyprotein and then cleaved by the viral protease (PRO) to produce CP1, CP2, and CP3 and an RNA-dependent RNA polymerase (POL). (B) RTBV has a double-stranded DNA genome that is transcribed to give an mRNA containing four open reading frames (ORFs). A polyprotein encoded by ORF3 contains coat protein (CP), protease (PRO), reverse transcriptase (RT), and RNase H domains. (C) RRSV contains 10 dsRNA segments. Each contains an ORF, except S4, which contains two ORFs, of which one encodes an RNA-dependent RNA polymerase (RDRP). At least six of them encode structural proteins to make up the complex virus particle and two encode for nonstructural (NS) proteins. (D) RSV genome is either dsRNA or negative-sense single-stranded RNA that is made up of four segments. RNA1 encodes RDRP (POL) and the other RNAs contain ORFs in both sense and negative-sense strands. (E) RYMV contains a single-stranded negative-sense RNA with four ORFs, with ORF2 encoding a polyprotein comprising a viral protease, the helicase, and the RDRP. ORF4 encodes the CP. ORF1 is required for viral replication and the function of ORF3 is yet to be determined.

the disease symptoms. Efforts from different groups to engineer resistance using RTBV PDR genes have not been successful to date (J. Fütterer, O. Azzam, and I. Potrykus, personal communication). This has been the case in dicots with other dsDNA viruses (see Waterhouse and Upadhyaya 1998).

Reoviruses

Seven different reoviruses infect rice (see Waterhouse and Upadhyaya 1998). Reoviruses are unusual plant viruses as they replicate in their insect vectors, have multisegmented genomes (10–12 segments) of double-stranded RNA, carry replicase molecules within their particles, and have very complex double-shelled isometric virus particles. The organization of the genes in reoviruses is relatively simple, with nearly all genome segments encoding for a single protein (Fig. 1C). From work on animal-infecting reoviruses, it is possible to present a scenario for the plant reovirus replication cycle (see Waterhouse and Upadhyaya 1998). Almost all steps of the life cycle occur within the particle. Upon infection of the plant cell, the outer shell is removed and the replicase within the inner particle produces mRNA transcripts (from some of the dsRNA genome segments), which exit the particle via channels. The proteins translated from these "early" transcripts form new core particles and specifically package one copy of each of all the "late" RNA transcripts (all 10 to 12 segments) and then produce a complementary RNA strand within the core particle. These particles are then either coated with the second outer shell and become inactive or remain single-shelled and repeat the replication process. An important feature of reoviruses is the need to package one copy of each of the genome segments per particle. It seems that the sequences at the termini of each segment form "panhandle" structures that allow their specific recognition during the core packaging process.

We have sequenced the entire genome (10 segments) of RRSV and identified four genes encoding structural proteins (S3, S5, S8, and S9), one polymerase gene (S4), and two genes encoding nonstructural proteins (S7 and S10) (Upadhyaya et al 1995, 1998a, unpublished). One other plant-infecting reovirus, RDV, has been sequenced completely (see Omura 1995) with the functions for genes encoded by 12 genomic segments tentatively assigned.

We have used the PDR approach for engineering RRSV resistance in rice. This was achieved by transforming japonica (cv. Taipei 309) and indica (cv. Chinsurah Boro II) rice with plant expression vectors carrying RRSV genes (in both sense and antisense orientations) encoded by genome segments S5, S7, S8, S9, and S10. Plants containing sense transgenes under the control of either the maize *Ubi1* promoter or rice *Actin1* (*Act1*) promoter produced detectable levels of the transgene protein. Preliminary glasshouse tests conducted in China indicate that resistance to RRSV has been achieved with transgenes derived from genome segments S5, S7, S8, S9, and S10. In some cases, the effective transgenes were in the sense orientation, whereas in other cases they were in the antisense orientation (Table 2). Resistance varied from complete immunity to delayed symptom development (Upadhyaya et al 1996, 1998b, unpublished).

Table 2. Summary of screening for rice ragged stunt virus (RRSV) resistance in transgenic rice containing RRSV pathogen-derived resistance (PDR) genes.

PDR gene	Promoter	Sense/ antisense	No. of lines tested	Resistance[a]		
				Low	Medium	High
RRSV *S5*	*rolC*	Sense	4	0	3	1
		Antisense	4	0	1	1*
RRSV *S7*	*rolC*	Sense	5	1	0	0
		Antisense	5	0	0	0
	Ubi1(I)	Sense	1	0	0	0
		Antisense	3	2	0	0
	Act1	Sense	1	0	0	0
		Antisense	2	0	1	0
	RTBV	Sense	5	1	2	0
		Antisense	3	1	1	0
	CaMV 35S	Sense	30	11	2	0
		Antisense	20	2	1	1
RRSV *S8*	*Ubi1(I)*	Antisense	1	0	1	0
RRSV *S9*	*rolC*	Sense	5	2	1	1
		Antisense	12	4	6	2*
	Ubi1(I)	Sense	4	1	1	1
		Antisense	4	0	0	1
RRSV *S9DI*	*rolC*	Sense	5	2	0	0
		Antisense	5	2	0	0
RRSV *S9*	*Act1*	Sense	23	5	0	4
		Antisense	17	7	3	2
RRSV *S9M*	*Act1*	Sense	4	0	2	0
RRSV *S9D*	*Act1*	Sense	25	4	3	7
RRSV *S10*	CaMV 35S	Antisense	1	0	1	0

[a]Resistance reconfirmed in T_2/T_3 generation. All transgenic lines were of cv. Taipei 309 except the lines containing RRSV *S5*, which were of cv. Chinsurah Boro II background. All were produced by biolistics except CaMV 35S-RRSV *S7* transgenic lines, which were produced by *Agrobacterium*-mediated transformation. The 10–14-d-old T_1 progeny seedlings were inoculated with viruliferous insects (*Nilaparvata lugens*) for 2 d and seedlings were then disinfected. Observations on symptom development and/or reverse transcriptase-polymerase chain reaction (RT-PCR) analysis were performed after 25–40 d. Percentage infection (RT-PCR and/or symptom positive) in nontransgenic control plants was taken as a benchmark to score resistance in transgenic plants as low (50–75% of that of control plants), medium (25–50%), and high (0–25%).

Transgenes derived from the *RDV S8* or *RRSV S9* genes have been placed under the control of the CaMV 35S promoter and transferred into rice, but no resistance was observed (Matsumura and Tabayashi 1995).

Tenuiviruses

Rice stripe virus (RSV), rice grassy stunt virus (RGSV), and rice hoja blanca virus (RHBV) are tenuiviruses with some unusual properties. They are similar to reoviruses in having the dsRNA genome, genome segments with panhandle structures, replicase associated with virus particles, and the capacity to replicate in their insect vectors as well as in their plant hosts. However, they differ from reoviruses in that they form filamentous virus particles, package both single (negative)-stranded and double-stranded forms of their four genome segments, and express their genes using

an ambisense coding strategy (Fig. 1D). Genome segment S1 encodes the replicase. Segments S2, S3, and S4 use ambisense expression to produce six proteins including a nucleoprotein (N protein), which is associated with the virus particle and sometimes referred to as the coat protein.

Rice stripe virus. Hayakawa et al (1992) first mentioned virus resistance in rice via PDR for RSV. Two japonica rice varieties (Kinuhikari and Nipponbare) were transformed with the N protein (referred to by the authors as the coat protein) gene of RSV under the control of the CaMV 35S promoter containing the castor bean catalase intron in the 5′ untranslated region of the transcript. Three out of five primary transgenic lines screened for RSV resistance were symptomless, which correlated with the transgene protein presence, and two lines showed segregation of transgene proteins and symptomless plants. Similar results were obtained when using the coat protein gene of a Chinese isolate of RSV (Yan et al 1997). This demonstration of N protein–mediated resistance for RSV suggests that it will be a successful approach for resistance to other tenuiviruses such as RGSV and RHBV. Because this mechanism seems to be acting before viral replication commences, the unusual nature of the tenuivirus replication strategy is perhaps irrelevant for this mechanism of resistance.

Rice hoja blanca virus. Encouraged by the *N* protein–mediated resistance in rice against RSV (Hayakawa et al 1992), a research group at the Centro Internacional de Agricultura Tropical (L. Calvert et al, personal communication) has investigated virus protection by expression of the RHBV *N* gene in rice. They have transformed the Latin American indica rice variety Circa-8, using the biolistic method of transformation, with a construct encoding the *N* gene under the control of the CaMV 35S promoter. Plant lines obtained show an extreme resistance to RHBV, Mendelian inheritance of this trait, and unaltered agronomic performance characteristics (in glasshouse tests) in the absence of the virus. The best lines hold great promise for breeding with rice containing conventional RHBV resistance to produce a superior rice variety for use in Latin America. The CIAT team has found that some of its resistant lines contain *N* genes with partial deletions. The most resistant lines had no detectable levels of N protein and very low levels of mRNA. This suggests that, in contrast to the RSV results, the mechanism is operating at the RNA level, is probably posttranscriptional degradation, and can operate against viruses using ambisense gene expression.

Sobemovirus

Rice yellow mottle virus. The disease caused by rice yellow mottle virus (RYMV) is a serious problem for African rice growers using large-scale irrigated systems. RYMV contains a 4,550-nt-long nonpolyadenylated single-stranded positive-sense RNA with four ORFs (Yassi et al 1994, Fig. 1E). The ORF1 product (P1) is required for viral replication (Bonneau et al 1998). The ORF2 product is a polyprotein comprising viral protease, the helicase, and the RNA-dependent RNA polymerase. ORF4 encodes the coat protein and the function of ORF3 is yet to be determined.

Kouassi and others (1997) have produced transgenic Taipei 309 rice expressing the RYMV-*cp* gene under the control of the *Ubi1* promoter and demonstrated transgene product accumulation and resistance to RYMV. In the majority of cases, the type of

resistance obtained was delayed symptom development but not complete immunity. By integrating a transgene encoding the RNA-dependent RNA polymerase of RYMV, Pinto and others (1999) have produced rice conferring resistance to RYMV strains from different African locations. Study of the most resistant line, using transcription analysis, indicated that the resistance is driven from an RNA-based mechanism associated with posttranscriptional gene silencing.

Recently, a self-cleaving circular satellite RNA has been found to be associated with RYMV (Collins et al 1998). It may be possible to make transgenes derived from this satellite RNA that will confer RYMV resistance.

Problems and prospects

At least two of the PDR approaches have proven to be successful in rice: the *cp*-mediated resistance to RTSV, RRSV, RSV, RHBV, and RYMV and replicase-mediated resistance to RYMV. To date, efforts by different research groups to engineer resistance to RTBV have not been successful. Although controlling RTSV may prevent the disease complex from spreading within the rice crop, resistance to RTSV will not protect rice from RTBV coming from other sources. Nevertheless, we believe that a cosuppression or PTGS approach may be successful, as the RTBV RNA transcripts may be vulnerable early in the infection cycle.

It seems very likely that rice with resistance to RGSV could be readily achieved using the *N* gene as this approach has worked well for the other tenuiviruses. The fact that introducing one or more of the RRSV PDR genes has yielded RRSV resistance is encouraging because it opens the door for delivering resistance for the other six rice-infecting reoviruses. Reoviruses spend most of their life within a particle core and it is therefore possible that they are less vulnerable to straightforward PDR strategies. Several strategies, special to reoviruses, could be pursued, such as overexpressing the protein that covers the channels of the core particle through which viral RNA is exported. Another approach might be to produce RNAs that contain prematurely terminated reading frames but have the "panhandle" sequences that label them as specific segments. Such molecules should compete with viral RNAs for packaging into new core particles but not produce wild-type protein in the next round of translation/replication.

An advantage of using a PDR transgene is that it can be followed in breeding programs using defined markers, whereas natural resistance genes are often polygenic and usually have, at best, only linked markers. However, polygenic resistance is likely to be more durable in the field. We will probably see, in the future, the use of combinations of natural virus and insect resistance genes with virus and insect resistance transgenes. In cases where there is no effective or easily transferred natural rice virus resistance gene, a synthetic resistance gene might be made based on the PDR approach and then introduced into a rice background containing natural virus resistance to other viruses.

The use of virus resistance transgenes poses concerns over issues other than their effectiveness. In the late 1980s, when transgenic plants containing viral genes were

first produced and shown to have virus resistance, little concern was raised over the possible environmental impact of these plants. However, many discussions have occurred since then (see Tepfer and Balazs 1997, McLean and Evans 1997) about the use of such transgenic plants on a larger scale. Some scientists believe that the benefits of transgenic virus-resistant plants greatly outweigh the minimal risks (Falk and Breuning 1994, Miller et al 1997), whereas others encourage a cautious approach (Gibbs et al 1997). The potential risks posed by plants expressing rice virus genes are similar to those expressing other plant virus genes and can be categorized into four types: (1) interaction of transgene proteins with an invading virus, thus causing a new disease, (2) selection by transgene-mediated resistance for the creation of a new resistance-breaking virus strain, (3) escape of a transgene into weed species, conferring on them greater fitness, and (4) recombination between a transgene mRNA and the genome of an invading virus to create a novel virus and a new disease. Although these risks are minimal and yet to be demonstrated, each needs to be addressed. Indeed, it may be necessary to modify some of the PDR approaches in order to ensure complete safety of such transgenic crops and satisfy governments, regulatory authorities, and the general public. We believe that careful and responsible uses of PDR transgenes in rice will minimize yield losses from viruses and that within a decade transgenic virus-resistant rice plants will have a widespread use in integrated disease management.

References

Bakker W. 1975. Rice yellow mottle virus. Descriptions of Plant Viruses No. 149. Wallingford (UK): Commonwealth Mycological Institute.

Baulcombe DC. 1996. Mechanisms of pathogen-derived resistance to viruses in transgenic plants. Plant Cell 8:1833-1844.

Beck DL, Guilford PJ, Voot DM, Andersen MT, Forster RLS. 1991. Triple gene block proteins of white clover mosaic potexvirus are required for transport. Virology 183:695-702.

Bonneau C, Brugidou C, Chen L, Beachy RN, Fauquet C. 1998. Expression of the rice yellow mottle virus P1 protein in vitro and in vivo and its involvement in virus spread. Virology 244:79-86.

Brunt AA, Crabtree K, Dallwitz MJ, Gibbs AJ, Watson L, Zurcher EJ. 1998. Plant viruses online: descriptions and lists from the VIDE database. URL http://biology.anu.edu.au/Groups/MES/vide/.

Carr JP, Zaitlin M. 1991. Resistance in transgenic tobacco plants expressing a nonstructural gene sequence of tobacco mosaic virus is a consequence of markedly reduced virus replication. Mol. Plant-Microbe Interact. 4:579-585.

Collins RF, Gellatly DL, Sehgal OP, Abouhaidar MG. 1998. Self-cleaving circular RNA associated with rice yellow mottle virus is the smallest viroid-like RNA. Virology 241:269-275.

Cooper B, Lapidot M, Heick JA, Dodds JA, Beachy RN. 1995. A defective movement protein of TMV in transgenic plants confers resistance to multiple viruses whereas the functional analogue increases susceptibility. Virology 206:307-313.

Falk BW, Bruening G. 1994. Will transgenic crops generate new viruses and new diseases? Science 263:1395-1396.

Finnegan J, McElroy D. 1994. Transgene inactivation: plants fight back! Bio/Technology 12:883-888.
Fitchden JH, Beachy RN. 1993. Genetically engineered protection against viruses in transgenic plants. Annu. Rev. Microbiol. 47:739-763.
Flavell RB. 1994. Inactivation of gene expression in plants as a consequence of specific sequence duplications. Proc. Natl. Acad. Sci. USA 91:3490-3496.
Garcia JA, Cervera MT, Riechmann JL, Lopez-Otin C. 1993. Inhibitory effects of human cystatin-C on plum pox potyvirus proteases. Plant Mol. Biol. 22:697-701.
Gerlach WL, Llewellyn D, Haseloff J. 1987. Construction of a plant disease resistance gene from the satellite RNA of tobacco ringspot virus. Nature 328:802-805.
Gibbs MJ, Waterhouse PM, Weiller GF. 1997. Analysis of natural viral recombination may assist the design of new virus resistance traits. In: McLean GD, Waterhouse PM, Evans G, Gibbs MJ, editors. Commercialization of transgenic crops: risk, benefit and trade considerations. Proceedings of the workshop, 11-13 March 1997, Canberra, Australia. Canberra (Australia): Australian Government Publishing. p 159-172.
Golemboski DB, Lomonossoff GP, Zaitlin M. 1990. Plants transformed with a tobacco mosaic virus nonstructural gene sequence are resistant to the virus. Proc. Natl. Acad. Sci. USA 87:6311-6315.
Green PJ, Pines O, Inouye M. 1988. The role of antisense in gene regulation. Annu. Rev. Biochem. 55:569-597.
Harrison BD, Mayo MA, Baulcombe DC. 1987. Virus resistance in transgenic plants that express cucumber mosaic virus satellite RNA. Nature 328:799-801.
Haseloff J, Gerlach WL. 1988. Simple RNA enzymes with new and highly specific endoribonuclease activities. Nature 334:585-591.
Hayakawa T, Zhu Y, Itoh K, Kimura Y, Izawa T, Shimamoto K, Toriyama S. 1992. Genetically engineered rice resistant to rice stripe virus, an insect-transmitted virus. Proc. Natl. Acad. Sci. USA 89:9865-9869.
Herdt RW. 1991. Research priorities for rice biotechnology. In: Khush GS, Toenniessen GH, editors. Rice biotechnology. Manila (Philippines): International Rice Research Institute, and Wallingford (UK): CAB International. p 19-54.
Huntley CC, Hall TC. 1996. Interference with brome mosaic virus replication in transgenic rice. Mol. Plant-Microbe Interact. 9:164-170.
Koller A, Dalmay T, Burgyan J. 1993. Defective interfering RNA-mediated resistance against cymbidium ringspot tombusvirus in transgenic plants. Virology 193:313-318.
Kouassi N, Brugidou C, Chen L, Ngon M, Yassi A, Beachy RN, Fauquet CM. 1997. Transgenic rice plants expressing rice yellow mottle virus coat protein gene. Int. Rice Res. Notes 22:14-15.
Lindbo JA, Dougherty WG. 1992. Pathogen-derived resistance to a potyvirus: immune and resistant phenotypes in transgenic tobacco expressing altered forms of a potyvirus coat protein nucleotide sequence. Mol. Plant-Microbe Interact. 5:144-153.
Lodge JK, Kaniewski WK, Tumer NE. 1993. Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. Proc. Natl. Acad. Sci. USA 90:7089-7093.
Malyshenko SI, Kondakova OA, Nazarova JV. 1993. Reduction of tobacco mosaic virus accumulation in transgenic plants producing non-functional viral transport proteins. J. Gen. Virol. 74:1149-1156.
Matsumura T, Tabayashi N. 1995. Transformation of rice plants by plant reovirus genes. Sem. Virol. 6:133-139.

McLean GD, Evans G. 1997. Some potential benefits from the use of transgenic crops with resistance to non-selective herbicides in Australian farming systems. In: McLean GD, Waterhouse PM, Evans G, Gibbs MJ, editors. Commercialisation of transgenic crops: risk, benefit and trade considerations. Proceedings of a workshop, 11-13 March 1997, Canberra, Australia. Canberra (Australia): Australian Government Publishing. p 293-300.

Miller WA, Koev G, Mohan BR. 1997. Are there risks associated with transgenic resistance to luteoviruses? Plant Dis. 81:700-710.

Mueller E, Gilbert J, Devenport G, Brigneti G, Baulcombe DC. 1995. Homology-dependent resistance: transgenic virus resistance in plants related to homology-dependent gene silencing. Plant J. 7:1001-1013.

Omura T. 1995. Genomes and primary protein structures of phytoreoviruses. Sem. Virol. 6:97-102.

Osbourn JK, Watts JW, Beachy RN, Wilson TMA. 1989. Evidence that nucleocaspid disassembly and a later step in virus replication are inhibited in transgenic tobacco expressing TMV coat protein. Virology 172:370-373.

Osbourn JK, Sarkar S, Wilson TMA. 1990. Complementation of coat protein-defective TMV mutants in transgenic tobacco plants expressing TMV coat protein. Virology 179:921-925.

Pinto YM, Kok RA, Baulcombe DC. 1999. Resistance to rice yellow mottle virus (RYMV) in cultivated African rice varieties containing RYMV transgenes. Nat. Biotechnol. 17:702-707.

Powell-Abel PA, Nelson RS, De B, Hoffmann N, Rogers SG, Fraley RT, Beachy RN. 1986. Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. Science 232:738-743.

Pressoir G, Albar L, Ahmadi N, Rimbault I, Lorieux M, Fargette D, Ghesquiere A. 1998. Genetic basis and mapping of the resistance to rice yellow mottle virus. II. Evidence of a complementary epistasis between two QTLs. Theor. Appl. Genet. 97:1155-1161.

Qu RD, Bhattacharya M, Laco GS, DeKochko A, Rao BLS, Kaniewska MB, Elmer JS, Rochester DE, Smith CE, Beachy RN. 1991. Characterization of the genome of rice tungro bacilliform virus: comparison with Commelina yellow mottle virus and caulimoviruses. Virology 185:354-364.

Ramasamy C, Jatileksono T. 1996. Intercountry comparison of insect and disease losses. In: Evenson RE, Herdt RW, Hossain M, editors. Rice research in Asia: progress and priorities. Manila (Philippines): International Rice Research Institute, and Wallingford (UK): CAB International. p 305-316.

Sanford JC, Johnston SA. 1985. The concept of parasite-derived resistance: deriving resistance genes from the parasite's own genome. J. Theor. Biol. 113:395-405.

Sebastian LS, Ikeda R, Huang N, Imbe T, Coffman WR, McCouch SR. 1996. Molecular mapping of resistance to rice tungro spherical virus and green leafhopper. Phytopathology 86:25-30.

Shen P, Kaniewska M, Smith C, Beachy RN. 1993. Nucleotide sequence and genomic organization of rice tungro spherical virus. Virology 193:621-630.

Sijen T, Wellink J, Hiriart JB, Van Kammen A. 1996. RNA-mediated virus resistance: role of repeated transgenes and delineation of targeted regions. Plant Cell 8:2277-2294.

Sivamani E, Huet H, Shen P, Ong CA, deKochko A, Fauquet C, Beachy RN. 1999. Rice plants (*Oryza sativa* L.) containing rice tungro spherical virus (RTSV) coat protein transgenes are resistant to virus infection. Mol. Breed. 5:177-185.

Stam M, de Bruin R, Kenter S, van der Hourn RAL, van Blokland R, Mol JNM, Kooter JM. 1997. Post-transcriptional silencing of chalcone synthase in *Petunia* by inverted transgene repeats. Plant J. 12:63-82.

Tacke E, Salamini F, Rohde W. 1996. Genetic engineering of potato for broad spectrum protection against virus infection. Nat. Biotechnol. 14:1597-1601.

Tavladoraki P, Benvenuto E, Trinca S, DeMartinis D, Cattaneo A, Galeffi P, De Martinis D. 1993. Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack. Nature 366:469-472.

Tenllado F, Garcia-Luque I, Serra MT, Diaz-Ruiz JR. 1995. *Nicotiana benthamiana* plants transformed with the 54-kDa region of the pepper mild mottle tobravirus replicase gene exhibit two types of resistance responses against viral infection. Virology 211:170-183.

Tepfer M, Balazs E. 1997. Virus-resistant transgenic plants: potential ecological impact. Heidelberg, Berlin (Germany): Springer-Verlag. 126 p.

Tien P, Wu G. 1991. Satellite RNA for the biocontrol of plant disease. Adv. Virus Res. 39:321-339.

Truve E, Aaspôllu A, Honkanen J, Puska R, Mehto M, Hassl A, Teeri TH, Kelve M, Seppanen P, Saarma M. 1993. Transgenic potato plants expressing mammalian 2′-5′ oligoadenylate synthetase are protected from potato virus X infection under field conditions. Bio/Technology 11:1048-1052.

Upadhyaya NM, Yang M, Kositratana W, Ghosh A, Waterhouse PM. 1995. Molecular analysis of rice ragged stunt oryzavirus segment S9 and sequence conservation among isolates from Thailand and India. Arch. Virol. 140:1945-1956.

Upadhyaya NM, Li Z, Yang M, Kositratana W, Waterhouse PM. 1996. Rice ragged stunt virus synthetic resistance genes and japonica rice transformation. In: Khush GS, editor. Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct. 1995, Manila (Philippines). Manila (Philippines): International Rice Research Institute. p 773-779.

Upadhyaya NM, Ramm K, Li Z, Kositratana M, Waterhouse PM. 1998a. Rice ragged stunt oryzavirus genome segment S4 could encode an RNA dependent RNA polymerase and a second protein of unknown function. Arch. Virol.143:1815-1822.

Upadhyaya NM, Li Z, Ramm K, Gaudron J, Waterhouse PM, Wang M-B, Kositratana W, Gong Z-X. 1998b. Production of transgenic rice with rice ragged stunt virus synthetic resistance genes. In: Drew RA, editor. Proceedings of the International Symposium on Biotechnology of Tropical and Subtropical Species. Acta Hortic. 461:393-400.

van den Elzen PJM, Huisman MJ, Posthumus-L, Willink D, Jongedijk E, Hoekema A, Cornelissen BJ. 1989. Engineering virus resistance in agricultural crops. Plant Mol. Biol. 13:337-346.

Wang M-B, Waterhouse PM. 2000. High efficient silencing of a β-glucuronidase gene in rice is correlated with repetitive transgene structure but is independent of DNA methylation. Plant Mol. Biol. 43:67-82.

Watanabe Y, Ogawa T, Takahashi H, Ishida I, Takeuchi Y, Yamamoto M, Okada Y. 1995. Resistance against multiple plant viruses in plants mediated by double-stranded RNA-specific ribonuclease. FEBS Lett. 372:165-168.

Waterhouse PM, Graham MW, Wang M-B. 1998. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc. Natl. Acad. Sci. USA 95:13959-13964.

Waterhouse PM, Smith NA, Wang M-B. 1999. Virus resistance and gene silencing: killing the messenger. Trends Plant Sci. 4:452-457.

Waterhouse PM, Upadhyaya NM. 1998. Genetic engineering of virus resistance. In: Shimamoto K, editor. Molecular biology of rice. Tokyo (Japan): Springer-Verlag. p 257-282.
Wilson TMA, Watkins PAC. 1986. Influence of exogenous viral coat protein on the cotranslational disassembly of tobacco mosaic virus (TMV) particles *in vitro*. Virology 149:132-135.
Wisniewski LA, Powell PA, Nelson RS, Beachy RN. 1990. Local and systemic spread of tobacco mosaic virus in transgenic tobacco. Plant Cell 2:559-567.
Yan Y-T, Wang J-F, Qiu B-S, Tien P. 1997. Resistance to rice stripe virus conferred by expression of coat protein in transgenic indica rice plants regenerated from bombarded suspension cultures. Virol. Sin. 12:260-269.
Yang X-C, Ye Y, Zhu F, Liu Y-L, Kang L-Y, Wang X-F, Tien P. 1997. Ribozyme-mediated high resistance against potato spindle tuber viroid in transgenic potatoes. Proc. Natl. Acad. Sci. USA 94:4861-4865.
Yassi MNA, Ritzenthaler C, Brugidou C, Fauquet C, Beachy RN. 1994. Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. J. Gen. Virol. 75:249-257.
Zheng HH, Li Y, Yu ZH, Li W, Chen MY, Ming XT, Casper R, Chen ZL. 1997. Recovery of transgenic rice plants expressing the rice dwarf virus outer coat protein gene (*S8*). Theor. Appl. Genet. 94:522-527.

Notes

Authors' addresses: N.M. Upadhyaya, Z. Li, M-B. Wang, P.M. Waterhouse, CSIRO Plant Industry, Canberra, ACT 2601, Australia; S. Chen, Zhejiang Academy of Agricultural Sciences, Hangzhou, China; Z-X. Gong, The Institute of Biochemistry and Cell Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 200030, China, E-mail: N.Upadhyaya@pi.csiro.au.
Acknowledgments: We would like to thank research teams from the International Laboratory for Tropical Agricultural Biotechnology (ILTAB), Scripps Research Institute, USA (E. Sivamani, P. Shen, C.A. Ong, H. Huet, A. de Kochko, C. Fauquet, and R. Beachy), from CIAT, Colombia (Z. Lentini, L. Calvert, I. Lozano, E. Tabres, M. Cuervo, and E. Restrepo), and from ITH Jurich, Switzerland (J. Fütterer and I. Potrykus), for allowing access to their unpublished information. We would also like to thank W. Kositratana, Kasetsart University, Thailand, for collaboration on the RRSV genome organization study, K. Ramm, J. Gaudron, J. Gellately, M. Yang, and E. Zinkowsky for technical support, Andrew Eamens for technical editing, and Roger Hull for valuable comments on the manuscript. We gratefully acknowledge financial support from the Rockefeller Foundation for our RRSV research.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Transgenic approaches for generating rice tolerant of dehydration stress

Z.Q. Cheng, J. Targolli, J. Su, C.K. He, F. Li, and R. Wu

Transgenic approaches offer new opportunities to improve tolerance for dehydration stress in rice by incorporating genes that are involved in stress tolerance. In this chapter, we present the results of transforming rice separately with plasmids containing genes that encode a group 3 LEA (late embryogenesis abundant) protein, a group 2 LEA protein, a group 1 LEA protein, and a delta 1-pyrroline-5-carboxylate synthetase (P5CS). In each experiment, the growth of transgenic plants under dehydration stress was shown to be faster than that of nontransformed control plants. To maximize gene expression and plant growth, we compared the results of using a constitutive promoter and a stress-inducible promoter in driving the expression of the *P5CS* gene. Next, the effect of using a matrix attachment region sequence on transgene expression was tested. Finally, we compared two commonly used methods for transforming rice.

Plant productivity is strongly influenced by dehydration stress induced by drought and high-salt and low-temperature conditions. Plants respond to these stresses by displaying complex quantitative traits that involve many genes. These responses result in a wide variety of biochemical and physiological changes, such as the accumulation of various low-molecular-weight organic compounds, collectively known as osmolytes, synthesis of different late embryogenesis abundant (LEA) proteins, and activation of several detoxification enzymes (Bajaj et al 1999, Reddy et al 1999). Abiotic stresses reduce the productivity of several major crops (Boyer 1982, Yancey et al 1982). Thus, understanding and improving stress tolerance in crop plants presents a challenging basic research problem that could also have a significant impact on agricultural productivity.

Ingram and Bartels (1996) and Bray (1997) have reviewed the molecular and physiological basis of dehydration stress tolerance. Gene expression and signal trans-

duction in response to dehydration stress (Shinozaki and Yamaguchi-Shinozaki 1997) and improving plant tolerance of stress by genetic transfer of transcription factor genes have been reported (Jaglo-Ottosen et al 1998, Kasuga et al 1999). Reviews have been published on genetic analysis of osmotic adjustment in crop plants (Zhang et al 1999) and on the need to use transgenic approaches to increase grain yield (Khush 1999, Reddy et al 1999). Cushman and Bohnert (2000) have summarized genomic approaches to plant stress tolerance.

Transgenic approaches offer new opportunities to improve tolerance for dehydration stress in rice by incorporating genes involved in stress tolerance. Genes that can be introduced into rice plants can be broadly classified into four classes. Class I includes genes that encode enzymes synthesizing osmoprotectants, such as proline and polyamines. Class II genes include late embryogenesis abundant (LEA) or LEA-related genes. Class III genes encode enzymes that may protect rice plants from oxidation stress, such as glutathione-S-transferase, and those that synthesize mannitol. Class IV genes encode transcription factors that regulate the expression of the many stress-responsive target genes.

The remaining part of this chapter is devoted to describing the progress that our laboratory has been making with the goal of increasing dehydration-stress tolerance in rice. We also briefly discuss future prospects in this area of research to increase rice yield by minimizing loss of crops because of salt, drought, and low-temperature stress.

Results

Development of a dehydration-stress-inducible promoter system

A strong and constitutive promoter is suitable for high-level expression of a gene of interest that is linked downstream of the promoter. However, a strong constitutive promoter is not always desirable for plant genetic engineering because the gene product may be toxic to the plant. Overexpression of a transgene may also compete for energy and building blocks that are required for protein and RNA synthesis, as well as plant growth and development under normal conditions. Thus, we developed a dehydration-stress-regulated promoter system to confer abscisic acid (ABA)- and/or stress-inducible gene expression. An ABA-response complex (ABRC1) from the barley *HVA22* gene was fused to a minimal rice *Act1* gene promoter (Act1-100P). We then constructed two plasmids with one or four copies of ABRC1 combined with the same Act1-100P and HVA22(I) of barley *HVA22* to drive the expression of *uidA* in transgenic rice plants. Three DNA-blot-positive plant lines were regenerated that showed the correct hybridization pattern for each construct. RNA blot analysis indicated that *uidA* expression is induced by ABA, water-deficit, and NaCl treatments. GUS activity assays in the transgenic plants confirmed that the induction of GUS activity varied from 3- to 8-fold with different treatments or in different rice tissues, and that transgenic rice plants harboring four copies of ABRC1 showed 50% to 200% higher absolute GUS activity both before and after treatments than those with one copy of ABRC1 (Su et al 1998).

LEA proteins and their effect on stress tolerance

LEA proteins were first characterized in cotton as a set of proteins that are highly accumulated in embryos at the late stage of seed development (Dure 1981). Subsequently, many LEA proteins or their genes have been characterized from different plant species (collated by Dure 1992). Based on their common amino acid sequence domains, LEA proteins have been classified into three major groups. The regions of homology among group 3 LEA proteins are composed of tandem repeats of an 11-amino-acid motif that may form an amphiphilic alpha-helix structure (Dure et al 1989, Dure 1993).

The correlation between LEA gene expression or LEA protein accumulation and stress tolerance in several nontransgenic plants provides additional evidence supporting a role of LEA proteins in stress tolerance. For example, in severely dehydrated wheat seedlings, the accumulation of high levels of group 3 LEA proteins was correlated with tissue dehydration tolerance (Ried and Walker-Simmons 1993). Studies of several indica varieties of rice (*Oryza sativa* L.) showed that the levels of group 2 LEA proteins and group 3 LEA proteins in roots were significantly higher, or induced by ABA and salt stress, only in naturally occurring salt-tolerant varieties as compared with salt-sensitive varieties (Moons et al 1995).

A barley group 3 LEA protein, HVA1 (related to the cotton D-7 protein and carrot Dc3 protein), accumulates to high levels in the aleurone layer and embryos during late seed development, correlating with the seed desiccation stage (Hong et al 1988). Expression of the *HVA1* gene is rapidly induced in young seedlings by ABA and by several stress conditions, including dehydration, salt, and extreme temperature (Hong et al 1992). A 26-kDa group 3 LEA protein was induced by ABA and salt stress in a salt-tolerant indica rice variety (Moons et al 1995).

We have taken a transgenic approach to investigating the function of the HVA1 protein in stress protection (Xu et al 1996). The *HVA1* from barley, joined to the rice *Act1* promoter in a suitable plasmid, was introduced into rice suspension cells using the biolistic-mediated transformation method, and a large number of independent transgenic rice plants were generated. Expression of the barley *HVA1* gene in transgenic rice led to high-level, constitutive accumulation of the HVA1 protein in both leaves and roots of transgenic rice plants. Second-generation transgenic rice plants showed significantly increased tolerance for water deficit and salinity. Transgenic rice plants maintained higher growth rates than nontransformed control plants under stress conditions. The increased tolerance was also reflected by delayed development of damage symptoms caused by stress and by improved recovery upon the removal of stress conditions. Thus, by using a transgenic approach, this study provides direct evidence supporting the hypothesis that LEA proteins play an important role in the protection of plants under water- or salt-stress conditions (Xu et al 1996).

Group 1 LEA proteins are characterized by their high glycine content and amino acids with charged R-groups (ca. 40%) (Stacey et al 1995). Group 1 LEA proteins also contain a hydrophilic 20-amino-acid sequence that occurs in tandem one to four times in the various proteins (Dure et al 1989). Group 2 LEA proteins, which are also referred to as dehydrins, are characterized by a highly conserved sequence

(KIKEKLPG) in the carboxyl terminus. These proteins have been predicted to function as chaperones (Dure 1992). Levels of the group 1 and 2 LEA proteins were significantly higher in vegetative tissues (such as roots) of salt- and drought-tolerant nontransgenic wheat, rice, and fingermillet genotypes than in those of sensitive genotypes under salt- or drought-stress conditions or after the application of ABA on plants (Morris et al 1995, Moons et al 1995). Thus, the physiological characteristics of group 1 and 2 LEA proteins support their function for providing cellular tolerance for conditions such as drought and salt.

Two wheat genes, PMA80 and PMA1959, have been demonstrated to be responsive to ABA and environmental stress. These two genes were cloned (Morris et al 1995). To test the function of a group 2 and a group 1 LEA protein gene in transgenic rice, we constructed two plasmids. One plasmid, pJPM44, contains a wheat LEA group 2 protein (PMA80). The other, pJPM45, contains a wheat LEA group 1 protein (PMA1959). The structures of the two plasmids are shown in Figure 1. Both plasmids were transferred into rice and fertile transgenic plants were obtained. Second-generation transgenic plants were subjected to dehydration or salt stress. The results showed that accumulation of either PMA80 or PMA1959 significantly increased the tolerance of transgenic rice plants for both dehydration stress and salinity stress (Table 1), as reflected by higher fresh and dry weight compared with NT (nontransformed) control plants.

Proline level and dehydration tolerance

Overproduction of proline in transgenic rice plants was found to improve tolerance for dehydration stress. Proline accumulation has been shown to be correlated with

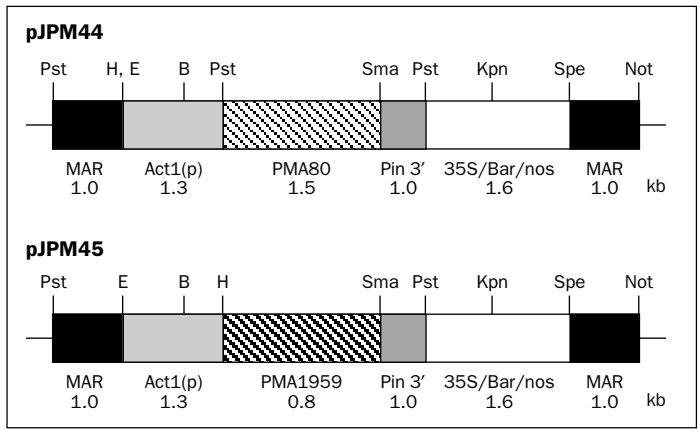


Fig. 1. Schematic diagram of plasmids pJPM44 and pJPM45. Act1(p) = rice *actin1* promoter with *actin1* first intron; *PMA80* = a wheat cDNA coding an LEA group 2 protein; *PMA1959* = a wheat cDNA coding an LEA group 1 protein; Pin3′ = potato proteinase inhibitor II 3′ noncoding sequence; 35S(p) = cauliflower mosaic virus 35S promoter; *nos* = 3′ terminator from the *nos* gene; MAR = the tobacco Rb7 matrix attachment region. B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII.

Table 1. Fresh and dry shoot weight of R_1 transgenic rice or NT (nontransformed) control plants after dehydration stress[a] or salt stress[a].

Plasmid	Fresh shoot weight (g)		Dry shoot weight (g)	
	Dehydration stress[b]	Salt stress[c]	Dehydration stress[b]	Salt stress[c]
pJPM44 (with group 2 LEA gene *PMA80*)	4.8	4.7	1.05	1.02
pJPM45 (with group 1 LEA gene *PMA1959*)	4.3	4.3	0.99	0.98
NT control	3.2	3.4	0.70	0.67

[a]R_1 plants from pJPM44-transformed plant lines were germinated. Six wk later, six plants from each plant line (total 36 R_1 plants) were used for a dehydration-stress-tolerance experiment and 36 R_1 plants were used for a salt-stress-tolerance experiment. Similarly, pJPM45-transformed plant lines and NT controls were grown and tested in the same way. [b]For dehydration-stress tolerance, 36 R_1 plants (from six plant lines) and 36 R_1 nontransgenic control plants were not watered for 8 d and then watered for 8 d. Then the fresh shoot weights and dry shoot weights were measured. The average value of the 36 plants was expressed as weight (g) per plant. [c]For salinity-stress tolerance, 36 R_1 plants grown in six pots were placed in a tray containing 5 L water plus liquid fertilizer and 200 mM NaCl. Water was added daily to maintain the volume in the trays. After 9 d of salt stress, water was added to the pots directly to wash out the salt. Eight days later, the fresh shoot weights and dry shoot weights of the 36 plants were measured and the average value was expressed as weight (g) per plant.

osmoregulation and tolerance for salt and drought stress in plants (Delauney and Verma 1993). Proline was believed to act as an osmoprotectant, as a sink for energy to regulate redox potentials, or as a compound that protects macromolecules from denaturation. Kavi Kishor et al (1995) demonstrated that overproduction of proline enhanced root biomass and flower development in transgenic tobacco under water-stress conditions.

Zhu et al (1998) introduced a delta 1-pyrroline-5-carboxylate synthetase (P5CS) cDNA from mothbean into rice cells by the biolistic method. Fertile transgenic rice plants were regenerated. It was found that expression of this P5CS transgene led to stress-induced overproduction of the P5CS enzyme and proline accumulation. Second-generation (R_1) transgenic rice plants showed an increase in biomass under salt-stress and water-stress conditions compared with the nontransformed control plants.

Cheng et al (2000) compared the extent of stress tolerance of transgenic rice plants by driving the P5CS transgene expression with either a constitutive promoter or a stress-inducible promoter. To test the expression of *p5cs* cDNA in transgenic plants, three plasmids were constructed (Fig. 1). pJS102 (Su et al 1998) contains a constitutive promoter (rice *Act1*) for driving the expression of the P5CS cDNA (*p5cs*). pJS112 contains an ABA-inducible promoter, the AIPC (Su et al 1998), for driving the expression of *p5cs*; the AIPC used in the present experiment contains four copies of ABRC1 from the barley *HVA22* gene, the rice *Act1* minimal promoter (180 bp), and the HVA22 intron. In our previous work, using *uidA* gene as the reporter, AIPC-directed transgene expression was shown to be induced 3- to 8-fold by ABA (Su et al 1998). pJS110 was used as a control because it contains the same components as pJS112 except that the reporter gene is *uidA*. All three plasmids contain the bacterial phosphinothricin acetyl-transferase gene *(bar)* driven by the *CaMV 35S* promoter for selection.

These three plasmids were used to transform rice suspension cells. Transgenic rice plants were regenerated. Transgene copy numbers were estimated by comparing the intensity of hybridization bands of *Bam*HI-digested genomic DNA from transgenic plants to bands of known quantities of plasmid DNA. The copy numbers of the transgene, including rearranged copies, varied between 1 and 9 in different plants (Cheng et al 2000). The *p5cs* expression was first analyzed at the mRNA level by RNA blots. Two-month-old R_2 plants were subjected to water-stress treatment by withholding water for 6 d, whereas nonstressed plants were supplied with water continuously and used for basal mRNA-level analysis. To detect the salt-stress-induced mRNA level, 200 mM NaCl solution was used to water *p5cs*-transgenic and *uidA*-transgenic plants (as control) for 48 h. Our results indicated that the *p5cs* mRNA production driven by the inducible promoter was low, but it did increase from 3- to 7-fold with different transgenic lines. Plants with constitutive *p5cs* expression gave high *p5cs* mRNA levels with or without stress, as expected.

Next, we measured the free proline level in transgenic rice plants. One R_2 line of *Act1-p5cs* plants (L1) and two R_2 lines of *AIPC-p5cs* plants (L5 and L7) were chosen. Before water-stress treatments (0-d sample), the proline level produced by the constitutive expression of the *p5cs* transgene (L1) reached 319%, the value of the L3 control plant. When a stress-inducible promoter (*AIPC*) was used, the proline level reached only 138% (L5) and 121% (L7) of that of L3. As the stress proceeded, however, the proline levels of lines L5 and L7 increased at relatively higher rates and reached final levels approaching that of line L1 after 8 d of water stress. For salt stress, the results are generally similar to those of the water-stress treatment (Cheng et al 2000).

Growth performance of transgenic plants under water-stress conditions. Cheng et al (2000) tested the growth performance of R_2 plants in soil. Under nonstress conditions in soil, no significant differences were observed between *p5cs*-containing transgenic plants and *AIPC-uidA* control plants in their growth performance during the entire period of the experiment. Upon withholding water from the trays, the absolute water content in the soil decreased from 35% to 12% after 7-d water stress. The plants were watered for 2 d and then stressed again by withholding water for another 7 d. Following two cycles of water stress, the leaves of *AIPC-uidA* control plants started to turn yellow and the *Act1-p5cs* plants showed a low growth rate, whereas the two *AIPC-p5cs* plants with a stress-inducible promoter showed healthy growth. After four cycles of water stress, more severe symptoms, such as leaf chlorosis (in both the L3 control and *Act1-p5cs* plants) or death of leaf tips (in control plants only), were found. The *AIPC-p5cs* plants still showed a relatively high rate of growth and less severe leaf chlorosis. Data in Table 2 show the average fresh shoot and root weights of the plants after four cycles of 7-d water stress. The results indicated that, under water stress, the *AIPC-p5cs* plants (L5 and L7), which contained a stress-inducible promoter to drive the *p5cs* expression, grew much faster than the *Act1-p5cs* plants (L1), which contained a constitutive promoter for driving the *p5cs* expression. The difference between using the two types of promoters was highly significant ($P<0.01$; $t = 5.88–7.64$).

Table 2. Growth performance of transgenic plants in soil under water-stress conditions.

Rice line	Major plasmid components	Fresh shoot wt[a] (mg plant^{-1})	Fresh root wt (mg plant^{-1})
JS110 (L3)	*Act1-uidA*	300 ± 20 (100)	90 ± 12 (100)
JS102 (L1)	*Act1-p5cs*	550 ± 60 (183)	130 ± 20 (144)
JS112 (L5)	*AIPC-p5cs*	940 ± 100 (310)	220 ± 30 (224)
JS112 (L7)	*AIPC-p5cs*	730 ± 70 (243)	170 ± 20 (189)

[a]Mean ± standard error represents the averages of 6 plants. Values in parentheses are the percentages of *p5cs*-transgenic plants compared with control plants (L3), represented by 100. The spread of data within each set of 6 plants was rather small.

Growth performance of transgenic rice plants under salt-stress conditions. To create high soil salinity, 300 mM NaCl solution was added to the trays in which the pots were placed. We found that, similar to the results under water-stress conditions, *AIPC-p5cs* transgenic rice plants under salt-stress conditions grew better than *Act1-p5cs* plants. In conclusion, stress-inducible transgene expression in *p5cs* plants shows significant advantages over constitutive expression of the *p5cs* transgene in the growth of rice plants under salt- and water-stress conditions (Cheng et al 2000).

Effects of matrix attachment region sequences on transgene expression

Transformation of cereal crop plants has become a relatively routine procedure in many laboratories. However, many unexpected and often undesirable findings have been observed in transgenic plants with respect to transgene expression. The more commonly observed phenomena include transgene silencing and variations in the expression levels of transgenes between individual transformants that are not correlated with copy number (Kumpatla et al 1998, Stam et al 1997). However, chromatin structure appears to affect the level of gene expression by exposing or not exposing DNA sequences to the transcriptional machinery (Stam et al 1997).

Matrix attachment regions (MARs), also known as scaffold attachment regions (SARs), are DNA sequences that bind to a cell's proteinaceous nuclear matrices and are thought to form DNA loop domains. MAR sequences are thought to form the boundaries of transcriptionally active DNA loop domains, thereby blocking the influence of neighboring sequences on gene expression. When a transgene flanked by MAR sequences is introduced into a cell, the transgene can presumably form an independent loop domain once it is integrated into the chromosome and thereby minimize position effects (Spiker and Thompson 1996). Different MAR sequences have been shown to increase the levels of transgene expression (for a review, see Holmes-Davis and Comai 1998) and reduce variations in transgene expression between transformants in dicotyledonous plants (Allen et al 2000, Mlynarova et al 1994). Transgene silencing in monocots is a widespread phenomenon and has been reviewed recently (Iyer et al 2000), but studies of the effects of MAR sequences on transgenic expression in monocots are limited (Vain et al 1999).

We now show the effects of the tobacco Rb7 MAR sequence on transgene expression in rice in terms of the accumulation of green fluorescent protein (GFP) by using three plasmids. Among these three, two plasmids contain two identical copies of the MAR sequence, which flanks a GFP gene. A control plasmid contains no MAR sequences. Figure 2 shows the structure of the three plasmids. Microprojectile bombardment was used to separately introduce these three plasmids into rice calli. Transgenic rice plants were regenerated and gene expression was measured in rice leaf extracts of 2-mo-old transgenic plants. By comparing transgenic plants with the corresponding control plants, transgenic plants harboring each of the two MAR-containing plasmids resulted in average gene expression levels enhanced by 3.3-fold and 18.4-fold, respectively, as shown in Table 3. More specifically, we compared GFP reporter gene expression in plasmids with and without the Rb7 MAR sequence to study the effects of the MAR sequence on transgene expression in rice. In nontransformed plants, the autofluorescence values averaged 5.8 FLU (fluorescent units) μg^{-1} protein. GFP values in 12 R_0 transgenic plants (pSBG700) ranged from 16.8 to 84.1, with an average of 47.1 FLU μg^{-1} (Table 3). When the GFP gene was flanked by the Rb7 MAR sequence (pJPM5), GFP values in 10 R_0 plants ranged from 605 to 1,333, with an average of 867 FLU μg^{-1} protein, an average that is 18-fold higher than that of those plants harboring a similar plasmid but lacking the MAR sequence. It appeared that there was no obvious relationship between copy number of the transgene and GFP level in rice plants containing the Rb7 MAR sequence (data not shown).

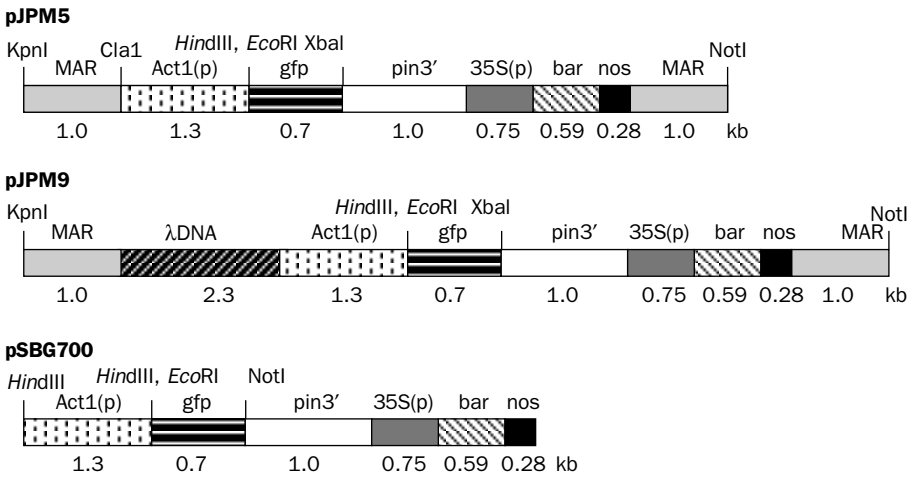


Fig. 2. Schematic diagram of plasmids pJPM5, pJPM9, and pSBG700. The pJPM5 and pJPM9 plasmids contain the Rb7 MAR sequence, whereas pSBG700 does not. MAR = tobacco Rb7 matrix attachment region sequence; Act1(p) = rice *actin1* promoter with *actin1* first intron; *gfp* = a modified gene coding a green fluorescent protein; *bar* = the bacterial phosphinothricin acetyltransferase gene; Pin3′ = 3′ terminator region of potato proteinase inhibitor 2 gene; 35S(p) = cauliflower mosaic virus 35S promoter; nos = terminator region of the *nos gene*; lambda DNA = 2.3-kb *Hin*dIII fragment containing lambda phage sequence from 25,100 bp to 27,479 bp isolated from *Hin*dIII-digested lambda DNA.

Table 3. The levels of green fluorescent protein (FLU, fluorescent units, per microgram protein) in transgenic and nontransformed rice plants.

GFP plasmid	Number of R_0 lines	Range of GFP values	Average GFP level (fold difference)[a]	Variation in transgene expression
pSBG700	12	16.8–84.1	47.1[b] (1.0)	5.0-fold
pJPM5/MAR	10	605–1,333	867[c] (18.4)	2.2-fold
pJPM9/MAR	6	66.8–356	155[d] (3.3)	5.3-fold
None	6	4.4–8.0	5.8	

[a]Fold difference is the difference in green fluorescent protein (GFP) level between plasmids with the matrix attachment region (MAR) sequence compared with those without the MAR sequence (pSBG700, with the value set at 1.0-fold). The values were not corrected for the background in the nontransformed control. [b]Standard error is 3.2; standard deviation is 12.5. [c]Standard error is 65.8; standard deviation is 132.2. [d]Standard error is 12.6; standard deviation is 54.6.

To study the effects of increased distance between the Rb7 MAR sequence and the *actin1* promoter, we used plasmid pJPM9, in which a 2.3-kb *Hin*dIII fragment from lambda DNA was inserted between the MAR sequence and the *actin1* promoter. The rice *actin1* promoter drives the GFP reporter gene expression in pJPM9. We selected this fragment after confirming that there was no homologous sequence corresponding to the 2.3-kb *Hin*dIII fragment in the rice genome by DNA-blot analysis. The average GFP value in six R_0 transgenic plants transformed with pJPM9 was 155 FLU μg^{-1} protein (Table 3), which was only 3.3-fold higher than that of the pSBG700 control plasmid without the MAR sequence. The average GFP values in pJPM9 plants were only 20% as high as those of pJPM5-transformed plants. Moreover, the GFP values in the different transgenic plants transformed with pJPM9 varied up to 5.3-fold, which is more than the 2.2-fold variation in plants transformed with pJPM5.

We have shown that the tobacco Rb7 MAR sequence significantly increased the average GFP values in leaves of transformed rice plants. To test whether the Rb7 MAR sequence also acts as an enhancer element in rice, we conducted transient assays using 14-d-old calli derived from mature rice embryos. Plasmids pJPM5 and pJPM9, with the Rb7 MAR sequence flanking the *gfp-bar* cassette, were introduced into calli by the biolistic method. We used plasmid pSBG700 (without the MAR sequence) as a control. Nonbombarded calli were used as an internal control as a reference while quantifying GFP expression of the bombarded calli. GFP expression was visualized in calli under a BH-2 Olympus Fluorescence Microscope (data not shown). The GFP transient expression values were measured 48 h after particle bombardment. The average relative GFP value in the pSBG700 control plasmid was 364 FLU μg^{-1} protein, whereas the values for pJPM5 and pJPM9 were 620 and 375 FLU μg^{-1} protein, respectively (data not shown). Thus, from these experiments, the tobacco Rb7 MAR sequence increased the GFP value by only 70%. This minimal stimulation can be attributed to an enhancer effect rather than an effect on the chromatin structure. It appears that, in order for the Rb7 MAR sequence to greatly enhance transgene expression in our system, integration of the plasmid into a chromosome is required, suggesting that this is mainly due to a boundary effect.

Since one major goal of plant scientists and breeders is to maximize expression levels of transgenes, especially in the production of transgene-encoded protein, inclusion of the Rb7 MAR sequence would be beneficial.

Comparison of the particle-bombardment and *Agrobacterium*-mediated transformation methods in rice

In rice, the particle-bombardment method was first used successfully to introduce plasmids into rice cells and show integration of the transgene (Cao et al 1990). However, no transgenic plants were regenerated. Christou et al (1991) reported the transformation of rice immature embryos and regeneration of fertile plants. The transgene copy number ranged between 1 and 6. Cao et al (1992) introduced into rice suspension culture cells a *bar*-gene-containing plasmid and regenerated fertile transgenic rice plants. The transgene copy number was between 1 and 5. Duan et al (1996) reported that there were 6–8 copies of transgenes in transgenic rice plants obtained by particle bombardment.

Chan et al (1993) were the first to successfully use *Agrobacterium*-mediated transformation to regenerate fertile transgenic rice plants, but only a few plants were produced. Hiei et al (1994) used a helper plasmid to supply *vir* gene function and established an efficient transformation method with many plants regenerated. The transgene copy number was between 1 and 5. Since then, this method has become an important transformation method in cereal plants, including rice.

Reports from different laboratories suggest that the transgene copy number is higher in transgenic plants obtained by particle bombardment than in those obtained by the *Agrobacterium*-mediated method. However, evidence based on well-controlled comparison is lacking because different laboratories have used different vectors and procedures. In our work, we used the same plasmid for both the *Agrobacterium*-mediated method and the particle bombardment transformation method to transform rice. The work was carried out by the same person using the same starting materials, media, etc. Thus, the results can be compared directly.

Results from past years have indicated that transgene copy number may affect the level of gene expression in transgenic plants, and the number of the rearranged copies of the transgene often results in gene silencing (Holmes and Comai 1998, Kumpatla et al 1998, Allen et al 2000, Iyer et al 2000). Therefore, an important consideration is to determine the transgene copy number of both intact and rearranged copies and to correlate this information with expression levels.

We have constructed a plasmid, pAc1PG-CAM, using the vector pCAMBIA1300 (from Richard Jefferson of CAMBIA). The plasmid contains the following components: *Arabidopsis Actin1* gene promoter/*Gus*/Nos 3′/pCAMBIA 1300. This plasmid was used for transformation using two different methods. In both methods, rice calli were induced in MS medium (see Zhang and Wu 1988) from the mature embryos of rice cv. TNG67. In the particle bombardment method, 15-d-old embryogenic calli were bombarded with gold beads coated with pAc1PG-CAM DNA according to the procedure of Cao et al (1992). Calli transformed with pAc1PG-CAM DNA were selected in plates containing MS medium supplemented with 50 mg L^{-1} hygromycin

(Zhang and Wu 1988). Plants were regenerated and transplanted into sterilized soil and grown in the greenhouse (30 °C day and 24 °C night with a supplemented photoperiod of 10 h).

In the *Agrobacterium*-based method, *A. tumefaciens* strain LBA4404(pAc1PG-CAM) was grown as described by Hiei et al (1994). Three-wk-old calli induced from mature rice seeds on MS medium were infected by LBA4404(pAc1PG-CAM) or LBA4404(pTOK233) by co-cultivation for 2 d. The remainder of the experiment followed the protocol of Hiei et al (1994).

Transformation efficiency. Transformation efficiency was calculated based on the number of transgenic plant lines divided by the total number of calli used in either the *Agrobacterium*-mediated or particle bombardment method of transformation. Each of the transgenic plant lines contained a plasmid with a specific gene of interest, which was confirmed by Southern blotting with the specific gene as a probe. We found that the transformation efficiency of the pAc1PG-CAM using the *Agrobacterium*-mediated method was between 11% and 16%, which was approximately twice as high as when using the particle bombardment method (6–7% efficiency).

Copy number of the specific transgene. Transgene copy numbers were estimated by Southern blot analysis of randomly chosen transgenic plants. The *gus* gene was used as a probe for Southern blotting of pAc1PG-CAM- or pTOK233-transformed plants. The *PM80* gene was used as a probe for Southern blotting of pJPM44-transformed plants. Single-cut enzymes were used to digest the genomic DNA isolated from transgenic plants. To more accurately determine transgene copy numbers, the digested genomic DNA was separated by electrophoresis on long agarose gel, with low voltage for longer periods of time, to separate DNA fragments completely. This avoided the overlapping of two DNA fragments with similar size. The hybridization pattern of each transgenic plant line gives the transgene copy number.

Table 4 lists the transgene copy numbers. Through *Agrobacterium*-mediated transformation, pAc1PG-CAM-transformed plants had 1–4 copies of the transgene, with an average of 2.3 among individual transgenic plant lines. By the particle bombardment method, pAc1PG-CAM-transformed plant lines had 1–10 copies of transgenes, with an average of 4.3. Thus, the transgene copy number in plants obtained by *Agrobacterium*-mediated transformation was much lower than in those obtained by the particle bombardment method. This conclusion is further confirmed by comparing the transgene copy numbers of plants transformed with pTOK233 or pJPM44. The pTOK233-transformed plants regenerated through *Agrobacterium*-mediated transformation had 1–4 copies of the transgene, with an average of 2.0. The pJPM44-transformed plants obtained by particle bombardment had 2–11 copies of the transgene, with an average of 4.5 (data not shown).

Frequency of DNA rearrangement of the transgene expression cassette. Our observation of DNA rearrangement based on Southern blotting focused on transgene expression cassettes containing a promoter, a specific gene of interest, and a terminator. The enzyme chosen for digestion of genomic DNA has two cut sites in the original plasmid located on either side of the expression cassette. This enables the enzyme to cut and release the expression cassette from the genomic DNA of transgenic rice

Table 4. Copy number of transgenes including rearranged copies.

Plasmid	Line no.	Copy no.	No. of hybridization bands[a]	No. of intact expression cassettes	No. of rearranged bands	GUS expression
pAc1PG-CAM	5	2	1	2	0	+
(*Agrobacterium*	6	4	1	4	0	+[b]
method)	7	2	1	2	0	+[b]
	8	1	1	0	1	–
	9	2	1	2	0	+
	10	4	1	4	0	+[b]
	11	1	1	1	0	+[b]
	12	3	2	2	1	+[b]
Average		2.3	1.1	2.1	0.25	
pAc1PG-CAM	14	5	5	2	4	+
(particle	15	1	1	0	1	–
bombardment	16	4	4	1	3	+
method)	17	5	3	3	2	+
	18	5	5	1	4	+
	19	10	6	5	5	+
	20	6	6	1	5	+
	21	2	2	1	1	+[b]
	22	2	1	1	1	+
	23	3	2	0	3	–
	24	2	3	1	2	+
	26	4	4	1	3	+
Average		4.3	3.5	1.4	2.8	

[a]The total number of hybridization bands when a two-cut-site enzyme was used to digest genomic DNA in Southern blotting. [b]High expression level of GUS was observed in the anther, including pollen of these lines.

plants. The size and number of the expression cassette showed the copy number of intact and rearranged copies of the transgene. Our results showed that the average number of rearranged copies of the transgenic expression cassette in transgenic plants with pAc1PG-CAM obtained by *Agrobacterium*-mediated transformation was only 0.25, much lower than the 2.8 obtained by using the particle bombardment method.

Transgene expression and DNA rearrangement of expression cassettes. Histochemical analysis of GUS expression was carried out on pAc1PG-CAM-transformed R_0 plants obtained by both the *Agrobacterium*-mediated transformation method and particle bombardment method. GUS expression levels were observed in young root tips of all plant lines transformed with pAc1PG-CAM by the *Agrobacterium*-mediated method except for line 8. Even though there was one copy of the *gus* gene integrated into the chromosome, there was no intact *gus* gene expression cassette in this line because of DNA rearrangement. GUS expression was observed in young root tips of all transgenic plant lines transformed with pAc1PG-CAM by the particle bombardment method except in lines 15 and 23. Even though there were three copies of the *gus* gene in line 23 and one copy in line 15, no histochemical GUS expression was observed, presumably because of gene silencing related to DNA rearrangement. In plasmid pAc1PG-CAM, *Act1*, which is mainly a pollen-specific promoter with low

expression levels in young root tips, was isolated from *Arabidopsis* (An et al 1996) Thus, randomly chosen plant lines 6, 7, 8, 10, 11, 15, 16, 17, 21, 23, and 24 were used for GUS histochemical analysis of pollen during the flowering stage. High GUS expression was observed in the pollen of plant lines 6, 7, 10, 11, 12, and 21. However, no GUS expression was observed in lines 8, 15, and 23. This further confirms the results observed in young root tips.

Future prospects

One goal of agricultural scientists is to produce superior transgenic rice plants with high levels of tolerance for drought and salinity stress. Based on current knowledge, this goal can be best achieved by using a combination of the following strategies:

1. Use suitable dehydration-stress-inducible promoters to drive transgene expression.
2. Construct plasmids that contain several stress-tolerance genes, as well as a transcription factor gene, each of which has been shown to increase stress tolerance.
3. Use suitable MAR sequences to flank the genes of interest to maximize high-level expression.
4. Use the *Agrobacterium*-mediated method to transform cells in 3-wk-old calli, regenerate transgenic rice plants, and choose at least 20 transgenic plants, each harboring a single copy of the input plasmid(s).
5. Choose homozygous transgenic plant lines with stable high-level expression of the transgene for at least six generations.
6. Carry out field tests on at least ten plant lines and measure yield of rice grains per plant under conditions with or without drought or salinity stress.

References

Allen GC, Spiker S, Thompson WF. 2000. Use of matrix attachment regions (MARs) to minimize transgene silencing. Plant. Mol. Biol. 43:361-376.

An Y, Huang S, McDowell M, McKinney EC, Meagher RB. 1996. Conserved expression of the *Arabidopsis* ACT1 and ACT3 actin subclass in organ primordia and mature pollen. Plant Cell 8:15-30.

Bajaj S, Targolli J, Liu L-F, Ho T-HD, Wu R. 1999. Transgenic approaches to increase dehydration-stress tolerance in plants. Mol. Breed. 5:493-503.

Boyer JS. 1982. Plant productivity and environment. Science 218:493-498.

Bray EA. 1997. Plant responses to water deficit. Trends Plant Sci. 2:48-54.

Cao J, Duan X, McElroy D, Wu R. 1992. Regeneration of herbicide resistant transgenic rice plants following microprojectile-mediated transformation of suspension cells. Plant Cell Rep. 11:586-591.

Cao J, Wang Y-C, Klein TM, Sanford J, Wu R. 1990. Transformation of rice and maize using the biolistic process. In: Lamb CJ, Beachy RN, editors. Plant gene transfer. Proceedings of UCLA Symposium. Park City, Utah (USA): Alan R. Liss Press. p 21-33.

Chan M-T, Chang H-H, Ho S-L, Tang W-F, Yu S-M. 1993. *Agrobacterium*-mediated production of transgenic rice plants expressing a chimeric alpha-amylase promoter/beta-glucuronidase gene. Plant Mol. Biol. 22:491-506.

Cheng Z, Targolli J, Su J, Wu R. 2000. The importance of plant biotechnology in the twenty-first century. In: Wu W-S, Chang S-T, Guan BT, editors. Proceedings Agriculture of the New Century: Managing Bio-resources and Bio-diversity, National Taiwan University Press, Taiwan. p 1-9.

Christou P, Ford TL, Kofron M. 1991. Production of transgenic rice (*Oryza sativa* L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. Bio/Technology 9:957-962.

Cushman JC, Bohnert HJ. 2000. Genomic approaches to plant stress tolerance. Curr. Opin. Plant Biol. 3:117-124.

Delauney AJ, Verma DPS. 1993. Proline biosynthesis and osmoregulation in plants. Plant J. 4:215-223.

Duan X, Li X, Xue Q, Abo-El-Saad M, Xu D, Wu R. 1996. Transgenic rice plants harboring an introduced proteinase inhibitor II gene are insect resistant. Nat. Biotechnol. 14:494-498.

Dure III L. 1981. Developmental biochemistry of cottonseed embryogenesis and germination: changing mRNA populations as shown in *in vitro* and *in vivo* protein synthesis. Biochemistry 20:4162-4168.

Dure III L. 1992. The LEA proteins of higher plants. In: Verma DPS, editor. Control of plant gene expression. Boca Raton, Fla. (USA): CRC Press. p 325-335.

Dure III L. 1993. A repeating 11-mer amino acid motif and plant desiccation. Plant J. 3:363-369.

Dure III L, Crouch M, Harada J, Ho T-HD, Mundy J, Quatrano RS, Thomas T, Sung ZR. 1989. Common amino acid sequence domains among the LEA proteins of higher plants. Plant Mol. Biol. 12:475-486.

Hiei Y, Ohta S, Khomari T, Kumashiro T. 1994. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. 6:271-282.

Holmes-Davis R, Comai L. 1998. Nuclear matrix attachment regions and plant gene expression. Trends Plant Sci. 3:91-96.

Hong B, Barg R, Ho T-HD. 1992. Developmental and organ-specific expression of an ABA- and stress-induced protein in barley. Plant Mol. Biol. 18:663-674.

Hong B, Uknes SJ, Ho T-HD. 1988. Cloning and characterization of a cDNA encoding a mRNA rapidly induced by ABA in barley aleurone layers. Plant Mol. Biol. 11:495-506.

Ingram J, Bartels D. 1996. The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:377-403.

Iyer LM, Kumpatla SP, Chandrasekharan MB, Hall TC. 2000. Transgene silencing in monocots. Plant Mol. Biol. 43:323-346.

Jaglo-Ottosen KR, Gilmour S, Zarka DG, Schabenberger O, Thomashow MF. 1998. *Arabidopsis* CBF1 overexpression induces COR genes and enhances freezing tolerance. Science 280:104-106.

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. 1999. Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat. Biotechnol. 17:287-291.

Kavi Kishor PB, Hong Z, Miao GH, Hu CAA, Verma DPS. 1995. Overexpression of delta 1-pyrroline-5-carboxylase synthetase increases proline production and confers osmotolerance in transgenic plants. Plant Physiol. 108:1387-1394.

Khush GS. 1999. Green revolution: preparing for the 21st century. Genome 42:646-655.

Kumpatla SP, Chandrasekharan MB, Iyer LM, Li G, Hall TC. 1998. Genome intruder scanning and modulation systems and transgene silencing. Trends Plant Sci. 3:97-104.

Mlynarova L, Loonen A, Heldens J, Jansen RC, Keizer P, Stiekema WJ, Nap JP. 1994. Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. Plant Cell 6:417-426.

Moons A, Bauw G, Prinsen E, van Montagu M, Straeten DVD. 1995. Molecular and physiological responses to abscisic acid and salts in roots of salt-sensitive and salt-tolerant indica rice varieties. Plant Physiol. 107:177-186.

Morris CF, Anderenberg RJ, Goldmark PJ, Walker-Simmons MK. 1995. Molecular cloning and expression of abscisic acid responsive genes in embryos of dormant wheat seeds. Plant Physiol. 95:814-821.

Reddy AR, Ramanathan V, Seetharama N, Bajaj S, Wu R. 1999. Genetic improvement of rice for drought and salt tolerance: molecular breeding and transgenic strategies. In: Ito O, O'Toole J, Hardy B, editors. Genetic improvement of rice for water-limited environments. Manila (Philippines): International Rice Research Institute. p 135-155.

Ried JL, Walker-Simmons MK. 1993. Group 3 late embryogenesis abundant proteins in desiccation-tolerant seedlings of wheat (*Triticum aestivum* L.). Plant Physiol. 102:125-131.

Shinozaki K, Yamaguchi-Shinozaki K. 1997. Gene expression and signal transduction in water-stress response. Plant Physiol. 115:327-334.

Spiker S, Thompson WF. 1996. Nuclear matrix attachment regions and transgene expression in plants. Plant Physiol. 110:15-21.

Stacey RAP, Espelund M, Saeboe LS, Hollum K, Jakobsen KS. 1995. Evolution of the group 1 late embryogenesis abundant (LEA) genes: analysis of the LEA B19 gene family in barley. Plant Mol. Biol. 28:1039-1054.

Stam M, Mol JNM, Kooter JM. 1997. The silence of genes in transgenic plants. Ann. Bot. 79:3-12.

Su J, Shen Q, Ho T-HD, Wu R. 1998. Dehydration-stress-regulated transgene expression in stably transformed rice plants. Plant Physiol. 117:913-922.

Vain P, Worland B, Kohli A, Snape JW, Christou P, Allen GC, Thompson WF. 1999. Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J. 18:233-242.

Xu D, Duan X, Wang B, Ho T-HD, Wu R. 1996. Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. Plant Physiol. 110:249-257.

Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. 1982. Living with water stress: evolution of osmolyte system. Science 217:1214-1222.

Zhang JX, Nguyen HT, Blum A. 1999. Genetic analysis of osmotic adjustment in crop plants. J. Exp. Bot. 50:291-302.

Zhang WG, Wu R. 1988. Efficient regeneration of transgenic plants from rice protoplasts and correctly regulated expression of the foreign gene in the plants. Theor. Appl. Genet. 76:835-840.

Zhu B, Su J, Chang M, Verma DPS, Fan Y-L, Wu R. 1998. Overexpression of a delta 1-pyrroline-5-carboxylate synthetase gene and analysis of tolerance to water- and salt-stress in transgenic rice. Plant Sci. 139:41-48.

Notes

Authors' address: Department of Molecular Biology and Genetics, Cornell University, 316 Biotechnology Building, Ithaca, NY 14853, USA.

Acknowledgments: This work was supported by the Rockefeller Foundation (Research Grant RF96001, Allocation 415, and RF98001, Allocation 606) to R.W., and by a Postdoctoral Fellowship to J.T. We thank Dr. R.B. Meager for plasmid pAc1PG-SB for our construction of pAc1PG-CAM.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

High-level expression of C_4 photosynthetic genes in transgenic rice

M. Matsuoka, H. Fukayama, M.S.B. Ku, and M. Miyao

Attempts have been made to transfer C_4 traits to C_3 plants by introducing a chimeric gene construct containing cDNAs for C_4 enzymes under the control of active promoters in C_3 plants. However, the levels of transcripts and proteins in these transformants were far below those in C_4 plants. Our studies have demonstrated that the promoters for maize C_4-specific genes encode phospho*enol*pyruvate carboxylase (PEPC) and pyruvate, orthophosphate dikinase (PPDK), which can drive the high-level expression of a reporter gene in transgenic rice plants in an organ-specific, mesophyll-specific, and light-dependent manner as in maize. These results suggest that the rice plant possesses the regulatory factors necessary for high-level expression of the C_4-specific genes, and imply that the introduction of the intact maize genes would lead to the high-level expression of the C_4 enzymes in rice leaves. The introduction of the intact maize C_4-specific genes containing all exons and introns and its own promoter and terminator sequences led to the high-level expression of the PEPC and PPDK proteins in the leaves of transgenic rice plants. The activities of PEPC and PPDK were 110- and 40-fold higher than those of nontransgenic rice, respectively. The high-level expression of each C_4 enzyme altered metabolism slightly but did not seem to increase the photosynthetic efficiency of transgenic rice leaves.

Land plants, including many important crops such as rice, wheat, and barley, assimilate atmospheric CO_2 through the C_3 photosynthetic pathway and are classified as C_3 plants. In C_3 plants, CO_2 is fixed into a C_3 compound, phosphoglycerate, in mesophyll cells, through carboxylation of ribulose bisphosphate (RuBP) catalyzed by RuBP carboxylase–oxygenase (Rubisco) (Fig. 1). This reaction is inhibited up to 50% by atmospheric O_2 that competes with CO_2 at the active site of Rubisco, and O_2 fixed in this way wastes energy via the process of photorespiration. C_4 plants such as maize and sugarcane have evolved a novel biochemical mechanism to overcome photorespiration. C_4 plants have succeeded in eliminating photorespiration by splitting photosynthetic reactions between two morphologically distinct cell types, the bundle sheath and mesophyll cells (Furbank and Taylor 1995). In a C_4 plant, maize, CO_2 is

initially fixed by phospho*enol*pyruvate carboxylase (PEPC) in the mesophyll cell to form a C_4 compound, oxaloacetate (OAA). OAA is subsequently converted to malate. Malate is then transferred to the bundle sheath cell and then decarboxylated by NADP-dependent malic enzyme (NADP-ME). CO_2 released from this decarboxylation reaction is refixed in the Calvin cycle by Rubisco, as in C_3 plants. To complete the cycle, pyruvate produced by the malate decarboxylation is transferred to the mesophyll chloroplast and phosphorylated to phospho*enol*pyruvate (PEP) via pyruvate, orthophosphate dikinase (PPDK) (Fig. 1).

The introduction of C_4 traits through genetic engineering into rice (C_3 plant) would improve photosynthetic efficiency. One of the major objectives in improving rice yield is to enhance the plant's photosynthetic efficiency. It is necessary to develop protocols to express C_4 enzymes at high levels in C_3 plants. We report here the mechanism of evolution of C_4 photosynthesis genes, and the strategy for the high-level expression of C_4 genes in C_3 plants based on characteristics of C_4 genes.

Evolution of a C_4-type PPDK gene

The genes involved in C_4 photosynthesis have been found in a variety of organisms, not only in C_4 and Crassulacean acid metabolism (CAM) plants, but also in C_3 plants

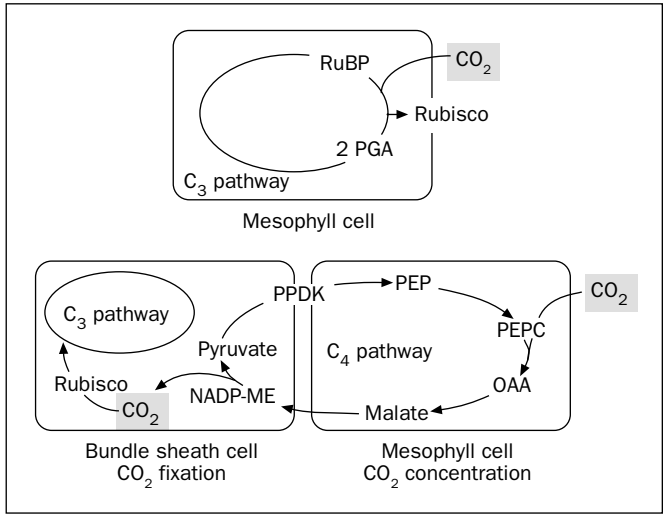


Fig. 1. Comparison of C_3 and C_4 pathways. The C_3 pathway is located in mesophyll cells in C_3 plants, whereas the C_3 cycle is located in bundle sheath cells in C_4 plants. The major function of the C_4 pathway is to concentrate CO_2 in the inner bundle sheath cells where Rubisco is located and thus suppress its oxygenase activity and the associated photorespiration. OAA = oxaloacetate, NADP-ME = NADP-malic enzyme, PEP = phospho*enol*pyruvate, PGA = phosphoglycerate, RuBP = ribulose-1,5-bisphosphate.

(Imaizumi et al 1997). However, the activities of C_4 enzymes such as PPDK are much lower in C_3 plants than in C_4 plants. It is thus important to identify C_4-type genes with high-level expression.

Glackin and Grula (1990) and Sheen (1991) found that the C_4-type PPDK gene in maize is transcribed from two different initiation sites under the control of two different promoters; the first promoter produces the larger transcript encoding the C_4-type enzyme with the chloroplast transit peptide and the second promoter produces the smaller transcript encoding the C_3-type enzyme that is localized in the cytoplasm. The larger transcript is started from exon 1 by the first promoter, whereas the smaller transcript is started from exon 2 by the second promoter (Fig. 2). We isolated and characterized the PPDK gene homologous to the maize C_4-type gene from rice.

The deduced amino acid sequence of rice PPDK is 88% homologous to the maize C_4-type PPDK in the mature peptide region (Imaizumi et al 1997). The rice gene is similar to that of maize except that it has two extra introns (Fig. 2).

The transcription activity of the promoters of the rice and maize PPDK was studied. In transgenic rice, both rice and maize promoters directed GUS expression in photosynthetic organs in a light-dependent manner. However, the maize promoter exhibited a much higher transcriptional activity than the rice promoter did (Matsuoka et al 1993, Nomura et al 2000). The high-level expression of C_4-specific genes in rice is not limited to the PPDK gene, but another C_4 photosynthetic gene for PEPC was also expressed in mesophyll cells in light-grown tissue (Matsuoka et al 1994). These results indicate that rice contains *trans*-acting elements that are required for expression of both the C_4 PPDK gene and C_4 PEPC gene. It is possible that rice contains most of the *trans*-acting elements required for the expression of C_4-specific photosynthetic genes.

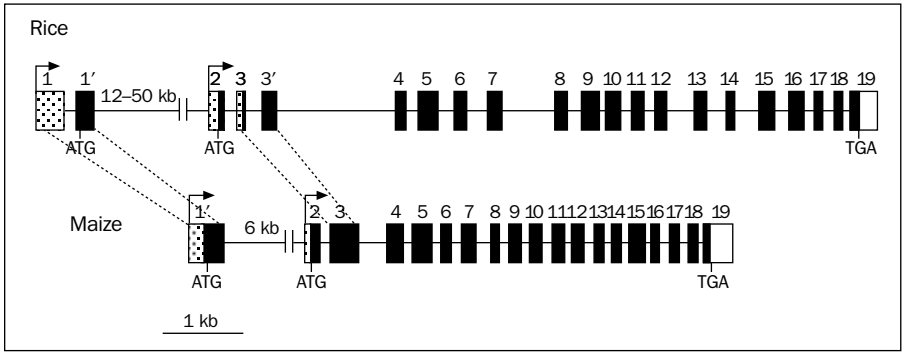


Fig. 2. Exon-intron structure of the genes for PPDK. In both rice and maize genes, transcription starts at the two sites indicated by curved arrows. The common coding regions of the large and small transcripts are represented by filled boxes and the 5′ and 3′ noncoding regions are represented by open boxes. Dotted boxes in rice exon 2 and maize exon 1 represent the region that encodes the transit peptide, and dotted boxes in rice exon 3 and maize exon 2 represent the coding region unique to the small transcripts. The positions of the initiation and termination codons are indicated by ATG and TGA, respectively. The broken lines linking the first and second diagrams indicate the presence of two extra introns in the rice gene. (From Matsuoka et al 2000.)

High-level expression of maize C_4-specific genes in rice

Recombinant DNA techniques were used to increase the activities of C_4 photosynthesis enzymes in C_3 plants (Hudspeth et al 1992, Kogami et al 1994, Gehlen et al 1996). However, the activities of the C_4 enzymes in transgenic C_3 plants were low. In these studies, cDNA clones encoding the C_4 enzymes and 35S promoter from CAMV or Cab promoters of target C_3 plants were used. We have demonstrated that the promoters in the maize PEPC and PPDK genes can drive the high-level expression of a reporter gene in rice in a manner similar to that in maize (Matsuoka et al 1994). The finding that rice possesses the necessary regulatory factors for high-level expression of the C_4-specific genes led us to examine the possibility that intact maize C_4-specific genes, containing all exons and introns and their own promoter and terminator sequences, can produce high amounts of C_4 enzymes in mesophyll cells of C_3 plants.

We transformed rice by introducing the intact gene for the C_4 PEPC enzyme of maize. More than 100 primary transformants were produced. The transgenic plants showed a wide range of PEPC activity. The majority (85%) had activities 2- to 30-fold higher than nontransformed plants, with the remaining 15% showing 30- to 110-fold higher activity, or 1- to 3-fold the maize PEPC activity on a protein basis (Fig. 3). Electrophoretic analysis of total leaf protein extracts revealed the presence of a novel 110-kDa polypeptide, corresponding to that of maize PEPC, in transgenic rice plants. Furthermore, the amount of the polypeptide, as judged from band intensity, was well correlated with the enzyme activity. Immunoblotting analysis with the antimaize PEPC antibody confirmed that the novel 110-kDa polypeptide was the product of the maize C_4-type PEPC gene. These results indicate that the elevated PEPC activities in transgenic rice plants were due to the expression of the maize enzyme. To our knowledge, such high-level expression of a transgene in plants has not been reported previ-

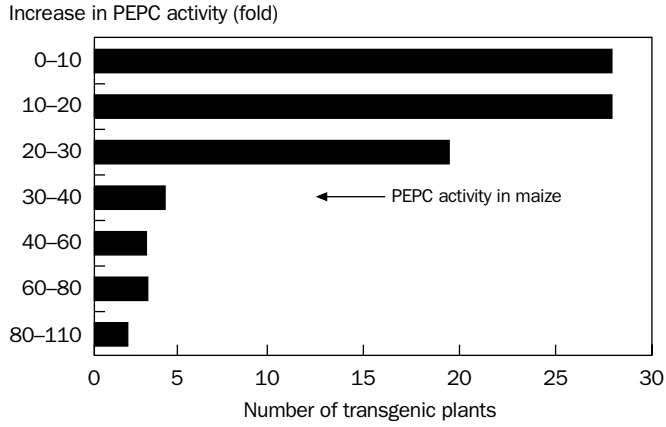


Fig. 3. The activities of PEPC in transgenic rice plants. (Note that PEPC activity of maize genes in transgenic rice is about 35-fold higher than that of nontransformed rice.) (From Matsuoka et al 2000.)

ously. Indeed, earlier attempts to increase PEPC activity in transgenic C_3 plants reported only a 0.5- to 3-fold increase in activity.

Similarly, transgenic rice carrying the PPDK gene from maize also showed 40-fold higher enzymatic activity than that of nontransformed rice. In some transformants, the level of the PPDK protein was extraordinarily high, and it amounted to 35% of total leaf-soluble protein. It was comparable with that of the large subunit of Rubisco, the most abundant protein in leaves of C_3 plants (Fig. 4). Such high-level expression is not solely caused by the transcriptional activity of the maize PPDK gene, since expression of the corresponding cDNA under the control of either the promoter of the maize C_4-specific PPDK gene or the rice Cab promoter increased the activity less than 5-fold. Therefore, it is possible that the presence of exons and introns and/or the terminator sequence in intact genes acts to confer the high-level expression. N-terminal amino acid sequencing indicated that the maize PPDK protein expressed in rice leaves was located exclusively inside the chloroplast. Analysis of transgenic rice plants showed that C_3 plants can be produced with high levels of C_4-specific enzymes.

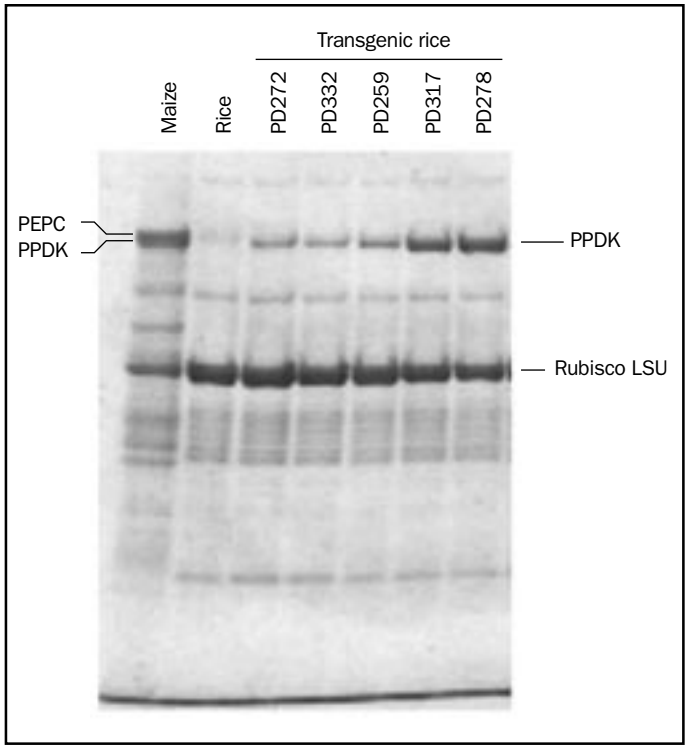


Fig. 4. Polypeptide profiles of leaf-soluble protein in maize, rice, and homozygous lines of transgenic rice carrying the intact maize *Pdk* gene. Protein was stained with Coomassie brilliant blue R-250. LSU = large subunit of Rubisco. (From Matsuoka et al 2000.)

The effect of illumination on PPDK was investigated in rice and maize. PPDK activity was much higher under illumination than in the dark (Fig. 5). However, large differences were observed between maize and rice in the extent of illumination. The light/dark activity ratios were 20 in maize, 6 in nontransformed rice, and 6–9 in transgenic rice. These findings indicate that endogenous rice PDRP can regulate the activity of maize PPDK, although to a limited level, possibly because of the low level of PDRP relative to PPDK in transgenic rice.

Maize and rice both belong to Gramineae. Our recent study indicated that the introduction of the intact gene from other C_4 gramineous plants could also lead to high-level expression of a C_4 enzyme in transgenic rice plants. The activity of the cytosolic form of AspAT in rice leaves was increased to reach 20-fold that of nontransformants by introduction of the intact gene from *Panicum miliaceum* (M. Nomura and M. Matsuoka, unpublished observations). *Panicum miliaceum* is classified in the NAD-ME subtype of C_4 plants and maize in the NADP-ME subtype. Thus, the intact genes from C_4 gramineous plants, irrespective of the C_4 subtype, will probably lead to high-level expression of the C_4 enzymes in mesophyll cells of C_3 gramineous plants.

This strategy, however, seems to have some limitation in that transgenes from plants closely related phylogenetically have to be used to achieve high-level expression of the C_4 enzyme in C_3 plants. The intact maize C_4-specific PEPC gene was not expressed at high levels in tobacco leaves because of incorrect transcription initiation

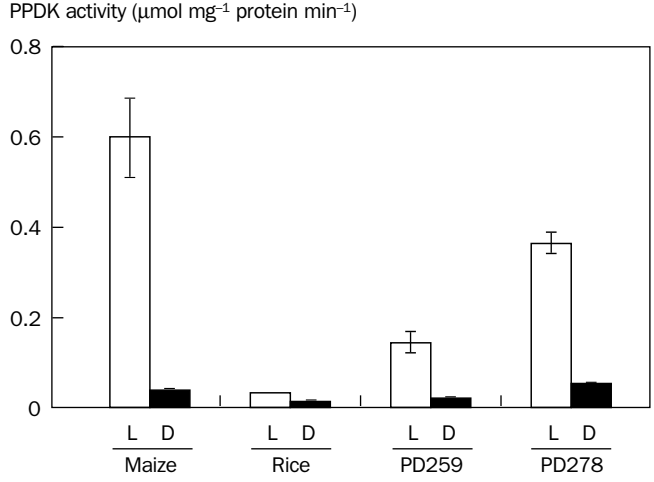


Fig. 5. Light-dependent regulation of PPDK activity in leaves. Maize and rice plants that had been grown in a growth chamber on the day/night cycle for 5 wk were illuminated for 8 h from the start of the daytime and then transferred to the dark for 10 h. The uppermost fully expanded leaves were harvested at the end of illumination (L) and after the subsequent dark incubation (D). The bars represent standard errors of three measurements. PD259 and PD278 are transgenic rice lines.

(Hudspeth et al 1992). Not only incorrect initiation and termination of transcription but also incorrect splicing could occur when genes from monocots are introduced into dicots (see Goodall and Filipowicz 1991). Thus, phylogenetic distance may hamper the expression of genes from C_4 plants in the leaves of C_3 plants.

Unlike the C_4 enzymes located in mesophyll cells of C_4 plants, those located in bundle sheath cells can be expressed at high levels in mesophyll cells of C_3 plants by introducing a chimeric gene containing full-length cDNA for the C_4 enzyme fused to the Cab promoter, which directs mesophyll cell-specific expression in C_3 plants (Sakamoto et al 1991). The expression of the maize C_4-specific NADP-ME cDNA under the control of the rice Cab promoter increased the activity of NADP-ME in rice leaves to 30- or 70-fold that of nontransformants (Takeuchi et al 2000, Tsuchida et al 2000). The level of the NADP-ME protein also increased several percentage points compared with that of total leaf-soluble protein (Takeuchi et al 2000, Tsuchida et al 2000). Such high-level expression was unique to cDNA for the C_4-specific NADP-ME, and the expression of cDNA for the C_3-specific isoform increased the activity only several fold (Lipka et al 1999, Tsuchida et al 2000). Recently, expression of the intact gene for C_4 enzymes located in bundle sheath cells of C_4 plants in C_3 plants has also been addressed. When the intact gene for the mitochondrial AspAT of *Panicum miliaceum*, which is located in bundle sheath cells, was introduced into rice, high AspAT activity was detected in vascular tissues and bundle sheath cells of transgenic rice plants (M. Nomura and M. Matsuoka, unpublished observations). Similar results were observed with the PEP-CK gene from *Zoysia japonica* and β-glucuronidase activity under the control of the PEP-CK promoter, which was selectively detected in vascular tissues and bundle sheath cells (M. Nomura and M. Matsuoka, unpublished information). These results demonstrate that the C_4-specific genes for the bundle sheath cell enzymes can retain their property of cell-specific expression even in a C_3 plant, rice, and therefore suggest that C_3 plants have a regulatory mechanism for gene expression of the bundle sheath cell-specific C_4 genes at their correct site. This fact is interesting in terms of the evolutionary aspect of C_4 plants, but also indicates that the strategy to introduce intact C_4-specific genes is not applicable to building the C_4 pathway solely in mesophyll cells of C_3 plants.

Physiological effects of overproduction of C_4 enzymes in C_3 plants

To examine the effects of high-level expression of maize PEPC on photosynthetic characteristics, O_2 sensitivity of photosynthetic CO_2 assimilation of transgenic rice plants was investigated by measuring CO_2 assimilation rates. O_2 inhibition of photosynthetic CO_2 assimilation was negatively correlated to the activity of PEPC in transgenic rice plants. This reduced O_2 inhibition was explained by direct CO_2 fixation by the maize PEPC in transgenic rice (Ku et al 1999). However, photosynthetic CO_2 assimilation of transgenic rice decreased with increasing PEPC activity at both O_2 concentrations (21% and 2%). In addition, the slope of the regression line was steeper in 2% O_2 than in 21% O_2. Thus, the observed reduction in O_2 inhibition might result from decreased CO_2 assimilation in 2% O_2 in transgenic rice. Since the photosynthetic

rate in 2% O_2 can be limited by Pi regeneration capacity (Sharkey 1985), high-level expression of PEPC might affect Pi regeneration capacity in transgenic rice.

Transgenic rice plants overexpressing the maize PPDK showed a higher photosynthetic rate (up to 35%) than untransformed plants. This increased photosynthetic capacity is at least in part due to an enhanced stomatal conductance and a higher internal CO_2 concentration. We are now crossing these transgenic rice plants to integrate both PEPC and PPDK genes into the same transgenic rice plants.

References

Furbank RT, Taylor WC. 1995. Regulation of photosynthesis in C_3 and C_4 plants: a molecular approach. Plant Cell 7:797-807.

Gehlen J, Panstruga R, Smets H, Merkelbach S, Kleines M, Porsch P, Fladung M, Becker I, Rademacher T, Hausler RE, Hirsch HJ. 1996. Effects of altered phospho*enol*pyruvate carboxylase activities on transgenic C_3 plant *Solanum tuberosum*. Plant Mol. Biol. 32:831-848.

Glackin CA, Grula JW. 1990. Organ-specific transcripts of different size and abundance derive from the same pyruvate, orthophosphate dikinase gene in maize. Proc. Natl. Acad. Sci. USA 87:3004-3008.

Goodall GJ, Filipowicz W. 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J. 10:2635-2644.

Hudspeth RL, Grula JW, Dai Z, Edwards GE, Ku MSB. 1992. Expression of maize phospho-*enol*pyruvate carboxylase in transgenic tobacco. Plant Physiol. 98:458-464.

Imaizumi N, Ku MSB, Ishihara K, Samejima M, Kaneko S, Matsuoka M. 1997. Characterization of the gene for pyruvate, orthophosphate dikinase from rice, a C_3 plant, and a comparison of structure and expression between C_3 and C_4 genes for this protein. Plant Mol. Biol. 34:701-716.

Kogami H, Shono M, Koike T, Yanagisawa S, Izui K, Sentoku N, Tanifuji S, Uchimiya H, Toki S. 1994. Molecular and physiological evaluation of transgenic tobacco plants expressing a maize phospho*enol*pyruvate carboxylase gene under the control of the cauliflower mosaic virus 35S promoter. Transgenic Res. 3:287-296.

Lipka V, Häusler RE, Rademacher T, Li J, Hirsch H-J et al. 1999. *Solanum tuberosum* double transgenic expressing phospho*enol*pyruvate carboxylase and NADP-malic enzyme display reduced electron requirement for CO_2 fixation. Plant Sci. 144:93-105.

Matsuoka M, Tada Y, Fujimura T, Kano-Murakami Y. 1993. Tissue-specific light-regulated expression directed by the promoter of a C_4 gene, maize pyruvate, orthophosphate dikinase, in a C_3 plant, rice. Proc. Natl. Acad. Sci. USA 90:9586-9590.

Matsuoka M, Kyozuka J, Shimamoto K, Kano-Murakami Y. 1994. The promoters of two carboxylases in a C_4 plant (maize) direct cell-specific light-regulated expression in a C_3 plant (rice). Plant J. 6:311-319.

Matsuoka M, Fukayama H, Tsuchida H, Nomura M, Agarie S, Ku MSB, Miyao M. 2000. How to express some C_4 photosynthesis genes at high levels in rice. In: Sheehy JE, Mitchell PL, Hardy B, editors. Redesigning rice photosynthesis to increase yield. Proceedings of the Workshop on The Quest to Reduce Hunger: Redesigning Rice Photosynthesis, 30 Nov.-3 Dec. 1999, Los Baños, Philippines. Makati City (Philippines): International Rice Research Institute and Amsterdam (The Netherlands): Elsevier Science B.V. p 167-175.

Nomura M, Sentoku N, Nishimura A, Lin J-H, Honda C, Taniguchi M, Ishida Y, Ohta S, Komari T, Miyao-Tokutomi M, Kano-Murakami Y, Tajima S, Ku MSB, Matsuoka M. 2000. The evolution of C_4 plants: acquisition of *cis*-regulatory sequences in the promoter of C_4-type pyruvate, orthophosphate dikinase gene. Plant J. 22:211-221.

Sakamoto M, Sanada Y, Tagiri A, Murakami T, Ohashi Y, Matsuoka M. 1991. Structure and characterization of a gene for light-harvesting Chl a/b binding protein from rice. Plant Cell Physiol. 32:385-393.

Sharkey TD. 1985. O_2-insensitive photosynthesis in C_3 plants: its occurrence and a possible explanation. Plant Physiol. 78:71-75.

Sheen J. 1991. Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. Plant Cell 3:225-245.

Takeuchi Y, Akagi H, Kamasawa N, Osumi M, Honda H. 2000. Aberrant chloroplasts in transgenic rice plants expressing a high level of maize NADP-dependent malic enzyme. Planta. 211:265-274.

Tsuchida H, Tamai T, Fukayama H, Agarie S, Nomura M, Onodera H, Ono K, Nishizawa Y, Lee B-H, Hirose S, Toki S, Ku MSB, Matsuoka M, Miyao M. 2000. High-level expression of C_4-specific NADP-malic enzyme in leaves impairs photoautotrophic growth of C_3 rice plant. Plant Cell Physiol. 42:138-145.

Notes

Authors' addresses: M. Matsuoka, BioScience Center, Nagoya University, Chikusa, Nagoya 464-8601, Japan, E-mail: j45751a@nucc.cc.nagoya-u.ac.jp; H. Fukayama, M. Miyao, National Institute of Agrobiological Resources, Tsukuba 305-8602, Japan, E-mail: fukayama@abr.affrc.go.jp; M.S.B. Ku, Botany Department, Washington State University, Pullman, Washington 99164-4238, USA.

Acknowledgment: This work was supported in part by a PROBRAIN grant from the Bio-Oriented Technology Research Advancement Institution (BRAIN) of Japan.

Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Transgene integration, organization, and expression in cereals

P. Christou, R.M. Twyman, Xiangdong Fu, E. Wegel, A. Kohli, and E. Stoger

As in other direct DNA transfer systems, transgenes delivered by particle bombardment integrate into the cereal genome by illegitimate recombination assisted by short regions of homology. Exogenous plasmid DNA tends to concatemerize prior to integration, resulting in tandemly arranged contiguous copies. However, recent research has shown that such concatemers may positively influence the integration of further transgenes nearby, perhaps by recruiting DNA repair complexes to the site of the original break. Transgenic loci therefore tend to comprise clusters of contiguous copies interspersed with short regions of genomic DNA. The analysis of transgenic wheat chromosomes by fluorescence *in situ* hybridization (FISH) indicates a further level of organization, where several transgene clusters integrate in the same region of the chromosome arm, but individual clusters produce separable FISH signals at metaphase, suggesting they are interspersed by large segments of genomic DNA. A segregating transgene locus therefore has a hierarchical structure, showing up to three levels of organization.

The analysis of transgene integration shows that recombination plays a strong role in determining integrated transgene structure. Particularly, certain sites in the transformation plasmid can provide hotspots for recombination, leading to particular types of transgene rearrangement. Since the failure of transgene expression often reflects such rearrangements, the modification of transformation vectors to eliminate troublesome sequences should improve the efficiency of transformation. The "clean DNA" system takes this strategy to its logical extreme by removing all unnecessary backbone elements and using just minimal linear cassettes (promoter, coding region, terminator) for transformation. Such experiments consistently generate transgenic plants with simple integration patterns, lower transgene copy numbers, fewer rearrangements, and stable transgene expression. The clean DNA system also allows direct transformation with multiple genes, but the integration patterns remain simple and silencing occurs very rarely.

Silencing often correlates with DNA methylation, and it is of particular interest to identify sequences in plant transgenes that induce methylation and result in unstable expression. There is evidence that methylation can be induced by interactions between homologous transgene copies (or between a transgene and homologous endogene) or may reflect genomic position ef-

fects. Recent data suggest, however, that more complex underlying processes may be involved. In a line of plants containing a single-copy three-gene transgenic locus, different forms of silencing could occur *de novo* and affect individual transgenes, even though the three were separated by only a few nucleotides. Remarkably, two completely different modes of silencing, associated with different methylation patterns, could be established in adjacent transgenes. Transposon mutagenesis is one of the most powerful tools available to gain an understanding of factors underlying transgene expression and stability. We have a population of more than 500 independent indica rice transformants containing the maize *Ac* transposon, representing a total of more than 1,500 plants. We have observed active excision and re-insertion of the *Ac* element in this population, which can now be scaled up to look for genes affecting genomic methylation and transgene expression.

The genetic manipulation of cereals is a key objective of both fundamental and applied research. This reflects the importance of cereals as a major food crop as well as the emergence of rice in particular as a model in functional genomics. It is therefore essential to gain a fundamental understanding of the mechanisms of transgene integration and how transgene structure and organization affect the stability of expression. Recent research carried out in our laboratory has shown that the mechanism of transgene integration at the molecular level is similar in cereals and all other plants (Kohli et al 1999). However, the use of direct DNA transfer methods such as particle bombardment to produce transgenic cereals appears to generate a unique higher-order transgene organization, which may have specific effects on transgene expression. The purpose of this review is to summarize our current knowledge of transgene integration and organization in cereals, particularly rice and wheat, and how this information has been used to improve the efficiency with which useful transgenic plants are generated.

DNA transfer to cereals

Since most monocots, including all cereals, were thought to be outside the natural host range of *Agrobacterium tumefaciens*, early attempts to transform cereals focused on direct DNA transfer methods. Originally, this involved the transformation of protoplasts, a technically demanding procedure requiring long periods of tissue culture (Krens et al 1982, Negrutiu et al 1987, Fromm et al 1987). The advent of particle bombardment, which is simple both conceptually and in practice, radically streamlined the production of transgenic cereals (Klein et al 1987, 1988, Christou et al 1988). Although advances in *Agrobacterium* methodology now permit the reliable transformation of cereals (Komari et al 1998), particle bombardment has emerged as the most robust method, since it is in principle unrestricted by either species or genotype barriers, relying on physical parameters rather than genetic ones. DNA can be introduced into a range of organized and easily regenerable tissues, and transgenic plants pro-

duced by organogenesis, somatic embryogenesis, or the germination of transformed seeds (Twyman et al, n.d.).

Another important advantage of particle bombardment is that it allows the stable and heritable introduction of many different genes at once using different plasmids, as these tend to cluster at a single locus. Conversely, transformation with more than two genes using conventional *Agrobacterium* systems currently requires the cointegration of all the genes in the same T-DNA. Chen et al (1998) reported the biolistic cotransformation of rice with 14 separate plasmids containing various genes and showed data confirming the integration of at least 13 of the plasmids in one plant. Cotransformation has also been used to introduce up to four agronomically important genes into rice, producing plants showing resistance to a spectrum of insect pests and plants with resistance pyramided against individual pests (Bano-Maqbool and Christou 1999, Tang et al 1999). Cointegration prevents different transgenes from segregating at meiosis. This is very important in breeding programs in which plants carry two or more transgenes required to generate a single protein, for example, in plants expressing recombinant human antibodies.

Transgene organization and expression

The stable expression of transgenes over multiple generations is essential if transgenic plants are to be of long-term value to producers and consumers. Transgene expression is influenced by several factors, many of which in some way reflect the structure of the transgenic locus. Such factors include the integrity of individual transgenes, the number of transgene copies, and their organization within the transgene locus. By investigating the relationship between transgene structure/organization and expression, it may be possible to determine the factors that are most important for stable transgene expression and to improve the design of transformation constructs to eliminate nonproductive transgene structures. This has been demonstrated, for example, by the success of the clean DNA transformation system discussed later in this chapter.

Three-level hierarchy of transgene organization in cereals

Transgene structure/organization can be investigated at different levels of resolution and this has provided evidence for a hierarchical organization. Polymerase chain reaction (PCR) amplification and sequencing provide the finest resolution and show that transgenic loci in rice often contain contiguous copies of the transforming plasmid (Kohli et al 1998, 1999). These may be arranged either head-to-tail or as inverted repeats and are thought to form prior to integration through the concatemerization (end-to-end joining) of exogenous DNA fragments.

Southern blot analysis is often used to characterize transgenic plants, but, as well as confirming transgene integration, it also demonstrates a second level of transgene organization (Kohli et al 1998, 1999, Fu et al 2000b; Fig. 1). Restriction enzymes that cut once in the transforming plasmid should generate different hybridizing bands for each transgene copy and thus provide an estimate of copy number. Restriction enzymes that cut either side of the transgene should release a cassette of diagnostic size

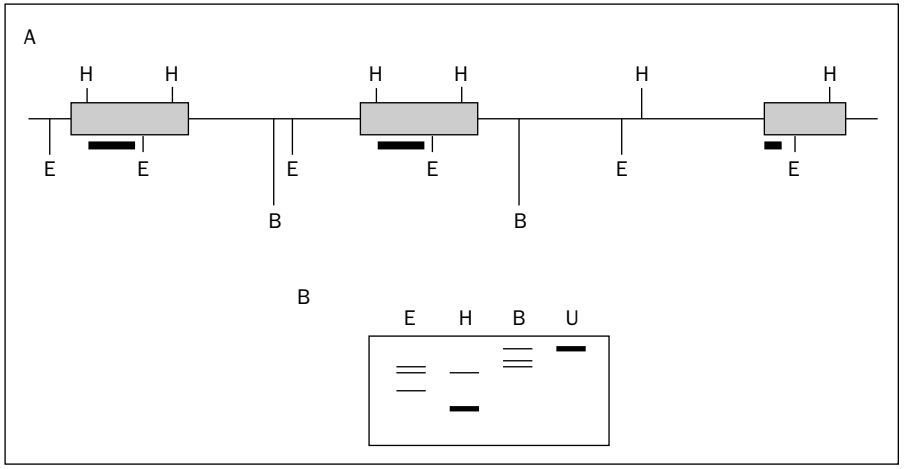


Fig. 1. How Southern blots can be used to characterize transgene organization. (A) A hypothetical transgenic locus comprising two intact transgenes and one truncated transgene, each separated by a short stretch of genomic DNA. Each intact transgene carries one site for the restriction enzyme *Eco*R1 (E), two sites for the enzyme *Hin*dIII (H), and no sites for the enzyme *Bam*H1 (B). (B) Typical Southern blot results. The *Eco*R1 digest reveals three bands. This indicates that there are three copies of the transgene since each hybridizing fragment is delineated by one *Eco*R1 site within the transgene and one randomly placed in the surrounding genomic DNA. The *Hin*dIII digest should generate a specific diagnostic hybridizing fragment of defined length, delineated by the two sites within the transgene. Additional fragments shorter or longer than expected indicate transgene rearrangements, in this case a truncation. The aberrant fragment is longer than expected because the 5′ *Hin*dIII site has been deleted in the truncation, so the fragment is formed by the next available *Hin*dIII site in the genomic DNA. The *Bam*H1 digest should generate a single high-molecular-weight fragment if there are no *Bam*H1 sites between the transgenes, often similar to the fragment seen in U (undigested DNA). If such sites do exist, individual fragments would be generated as shown. Experiments with several such "noncutter" enzymes reveal the existence of genomic DNA between transgene copies.

that can be used to confirm transgene integrity at the gross level (i.e., point mutations will not be identified). Restriction enzymes that do not cut in the transformation plasmid should release the transgenic locus as a single high-molecular-weight fragment regardless of the number of copies, as dictated by the positions of restriction sites in the surrounding genomic DNA. However, such digests often produce two or more hybridizing bands, providing evidence for the presence of variable-length stretches of genomic DNA between transgenes or contiguous transgene arrays. Such genomic stretches can be isolated by long PCR, indicating that they are tens rather than hundreds of kilobase pairs in length. This indicates that individual transgene copies and contiguous arrays are interspersed with genomic DNA and organized into transgene clusters.

In wheat, we began an investigation of the distribution of transgene loci to determine whether there was any bias to integration sites in terms of specific chromosomes or chromosome regions. This series of experiments was carried out using fluorescence *in situ* hybridization (FISH) and provided some surprising findings concern-

ing transgene organization. FISH analysis of wheat lines shown to have a single segregating transgenic locus revealed that each locus could comprise several separable signals, suggesting that individual transgenes and transgene clusters could be interspersed with sections of genomic DNA visible at the cytogenetic level, that is, in the order of megabase pairs (Abranches et al, n.d.). Analysis of rice and wheat transgenic loci has therefore demonstrated a three-tier organization of transgenes (Fig. 2), starting with individual copies and contiguous arrays, organized at a second level in clusters containing short regions of genomic DNA, which are in turn distributed in a local region of the chromosome and interspersed with longer regions of DNA, so that they generate discrete FISH signals at the cytogenetic level. Each of these levels of organization provides clues as to the mechanisms involved in transgene integration by direct DNA transfer, as discussed below.

Transgene expression

Transgene organization reflects three properties of a transgenic locus: the transgene copy number, the spatial arrangement of transgenes within the locus, and their structural integrity. In animal systems, evidence has accumulated that all three of these properties can profoundly affect transgene expression. For example, where two transgenes are arranged as an inverted repeat, the copies can form ectopic pairs or

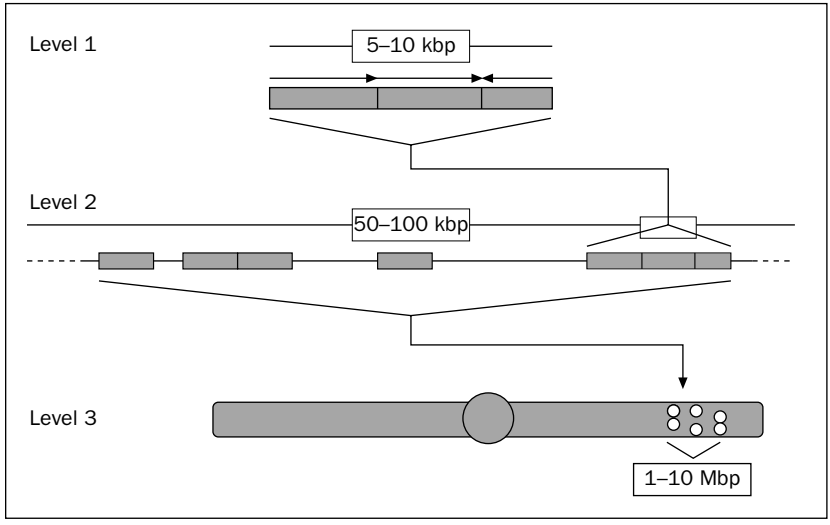


Fig. 2. Three-tier hierarchical transgene organization in cereals. Level 1 comprises individual transgenes and contiguous transgene copies (either tandem or inverted repeats, and containing intact and/or truncated copies), which concatemerize prior to integration. Level 2 comprises transgene clusters, i.e., groups of individual transgenes and contiguous transgene copies, interspersed by short regions of genomic DNA that can be isolated by polymerase chain reaction. Level 3 comprises groups of separable fluorescence *in situ* hybridization (FISH) signals, each corresponding to a transgene cluster, separated by megabases of genomic DNA. The groups of FISH signals behave as a single locus in genetic segregation analysis.

transcription through one transgene and into the next can generate antisense RNA (Twyman and Whitelaw 2000). Both these effects can lead to transgene silencing. In plants, the influence of the spatial arrangement of transgenes has not received much attention. A great deal of evidence suggests that multiple transgene copies provoke silencing, either by ectopic pairing in *cis* or in *trans* or by posttranscriptional mechanisms involving aberrant RNA species (reviewed by Gallie 1998, Grant 1999). All such studies, however, have been carried out in model dicot plants such as petunia, tobacco, and *Arabidopsis*. Work carried out in our laboratory has provided no evidence for a link between transgene copy number and expression level. We have generated hundreds of cereal plants ranging from those with single-copy transgenes to those carrying tens of copies, and there appears to be no correlation between copy number and silencing. Therefore, it appears that transgene integrity and spatial arrangement are most important when it comes to the effect of transgene organization on expression, and that even when intact single-copy transgenes are analyzed, there can still be pronounced and inexplicable *de novo* silencing effects at both the transcriptional and posttranscriptional levels even in plants with intact, single-copy transgenes (Fu et al 2000a).

Mechanisms of transgene integration

Level 1: integration at the molecular level

Studies in dicots have identified a possible link between double-strand break repair (DSBR) and transgene integration, both requiring illegitimate recombination. This form of recombination is a ubiquitous repair mechanism in eukaryotes, involving the covalent joining of sequences with either microhomology or no homology (Tsukamoto and Ikeda 1998). Illegitimate recombination junctions are often characterized by one or more of the following properties: (1) deletion of nucleotides at one or both of the recombining ends; (2) microhomology between the recombining ends, involving up to nine nucleotides; and (3) the presence of short stretches of additional DNA, known as filler DNA, at the junction. Equivalent studies in cereals have been carried out only recently. Takano et al (1997) used calcium phosphate-mediated transformation to study plasmid-genomic junctions in rice protoplasts, whereas, in our laboratory, we have carried out a series of studies involving plasmid-genomic and plasmid-plasmid junctions in transgenic rice and maize plants generated by particle bombardment. The sequences of novel recombination junctions suggest that the transforming plasmid had undergone rearrangements involving illegitimate recombination in a fashion similar to that of monocot and dicot species transformed by other methods. Therefore, at the molecular level, transgene integration following direct DNA or *Agrobacterium*-mediated T-DNA transfer is likely to involve the same components of the DNA-repair machinery.

Illegitimate recombination is the predominant mechanism of DSBR in higher eukaryotes, probably because the large genome size prevents effective homology searching and because the higher-order chromatin structure would be expected to hold broken DNA ends in close proximity. However, where unorthodox substrates are involved, illegitimate recombination can lead to large-scale genome rearrange-

ments and the integration of exogenous DNA (Salomon and Puchta 1998). In both *Agrobacterium*-mediated DNA delivery and the direct physical DNA transfer into plant cells, a wound response is elicited, which involves the activation of nucleases and DNA-repair enzymes to maintain the integrity of the host genome (Hunold et al 1994). Any exogenous DNA entering the cell is therefore simultaneously exposed to breakdown and repair enzymes. As a consequence, much of the exogenous DNA may be degraded but some may be used as a substrate for DNA repair, resulting in its rearrangement and/or incorporation into the genome (Matsumoto et al 1990, Gheysen et al 1991, Mayerhofer et al 1991, Takano et al 1997, Gorbunova and Levy 1997). Intact plasmids are relatively poor substrates for illegitimate recombination, so DNA ends may be provided by the activity of nucleases, which create single-strand nicks or double-strand breaks in the DNA.

The plasmid-plasmid recombination junctions we characterized in rice and maize plants generated by particle bombardment were similar to those found in tobacco following polyethylene glycol-mediated (Gorbunova and Levy 1997) or *Agrobacterium*-mediated transformation (De Groot et al 1994, Salomon and Puchta 1998). The analysis revealed typical features of illegitimate recombination, including (1) short patches of 4–8 homologous nucleotides at the junctions, (2) filler DNA, and (3) deletions at one of the recombining ends. Junctions were predominantly formed by microhomology-dependent illegitimate recombination. The favored mechanism for this type of reaction involves interaction between DNA ends with short single-stranded tails, the annealing of complementary tails, and repair synthesis over the remaining gaps. In addition, weak homology was often seen surrounding the junction site—this may have stabilized the initial interaction between the recombination substrates.

Illegitimate recombination junctions studied in animals suggest that homology with more than four nucleotides is rare, whereas the majority of the junctions we characterized in rice involved homology with more than four nucleotides. We found that junction sites preferentially comprised purine-rich tracts. In only two cases did junction formation appear not to involve the above mechanism, that is, there was no homology between the recombining partners and recombination may have occurred simply by blunt-end ligation of two DNA molecules. Gorbunova and Levy (1997) found only three out of 19 tobacco protoplast-derived clones in which the blunt or cohesive ends of their restriction-digested substrate joined without any degradation. Salomon and Puchta (1998) found two blunt-end ligation events among the 28 junctions they characterized in tobacco.

In one line, the junction contained an insertion of a 23-bp fragment of filler DNA. The presence of direct repeats (5′ TCCGG 3′) flanking this insert suggested one of two possible mechanisms. Illegitimate recombination may have occurred between the two ends and may have involved the synthesis of untemplated nucleotides. The direct repeats may represent short tails of imperfect complementarity responsible for the initiation of recombination. Alternatively, the insertion may represent a transposition event. The presence of staggered breaks in a target DNA molecule (in this case represented by two plasmids) may have acted as a substrate for the transposase or integrase encoded by an endogenous plant transposable element, resulting in the in-

sertion of an adventitious DNA fragment. Such events are thought to be responsible for the generation of processed pseudogenes, which make up a significant component of many higher eukaryotic genomes (Berg and Howe 1989). The direct repeats would in this case represent repair synthesis over the staggered break, generating the "target site duplication" characteristic of transposition events in all species, including the integration of T-DNA into the plant genome (Tinland 1996). Gorbunova and Levy (1997) found insertions ranging from 2 bp to 1.2 kbp in nearly 30% of the plasmid junctions they studied. This so-called "filler DNA" was sometimes genomic in origin, sometimes it appeared to derive from the transforming plasmid, and in other cases the origin was uncertain. The similarities between recombination junctions generated by particle bombardment in our study and those generated by alternative transformation methods in both monocots and dicots in other studies strongly suggest that the underlying mechanisms controlling plasmid rearrangement and transgene integration in plants are likely to be the same. The evidence we have provided suggests that several categories of illegitimate recombination are likely to be involved in this process, although microhomology-mediated recombination predominates (Kohli et al 1999; Fig. 3).

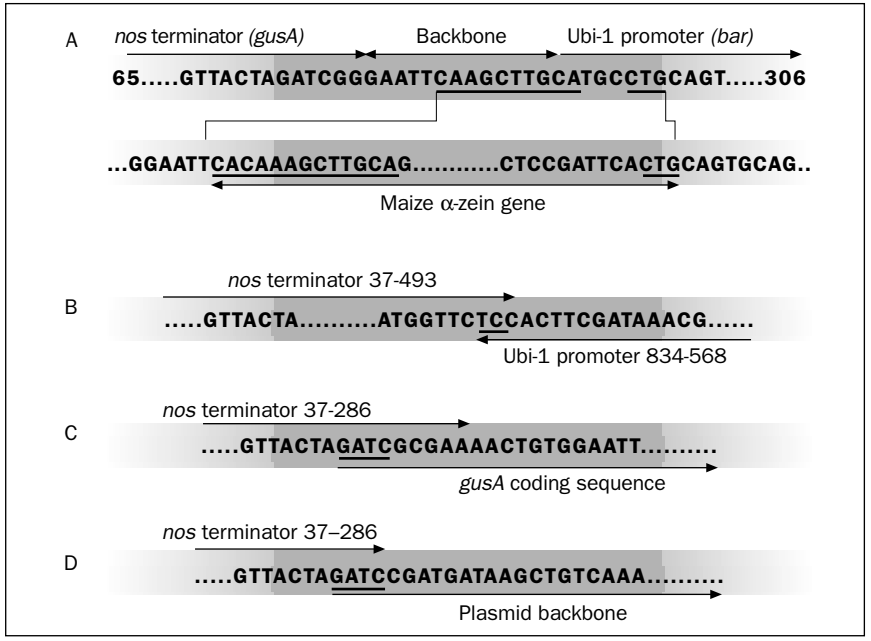


Fig. 3. Microhomology-mediated recombination in transgenic maize. The figure shows transgene junction sequences from transgenic loci in four transgenic maize plants. In each case, short regions of homology (3–10 nt) are involved in the recombination process and these regions of common sequence between the two recombining partners are shown as underlined. This can result in (A) the insertion of genomic DNA into transgene (B) "head-to-head" junctions or (C, D) "head-to-tail" junctions.

Level 2: activation of local repair complexes

Most of the transgenic plants we have analyzed have only a single transgenic locus as defined by genetic segregation analysis. However, as discussed above, individual transgenes and contiguous transgene arrays have been shown to be arranged in a local cluster, also containing significant stretches of genomic DNA. This is not the filler DNA described above, which is usually only a few nucleotides in length.

The presence of significant stretches of genomic DNA within a transgene cluster suggests that there is a pronounced tendency for fragments of exogenous DNA to integrate at one site. Since transgenic loci appear to arise randomly, probably reflecting naturally occurring DNA breaks, we proposed that a primary integration event promoted secondary integration events nearby (Kohli et al 1998). This two-phase mechanism could reflect the recruitment of DNA-repair complexes to the original integration site and the consequent introduction of additional double-strand DNA breaks in the local region. This mechanism would result in a concentrated core of DNA repair and transgene integration, which would likely cause a significant amount of transgene rearrangement as well as the loss of stretches of genomic DNA. Indeed, the deletion of genomic DNA around transgene integration sites is a well-known phenomenon in both transgenic plants and animals.

Level 3: integration at the chromatin level

The third level of transgene organization involves the dispersion of transgene copies and clusters such that individual signals can be visualized by FISH on metaphase chromosomes. This demonstrates that transgene clusters may be separated by megabases of DNA. How is this different from the level 2 organization discussed above? Could it be that repair complexes are being activated over a greater area to generate transgene sites separated by larger intervening regions of genomic DNA? Two pieces of evidence suggest not. First, the intervening regions at level 2 organization can be isolated by long PCR, a reasonable distance to be spanned by locally recruited repair complexes. However, repair complexes would unlikely be recruited to sites megabases away from the initial integration site. Second, the analysis of the same wheat nuclei at interphase shows the remarkable phenomenon that the FISH signals that are separated at metaphase are brought back together at interphase (Abranches et al, n.d.). Since this phenomenon occurs in multiple nuclei from somatic tissues of a given transgenic plant, and in progeny thereof, it confirms that the interphase chromatin is highly and reproducibly organized in the nucleus. We have put forward three models to explain this (Abranches et al, n.d.). First, it is possible that the homologous DNA sequences are associating in *trans*. Second, copies of the same promoter may be recruited to a common transcription factory in the nucleus. Our third model, which provokes the most thought with respect to the mechanisms of transgene integration, is that the FISH signals permit visualization of the three-dimensional configuration of the nucleus at the moment of transformation. It has been shown that cells successfully transformed by particle bombardment usually have the metal particle lodged in the nucleus. Therefore, the metal particle may cause localized damage to DNA in a particular region of the nucleus. The way that chroma-

tin is folded into loops may result in damage to several discrete sites that are separated in *cis* by hundreds of kilobase pairs of DNA but are close together in *trans*.

Transgene rearrangements and the role of recombination "hotspots"

The importance of transgene rearrangements

We have found that transgene rearrangements have profound effects on transgene expression. Transgene rearrangement is a pitfall of all direct DNA transfer methods, but is perhaps more acute in particle bombardment because the forces involved may cause more DNA fragmentation than other methods and because bombarded plant cells may be induced to produce DNA degradation and repair enzymes in response to their injury (Hunold et al 1994). Two important recent discoveries in this area are that transgene rearrangements can be subtle, yet still contribute to loss of activity, and that rearrangements often involve so-called recombination hotspots in the transformation construct.

Detection of subtle rearrangements

We generated several transgenic maize plants carrying three transgenes: *hpt*, *bar*, and *gusA*. We examined all plants by Southern blot analysis and found that even those with apparently intact transgene cassettes could be inactive. To investigate whether transgene inactivity reflected mutation or epigenetic phenomena, we used long PCR to characterize the transgene loci (Mehlo et al, n.d.). We found that, in most cases where intact transgenes gave no expression, there were subtle rearrangements that could not be detected by Southern blot analysis. For example, some lines showed small internal deletions, point mutations, or external erosions. These had the effect of removing PCR primer sites so that amplification of the entire expression cassette failed, whereas typical PCR testing of a small region within the transgene showed no abnormalities. From these experiments, we concluded that copy number/expression correlations should be interpreted with caution since undetected subtle mutations could account for many instances of transgene inactivity.

Recombination hotspots

Our analysis of rice plants concentrated on the structure of plasmid-plasmid junctions. As discussed above, this showed that the formation of contiguous plasmid arrays and the integration of exogenous DNA both involved predominantly microhomology-mediated recombination. Remarkably, however, more than 50% of the junctions we analyzed involved the same region of the plasmid, a palindromic element from within the CaMV 35S promoter (Kohli et al 1999). The same region in the full-length CaMV RNA is known to promote recombination between the genomes of different strains of the virus in plants (Gal et al 1992, Swoboda et al 1993). It has been suggested that virus-specific *cis*-acting elements and virus-encoded enzymes might play a role in determining the nature of recombination events (Schoelz et al 1993). Even in the absence of these factors, however, the CaMV 35S DNA fragment underwent recombination at a high frequency, an observation that

strongly suggested that plant factors can direct recombination events by recognizing and using the highly recombinogenic viral sequences. We established that, at the 3′ end of the CaMV 35S promoter, an imperfect palindrome of 19 bp, in conjunction with specific flanking sequences derived from the transforming plasmid, acts as a hotspot for recombination.

Clean DNA transformation

In all transformation methods, exogenous DNA tends to undergo rearrangement and recombination events leading to the integration of multiple fragmented, chimeric, and rearranged transgene copies. Such clusters are prone to internal recombination and may therefore be unstable as well as promote transgene silencing.

A major breakthrough has been achieved in the last year in the production of transgenic plants without integrated vector backbone sequences. Since the backbone is a major source of recombinogenic elements, such as the origin of replication, a logical approach is to remove these sequences from the transforming vector. The "clean DNA" system (Fu et al 2000b) takes this strategy to its logical extreme by removing all unnecessary backbone elements and using just minimal linear cassettes for transformation: the promoter, open reading frame, and terminator. Transgenic rice plants regenerated from callus bombarded with such cassettes showed very simple integration patterns, with one or a few hybridizing bands on Southern blots. Comparative transformation experiments using whole-plasmid DNA resulted in much more complex integration patterns, with multiple bands of different sizes (Fig. 4). Furthermore, release of a diagnostic fragment from the transgene showed that fewer rearrangement events had occurred in the plants carrying the linear cassettes. The progeny of these plants were examined for transgene expression and there were no instances of silencing among the transgenes studied. Conversely, whole-plasmid transformation by particle bombardment can lead to silencing in up to 20% of the resulting transgenic plants. The simplicity of clean DNA transformation is probably achieved by reducing the overall recombinogenicity (propensity to undergo recombination) of the construct and thereby simplifying the structural hierarchy of the locus. This work showed that minimal linear cassettes promoted the same efficiency of transformation as normal plasmids. Similarly, multiple linear fragments showed the same efficiency of cotransformation as multiple plasmids, so the clean DNA strategy is a promising technique for the generation of transgenic plants carrying multiple transgenes.

Transgene silencing and DNA methylation

Silencing often correlates with DNA methylation, and it is of particular interest to find sequences in plant transgenes that induce methylation and result in unstable expression. There is evidence that methylation can be induced by interactions between homologous transgene copies (or between a transgene and homologous endogene) or may reflect genomic position effects (Gallie 1998, Grant 1999). Recent data suggest, however, that more complex underlying processes may be involved. We addressed

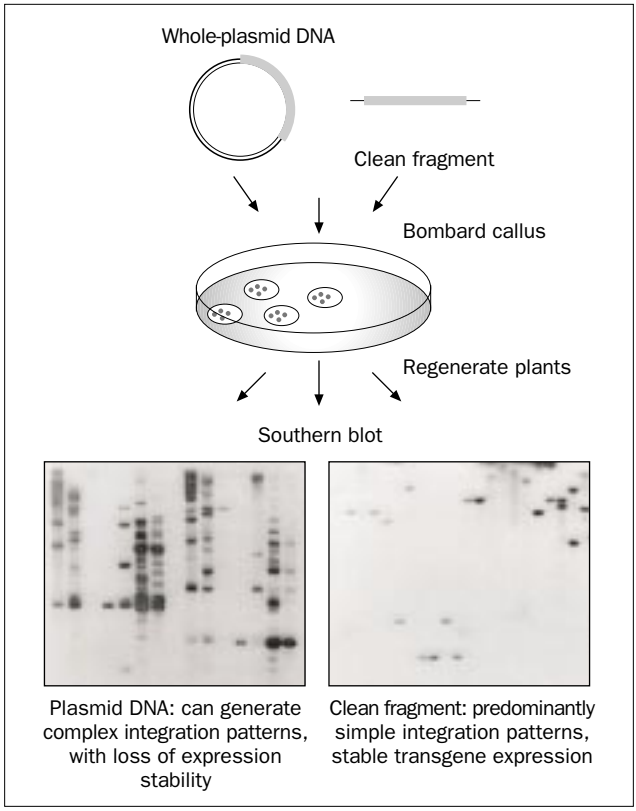


Fig. 4. Clean DNA transformation. Clean DNA is the linear minimal transgene expression cassette, cleaved from the plasmid using restriction enzymes or generated by polymerase chain reaction. In comparative experiments, clean DNA generates simpler integration patterns and a greater proportion of plants with stable transgene expression, even though overall transformation frequencies are similar using the two methods.

this issue by studying a single transgenic rice line containing a single copy of a construct comprising three heterologous transgenes: *hpt*, *gusA*, and *bar* (Fu et al 2000a). By concentrating on this line, we eliminated variation caused by different position effects and different transgene copy numbers, since the transgene locus was identical in structure in each plant we analyzed. Furthermore, because there was only a single copy of the locus and because none of the transgenes were homologous to endogenous rice genes, we eliminated the possibility that ectopic pairing could influence transgene expression.

Despite these precautions, we found that silencing still occurred *de novo* in several R_1, R_2, and R_3 plants. In all cases, silencing was associated with DNA methylation of the silenced locus, but in some cases this was heritable through meiosis and in

other cases it was not. We observed a novel form of silencing in the *hpt* transgene, involving the methylation of one DNA strand, with the result that the silencing phenotype was inherited by 50% of the progeny in each generation. We also observed developmentally regulated silencing of *gusA*, which was active at 3 weeks postgermination but inactive at 6 weeks. We found that silencing could affect the *hpt* gene without influencing the adjacent *gusA* gene, and vice versa, even though the two transgenes were separated by less that 50 bp of intervening DNA. Most remarkably of all, we found that the *hpt* and *gusA* genes could be differentially silenced in the same plant, with each silencing mode associated with an entirely distinct and nonspreading type of DNA methylation (Fig. 5). For all these variations to occur in sibling plants with the same transgenic locus suggests that there is still much to learn about the factors responsible for DNA methylation and silencing.

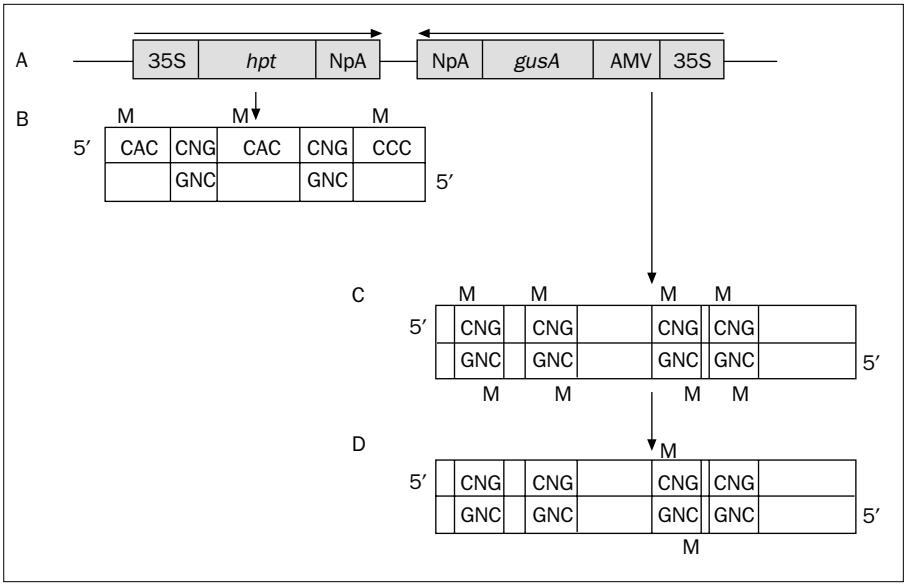


Fig. 5. Methylation and transgene silencing in a single rice plant. (A) Transgenic rice line C549 carried a single-copy transgenic locus comprising three heterologous transgenes: *bar* (not shown), *hpt*, and *gusA*. Each transgene was separated by less than 100 bp of DNA. In one transgenic R_1 plant, the *hpt* and *gusA* genes simultaneously underwent different modes of silencing, associated with distinct, nonspreading methylation patterns. (B) The *hpt* gene underwent transcriptional silencing, associated with hemimethylation of the coding region at nonconventional (non-CNG) sites. The silencing was transmitted to 50% of progeny plants, presumably those inheriting the methylated DNA strand. (C) The *gusA* gene concurrently underwent posttranscriptional silencing, associated with methylation at conventional CNG sites in the coding region. However, while these CNG sites were methylated at 3 weeks postgermination (C), most were demethylated at 6 weeks postgermination (D), corresponding to the onset of *gusA* gene expression.

Concluding comments

Experiments involving cotransformation with multiple transgenes, in combination with FISH analysis, have given us a unique perspective of how hierarchical organization affects the level and stability of transgene expression in cereals. Undoubtedly, further studies need to be carried out to completely unravel the factors that permit stable and predictable transgene expression, and these studies may offer a glimpse of the complex interacting mechanisms involved. In the future, it may be possible to design transformation constructs that permit full predictable control over transgene integration and therefore provide guaranteed routes to expression. Until this time, it is important to continue investigating the mechanisms that control transgene integration, organization, and expression in the cereals we all rely on so heavily.

References

Abranches R, Santos AP, Williams S, Wegel E, Castilho A, Christou P, Shaw P, Stoger E. n.d. Widely-separated multiple transgene integration sites in wheat chromosomes are brought together at interphase. Plant J. (In press.)

Bano-Maqbool S, Christou P. 1999. Multiple traits of agronomic importance in transgenic indica rice plants: analysis of transgene integration patterns, expression levels and stability. Mol. Breed. 5:471-480.

Berg DE, Howe M. 1989. Mobile DNA. Washington, D.C. (USA): American Society of Microbiology.

Chen LL, Marmey P, Taylor NJ, Brizard JP, Espinoza C, de Cruz P, Huet H, Zhang SP, de Kochko A, Beachy RN, Fauquet CM. 1998. Expression and inheritance of multiple transgenes in rice plants. Nature Biotechnol. 16:1060-1064.

Christou P, McCabe D, Swain WF. 1988. Stable transformation of soybean callus by DNA-coated gold particles. Plant Physiol. 87:671-674.

De Groot MJ, Offringa R, Groet J, Does MJ, Hooykaas PJ, van dan Elzen PJ. 1994. Non-recombinant background in gene targeting: illegitimate recombination between *hpt* gene and defective 5′ deleted *nptII* gene can restore a Km^r phenotype in tobacco. Plant Mol. Biol. 25:721-733.

Fromm M, Callis J, Taylor LP, Walbot V. 1987. Electroporation of DNA and RNA into plant protoplasts. Methods Enzymol. 153:351-366.

Fu X, Kohli A, Twyman RM, Christou P. 2000a. Alternative silencing effects involve distinct types of non-spreading cytosine methylation at a three-gene single-copy transgenic locus in rice. Mol. Gen. Genet. 263:106-118.

Fu X, Duc LT, Fontana S, Bong BB, Tinjuangjun P, Sudhakar D, Twyman RM, Christou P, Kohli A. 2000b. Linear transgene constructs lacking vector backbone sequences generate low-copy-number transgenic plants with simple integration patterns. Transgenic Res. 9:11-19.

Gal S, Pisan B, Hohn T, Grimsley N, Hohn B. 1992. Agroinfection of transgenic plants leads to viable cauliflower mosaic virus by intermolecular recombination. Virology 187:525-533.

Gallie DR. 1998. Controlling gene expression in transgenics. Curr. Opin. Plant Biol. 1:166-172.

Gheysen G, Villarroel R, van Montagu M. 1991. Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev. 5:287-297.

Gorbunova V, Levy AA. 1997. Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res. 25:4650-4657.
Grant SR. 1999. Dissecting the mechanisms of post-transcriptional gene silencing: divide and conquer. Cell 96:303-306.
Hunold R, Bronnel R, Hahne G. 1994. Early events in microprojectile bombardment—cell viability and particle location. Plant J. 5:593-604.
Klein TM, Wolf ED, Wu R, Stanford JC. 1987. High-velocity microprojectiles for delivering nucleic acids into living cells. Nature 327:70-73.
Klein TM, Fromm M, Weissinger A, Tomes D, Schaaf S, Sletten M, Stanford JC. 1988. Transfer of foreign genes into intact maize cells with high-velocity microprojectiles. Proc. Natl. Acad. Sci. USA 85:4305-4309.
Kohli A, Leech MJ, Vain P, Laurie DA, Christou P. 1998. Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot-spots. Proc. Natl. Acad. Sci. USA 95:7203-7208.
Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P. 1999. Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology-mediated recombination. Plant J. 17:591-601.
Komari T, Hiei Y, Ishida Y, Kumashiro T, Kubo T. 1998. Advances in cereal gene transfer. Curr. Opin. Plant Biol. 1:161-165.
Krens FA, Molendijk L, Wullems GJ, Schilperoort RA. 1982. In vitro transformation of plant protoplasts with Ti-plasmid DNA. Nature 296:72-74.
Matsumoto S, Ito Y, Hosoi T, Takahashi Y, Machida Y. 1990. Integration of T-DNA into tobacco chromosome: possible involvement of DNA homology between T-DNA and plant DNA. Mol. Gen. Genet. 224:309-316.
Mayerhofer R, Koncz-Kalman Z, Nawrath C, Bakkeren G, Crameri A, Angelis K, Redei GP, Schell J, Hohn B. 1991. T-DNA integration: a mode of illegitimate recombination in plants. EMBO J. 10:697-704.
Mehlo L, Mazithulela G, Twyman RM, Boulton MI, Davies JF, Christou P. n.d. Structural analysis of transgene rearrangements and effects on expression in transgenic maize plants generated by particle bombardment. Maydica. (In press.)
Negrutiu I, Shillito R, Potrykus I, Biasini G, Sala F. 1987. Hybrid genes in the analysis of transformation conditions. I. Setting up a simple method for direct gene transfer in plant protoplasts. Plant Mol. Biol. 8:363-373.
Salomon S, Puchta H. 1998. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J. 17:6086-6095.
Schoelz JE, Wintermantel WM. 1993. Expansion of viral host range through complementation and recombination in transgenic plants. Plant Cell 5:1681-1692.
Swoboda P, Hohn B, Gal S. 1993. Somatic homologous recombination *in planta*: the recombination frequency is dependent on the allelic state of the recombining sequences and may be influenced by genomic position effects. Mol. Gen. Genet. 273:33-40.
Takano M, Egawa H, Ikeda JE, Wakasa K. 1997. The structures of integration sites in transgenic rice. Plant J. 11:353-361.
Tang K, Tinjuangjun P, Xu Y, Sun X, Gatehouse JA, Ronald PC, Qi H, Lu X, Christou P, Kohli A. 1999. Particle bombardment-mediated co-transformation of elite Chinese rice cultivars with genes conferring resistance to bacterial blight and sap-sucking insect pests. Planta 208:552-563.

Tinland B. 1996. The integration of T-DNA into plant genomes. Trends Plant Sci. 1:178-184.
Tsukamoto Y, Ikeda H. 1998. Double-strand break repair mediated by DNA end-joining. Genes Cells 3:135-144.
Twyman RM, Whitelaw CAB. 2000. Animal cell genetic engineering. In: Spier RE, editor. Encyclopedia of cell technology. New York (USA): John Wiley & Sons. p 737-819.
Twyman RM, Christou P, Stoger E. n.d. Genetic transformation of plants and their cells. In: Oksman-Caldentey KM, Barz W, editors. Plant biotechnology and transgenic plants. New York (USA): Marcel & Dekker. (In press.)

Notes

Authors' address: Molecular Biotechnology Unit, John Innes Center, Colney Lane, Norwich NR4 7UH, U.K.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Gene silencing and its reactivation in transgenic rice

T.C. Hall, S.P. Kumpatla, P. Kharb, L. Iyer, M. Cervera, Y. Jiang, T. Wang, G. Yang, P. Teerawanichpan, J. Narangajavana, and J. Dong

Genome surveillance systems of higher organisms protect not only against intragenomic parasites such as retroposons and transposons but also against invasive DNA introduced via genome transformation, resulting in (trans)gene silencing at both the transcriptional and posttranscriptional levels. Transcriptional gene silencing of both bialaphos resistance and *Btt CryIIIA* crystal toxin genes was encountered in experiments targeted to provide resistance to the rice water weevil. Extensive methylation of the transgenes, especially of the promoter elements, was evident. Germination of seedlings from silenced lines in the presence of 5-azacytidine (which prevents cytosine methylation) led to the reactivation of bialaphos resistance in progeny of silenced plants. The *CaMV 35S* promoter has been implicated in many instances of silencing and the possibility that some promoters are especially prone to silencing is supported by the finding that *RCg2*, a rice root-specific promoter, is silenced in more than 80% of transformants. Genes flanking *RCg2/uidA* are rarely silenced, showing that silencing can be highly targeted. Multicopy insertions are very susceptible to silencing and epistatic interactions between multicopy and single-copy inserts are documented. Several strategies to alleviate silencing have been considered, such as sequence diversification and flanking transgene inserts with matrix attachment regions.

The ability to isolate genes and transfer them among organisms so that they are expressed in an alien genomic background is surely one of the most exciting discoveries of the 20th century. Such expression has provided great insight into the fundamental principles of biology as well as opening new practical opportunities for the improvement of crops and the development of novel pharmaceutical products. The realization that some transgenes do not express as anticipated is a significant challenge for genetic engineering of crops, but is also leading to an unanticipated understanding of epigenetic phenomena that were previously perplexing (Wolffe and Matzke 1999). Indeed, over the past decade it has become abundantly clear that some genome surveillance systems complement and may even evolutionarily precede better known

immune systems. Thus, restriction endonucleases may reflect ancient prokaryotic defense systems against virus infection, elaborated by methylation strategies to avoid degradation of self DNA, that have evolved to provide protection against intragenomic parasites such as retroposons and transposons in higher organisms (Bestor and Tycko 1996). The introduction of novel genetic information through genome transformation is subject to these surveillance mechanisms, resulting in (trans)gene-silencing phenomena.

The failure of transgenes to express is now recognized to result from multiple mechanisms, but the functional distinction between transcriptional gene silencing (TGS) and posttranscriptional gene silencing (PTGS) is widely accepted (Fagard and Vaucheret 2000). Promoter methylation is highly characteristic of TGS (Bestor and Tycko 1996, Kumpatla et al 1997). However, the occurrence of TGS in organisms such as yeast and *Drosophila* that do not have detectable DNA methylation suggests that it is not necessarily the causative event. Indeed, repression of methylated promoters probably results most frequently through methylated DNA binding proteins (such as MeCP2) that recruit chromatin-modifying factors (such as histone deacetylases) and remodeling factors (such as SNF2 helicases) that yield a heterochromatin-like promoter status (Bird and Wolffe 1999, Wolffe and Matzke 1999), thus preventing access by the transcription machinery to the promoter DNA.

PTGS was originally described as cosuppression in plants (Napoli et al 1990, van der Krol et al 1990) and is responsible for various forms of viral gene silencing (Baulcombe 1999, English et al 1996, Lindbo et al 1993). The recognition that RNA-dependent RNA polymerase is involved in PTGS revealed links with gene-silencing phenomena in other organisms, such as quelling in *Neurospora crassa* (Cogoni and Macino 1999a) and RNAi in *Caenorhabditis elegans* (Fire 1999), and has provided insight into the way in which sequence specificity is attained in silencing (Zamore et al 2000).

In a recent review of transgene silencing (Iyer et al 2000), it was noted that relatively few silencing events have been unequivocally identified in monocots as PTGS. Although the number of instances of PTGS will undoubtedly grow rapidly, we have chosen here to focus on TGS and specifically on the association of promoter region methylation with silencing events.

Gene silencing encountered in a practical application

The rice water weevil (RWW), *Lissorhoptrus oryzophilus* K., is the most damaging and ubiquitous insect pest of rice in the United States. Adult weevils move to rice fields in spring from nearby overwintering habitats, feed on young expanding rice leaves, and lay eggs in submerged rice stems and leaf sheaths. The eggs hatch in 4–9 d and shortly thereafter the larvae move to and feed on the roots of the young rice plants for about 27 d as they go through four stages of larval development. Extensive root pruning by feeding larvae decreases the plant's ability to take up nutrients from the soil, leading to stunted growth, delayed maturity, and reduced yield.

Carbofuran (trade name, Furadan) is a highly toxic, systemic carbamate (cholinesterase inhibitor) insecticide. It is effective in protecting the crop from substantial yield losses, but its toxicity to nontarget organisms in the rice ecosystem is a serious problem and its use has been discontinued in the U.S. Although alternative effective chemical insecticides are now approved, their efficacy against RWW is still being evaluated and this, combined with a lack of viable traditional sources of resistance to the RWW, makes biotechnology an attractive option for combating this pest. In collaboration with the group of John Kemp (New Mexico State University, Las Cruces), we are exploring the use of a codon-optimized version of the *Btt CryIIIA* crystal toxin gene, which is effective against coleopteran insects (Sutton et al 1992).

Because of the 4-wk feeding period during which larvae feed on the roots, we initially biolistically cotransformed rice (Buchholz et al 1998) with *p35S-CryIIIA*, which uses a *35S* promoter to drive expression of the *Btt CryIIIA* coding region (Fig. 1A), and *mUbi1/bar* (Fig. 1B) as a selectable marker. Resulting plants were designated JKA lines. Unfortunately, an extensive investigation in our laboratory of foreign gene expression in several hundred transgenic rice plants generated by electroporation or biolistics revealed that transformants bearing multicopy inserts frequently displayed non-Mendelian segregation (Kumpatla and Hall 1999, Kumpatla et al 1997). For example, characterization of several R_1 progeny revealed that *mUbi1/bar* and *35S/Btt* were extensively methylated and transcriptionally inactivated in the silenced lines (Fig. 2). The epigenetic modification of the transgene sequences was further confirmed by the reactivation of the *bar* gene expression in R_2 seedlings (from the silenced lines) germinated on medium containing 5-azacytidine (AzaC) (Kumpatla et al 1997).

It appears that, once a locus is sensitized to methylation, it is very prone to methylation. Thus, in an investigation of 34 R_2 plants in which bialaphos resistance was reactivated by seedling exposure to AzaC, all except one were re-silenced by 50 d after germination. The single line that retained resistance to maturity is male sterile and yields no transgenic seed when crossed with wild-type pollen. Reestablishment of silencing was found to correlate in each case with methylation of the ubiquitin promoter (Kumpatla and Hall 1998a). When following JKA52 progeny of a homozygous, nonsilenced R_1 line, we found that promoter methylation and silencing may arise in the R_3 generation (Kumpatla and Hall 1998b). The stochastic nature of initiation of methylation is intriguing: in plants grown under identical conditions, all progeny of two lines remained nonmethylated and unsilenced, whereas three other lines gave progeny that were 5% to 40% silenced. It is important to recognize that chromosomal location does not participate in these differences since all lines studied are progeny of a single ancestral plant.

Gene-specific silencing

Are some promoters especially prone to silencing?

In early investigations of monocot transformation, comparison was made between the efficacy of using circular and linearized forms of plasmid (Battraw and Hall 1991).

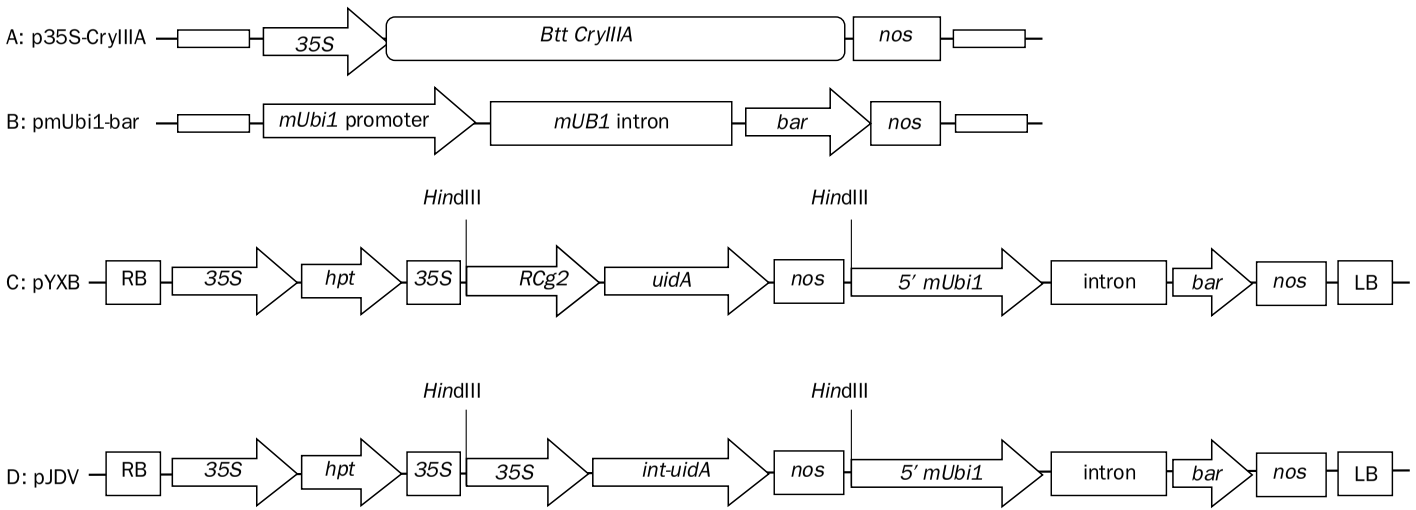


Fig. 1. Plasmid maps for transgenes used in these studies. Plasmids p35S-CryIIIA (A: *35S/Btt CryIIIA/nos*) and pmUbi1-bar (B: *mUbi1/bar/nos*) were used to biolistically generate transgenic rice for rice water weevil resistance. Binary vector pYXB (C: *35S/hpt/35S::RCg2/uidA/RCg2::mUbi1/bar/nos*) was used to generate YXB rice transformants for functional analysis of the *RCg2* promoter sequence. Plasmid pJDV (D: *35S/hpt/35S::35S/int-uidA/nos::mUbi1/bar/nos*) was used for analysis of transgene expression driven by the *35S* promoter. The *Hin*dIII sites used in the blot analysis are shown. RB = right border, LB = left border.

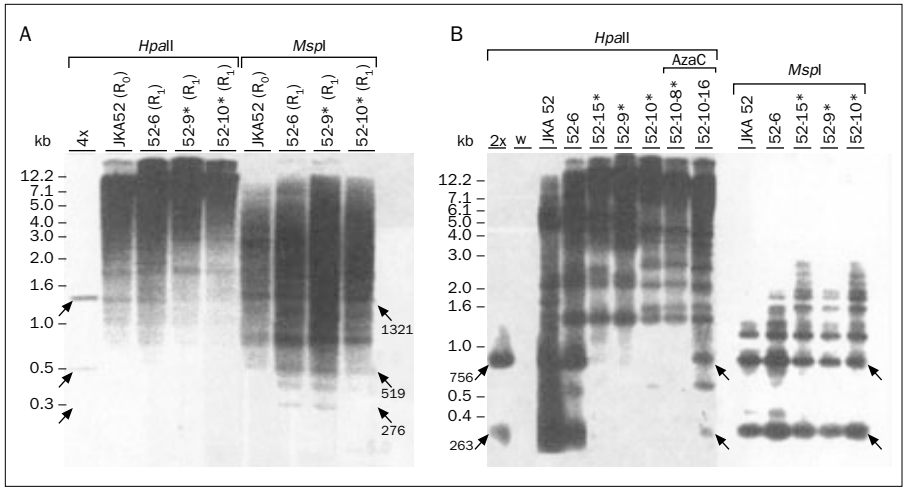


Fig. 2. Methylation and transgene expression in biolistically generated rice. Genomic DNA from nonsilenced parent (JKA52) and progeny (JKA52-6) lines and silenced lines (JKA52-9 and JKA52-10) was digested with *Hpa*II or *Msp*I and hybridized with (A) a *Btt CryIIIA* coding region probe or (B) an *mUbi1* promoter probe. 2x = 2-copy and 4x = 4-copy reconstruction of *p35S/CryIIIA* DNA; arrows indicate locations of expected fragments. * denotes a silenced line. Modified from Kumpatla et al (1997) and Kumpatla and Hall (1999), with permission.

Although linearized plasmid was found to yield twice the number of resistant calli as circular plasmid, the additional expense and effort of linearization and purification were not deemed worthwhile. However, insertion of the linear transgene requires that fragmentation occur at some point. The relative physical fragility of sequences within the plasmid, or susceptibility to plant nucleases, or both, are probably important parameters. Since such sequences may be within the transgene element rather than in the vector elements, it is not surprising that direct (biolistic and electroporation) transformation approaches rarely yield discrete inserts if circular DNA is used. The *35S* promoter has been widely used for plant transformation, but evidence is accumulating that it contains several sites that are very susceptible to fragmentation (Kohli et al 1999, Kumpatla and Hall 1999). Other DNA features, such as the possible sequence or structural differences between eukaryotic and prokaryotic DNA, have been suggested as signals that could lead to transgene sequences being recognized by host surveillance systems as intrusive (Kumpatla et al 1998). Interestingly, Fu et al (2000a) have recently shown that linear transgene constructs lacking vector backbones are very effective for rice transformation and reduce the occurrence of transgene silencing.

The above considerations indicate that many parameters contribute to variability in efficacy of expression from any transgene construct. However, it is tempting to consider the possibility that some sequences are especially prone to silencing. We have accumulated substantial evidence that the rice *RCg2* promoter (Xu et al 1995) may be such a sequence. The use of a strong root-specific promoter is especially

appealing in regard to our efforts to combat the RWW in its larval stages. However, in studying more than 100 plants independently transformed with an *RCg2/*(*β-glucuronidase*) *uidA* construct (pYXB, Fig. 1C), we have found that less than 20% yield GUS expression and that few of these plants express strongly (Table 1, Fig. 3), even though *Agrobacterium*-mediated transformation was used (Dong et al 1996) to generate plants that have simple, low-copy-number integration patterns. We have confirmed that the majority of the plants are indeed silenced by observing reactivation of GUS in mature roots (Fig. 3B) and seedlings treated by AzaC (50 mg L^{-1}). The observation of brief reactivation of GUS activity from the *RCg2-uidA* in germinating seedlings transgenic for pYXB also confirmed silencing of *RCg2-uidA*. Although mature roots exhibited weak GUS expression, many of the young seedlings showed strong expression, indicating meiotic resetting of silencing. However, activity decayed after about 5 to 7 d. The reactivation by AzaC and meiotic resetting of *RCg2/uidA* are not uniform among the lines or even throughout individual plants, indicating that various silencing mechanisms may operate and that the chromosomal position of the transgene may affect the silencing processes. We have established fluorescent *in situ* hybridization (FISH) techniques for rice transgenes (Fig. 3D) and hope to gain insight into the possibility that position affects the susceptibility of a transgene for silencing.

Table 1. A summary of transgene expression analysis of YXB plants. More than 100 independent rice lines were analyzed for the presence and expression of *mUbi1/bar* and *RCg2/uidA*. More than 80% of the YXB plants showed silencing of *RCg2/uidA* while the flanking *mUbi1/bar* and *35S/hpt* genes remain active, indicating that the silencing is highly efficient and targeted specifically to *RCg2/uidA*.

Protein	Transgene analysis			
	Expressing		Nonexpressing[a]	
GUS	Strong (8)	Weak (19)	Gene* present (55)	Absent (11)
	49a, 68, 69a, 91–93, 98, 138	56–59, 61, 79, 122, 126–128, 130, 135, 147, 149, 151, 153, 156–158	2, 12, 19, 24, 25, 33, 37, 44, 49b, 51–52, 54, 55, 60, 62, 63, 65–66, 69b, 71–73, 75, 80, 82, 84, 88–90, 100–102, 104, 106, 108–109, 114–115, 119, 123–125, 131–134, 136–137, 139–141, 143–146	1, 32, 50, 64, 76, 78, 86, 96, 105, 121, 129
BAR	Resistant (74)		Gene* present (7)	Absent (12)
	2, 12, 24, 25, 33, 37, 44, 49a, 49b, 51, 52, 54–61, 63, 66, 68, 69a, 69b,71–73, 75, 79, 80, 82, 84, 88–93, 98, 100, 102, 104, 106, 108, 114, 115, 119, 121–128, 130, 131, 133, 135, 137–141, 143–147, 151, 153, 156–158		19, 62, 76, 96, 101, 136, 149	1, 32, 50, 64, 65, 78, 86, 105, 109, 129, 132, 134

[a]* = intact transgene.

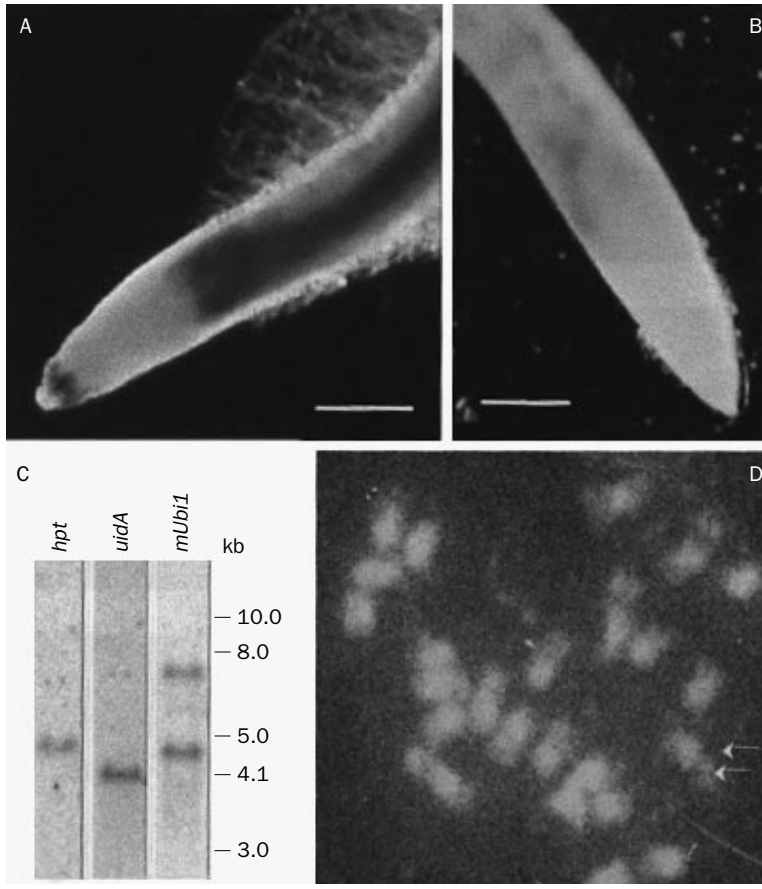


Fig. 3. Transgene organization and expression in YXB rice. In a few lines, *RCg2/uidA* yielded strong activity in the root cap and elongation region of rice stably transformed via *Agrobacterium* (A). Roots of some *RCg2/uidA*-silenced lines (e.g., YXB37) can be reactivated by AzaC treatment (B). Genomic DNA blot of YXB37 showing a low-copy transgene integration event (C). *Hin*dIII-digested genomic DNA was blot-analyzed with ^{32}P-labeled DNA fragments for right border (*hpt*), left border (*mUbi1*), and internal (*uidA*) sequences. Metaphase chromosomes showing FISH localization of the *RCg2/uidA* transgene: digoxygenin (DIG)-labeled T-DNA of pYXB was hybridized to the chromosomes and detected by Cy3 (red spots, indicated by yellow arrows) on 4′, 6-diamidine-2′-phenylindole dihydrochloride (DAPI)-stained chromosomes (D).

Fu et al (2000b), who studied a rice line biolistically transformed with a three-gene construct containing two functional transgenes (*hpt*, *uidA*), found that the third gene (*bar*) was inactive as its *35S* promoter suffered rearrangement. Using this single-parent plant, the progeny were followed and various silencing effects were claimed. This was not surprising as each of the transgenes contained the *35S* promoter, which is known to be a target for rearrangement or a hot spot for recombination (Kohli et al

1998, Kumpatla and Hall 1999). The presence of repeated sequences is also a well-known signal for methylation and silencing (Assaad et al 1993, Dorer and Henikoff 1994, Garrick et al 1998, Henikoff 1998, Meyer 1996, Selker 1999, Ye and Signer 1996).

Epistatic interactions can silence a single gene

Epistatic regulation of transgene expression was first recorded in tobacco by Matzke et al (1989) and both *cis*-inactivation (between closely linked genes) and *trans*-inactivation (between nonlinked transgenes or a transgene and an endogenous gene) occur (Matzke et al 1993). As we were unaware of any previous reports of epistatic interactions in transgenic monocots, we thought it would be of interest to explore gene silencing in rice lines bearing single and multicopy transgene inserts.

A series of transgenic rice lines was obtained using a construct (pJDV, Fig. 1D) that bears three different genes in a single T-DNA element. Silencing of expression of the central *35S/int-uidA* gene was found to be rare in the R_0 and R_1 generations of lines bearing a single-copy insert, but multicopy insertion often yielded plants that were silenced in at least one transgene (Table 2). Silencing could be reversed by treatment with AzaC (20 or 50 mg L^{-1}) or trichostatin A (TSA) (10 or 20 nM) for all R_1 progeny of lines JDV88 (one copy), JDV92 (five copies), JDV83 (eight copies), and JDV 90 (eight copies) but not for line JDV95 (two copies).

The silencing seen in JDV95 is most likely transcriptional because it is briefly reset at meiosis, giving rise to a brief reactivation of GUS expression (48–56 h postgermination), followed by immediate re-silencing in the R_1 seedling (Fig. 4). This may represent a case in which either the transgene is inserted into a locus that is highly subject to silencing or one at which gene insertion stimulates an as yet unidentified and potent silencing mechanism.

Line JDV92 has a total of five transgene copies at two separate transgene loci (Fig. 5). The single-copy gene at one locus was silenced in the presence of the locus bearing multiple transgene copies. Segregation from the multiple transgene locus re-

Table 2. Analysis of transgene expression in JDV rice lines. Expression of *35S/uidA* and *mUbi1/bar* was assayed using histochemical staining (Jefferson et al 1987) and bialaphos dipping (Kumpatla et al 1997) on representative transformants having a low (JDV85, 86, 88, and 92) or multicopy (JDV83, 90, and 92) transgene insert. With the exception of line JDV95, all *35S/uidA*-silenced lines were reactivated following 7 d of AzaC or TSA treatment in the dark.

Transgene		R_0		R_1		AzaC (mg L^{-1})		TSA (nM)	
Line	Copy	GUS	BAR	GUS	BAR	20	50	10	20
JDV85	1	+	+	+	+	na[a]	na	na	na
JDV86	1	+	+	+	+	na	na	na	na
JDV88	1	+	+	+/–	+	+	+	+	+
JDV95	2	–	–	–	–	–	–	–	–
JDV83	8	–	+	–	+	+	+	+	+
JDV90	8	–	+	–	+	+	+	+	+
JDV92	5	–	+	–	+	+	+	+	+

[a]na = not applicable.

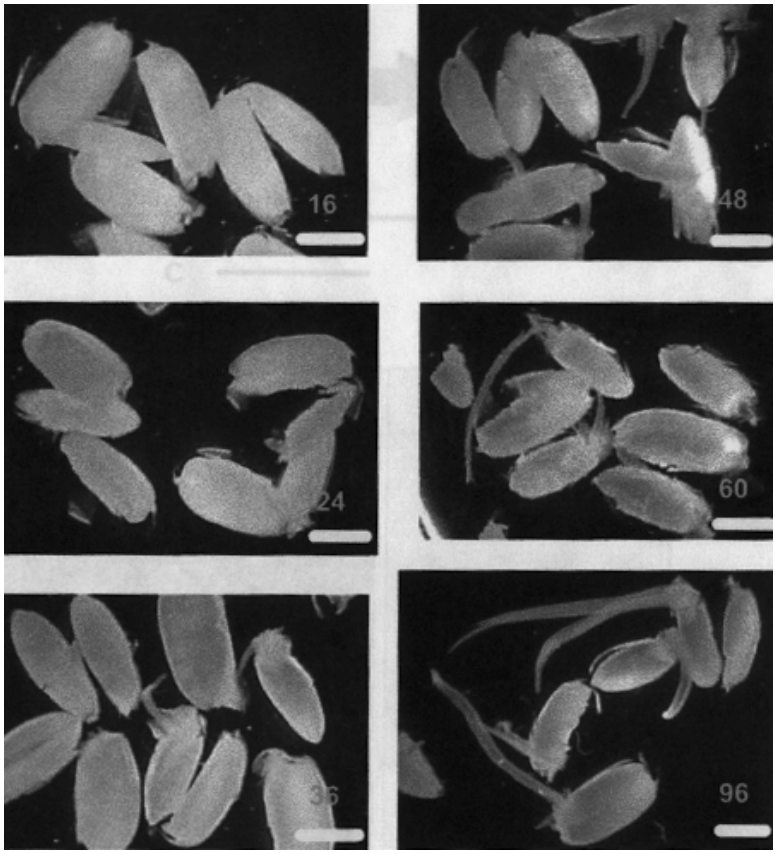


Fig. 4. X-gluc staining of germinating R_1 JDV95 seedlings. GUS activity, detected by histochemical staining (Jefferson et al 1987), was absent for the first 36 h of germination, but was evident between 48 and 60 h and inactivated afterwards. The yellow bar represents 1.5 mm.

sulted in reactivation of the gene at the single-copy locus (Table 2). This is reminiscent of paramutation where a paramutated locus regains expression when segregated away from a paramutagenic locus (Matzke et al 1996). Crosses between JDV92 and other functional locus-containing lines is being conducted to determine if its multigene locus can silence other functional single transgene loci.

To study the epistatic effects of silenced loci on functional loci, sexual crosses were made. As observed for tobacco (Matzke et al 1993, 1994, Nap et al 1997, Neuhuber et al 1994, Schmulling and Rohrig 1995, Vaucheret 1993, 1994), different results were obtained for individual crosses. For example, there was no silencing when single-copy lines JDV85 or JDV86 were crossed to the silenced two-copy line JDV95, but when they were crossed to multicopy lines JDV83 or JDV90, GUS expression was silenced in the resulting hybrid plants (Table 3). The differences in outcome may

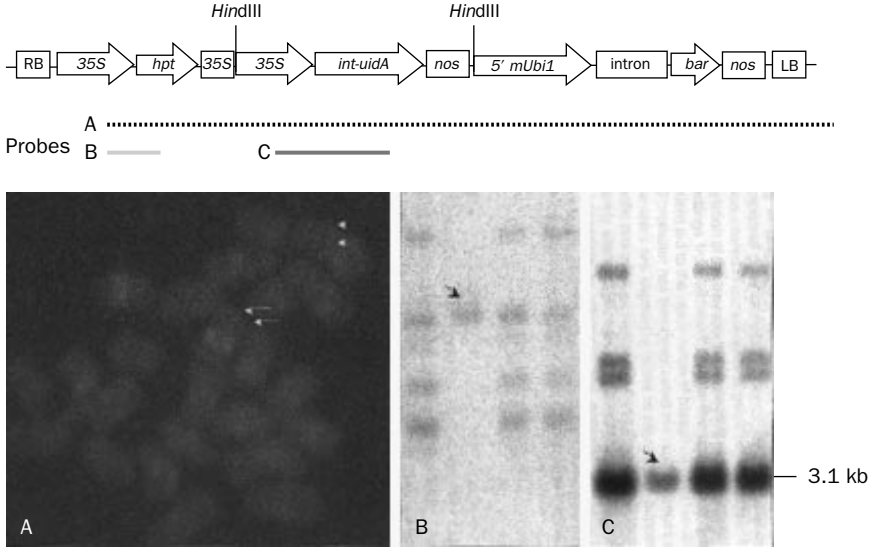


Fig. 5. FISH detection of the chromosomal location of T-DNA in JDV92. Denatured chromosomes of R_1 seedlings were hybridized with digoxygenin (DIG)-labeled T-DNA sequences of pJDV (upper panel). The hybridization signals (yellow arrow) were visualized using the fluorescent dye Cy3 in a background of 4′, 6-diamidine-2′-phenylindole dihydrochloride (DAPI)-stained chromosomes. FISH analysis revealed two chromosomal transgene integration sites in JDV92 (A). Genomic DNA blot analysis of DNA from R_1 seedlings of JDV92, digested with *Hin*dIII and probed with ^{32}P-labeled *hpt* (B) or *uidA* (C), revealed the segregation of two loci (yellow arrows in panel A), one of which contains multiple copies of the insert and the other a single-copy transgene (arrows in panels B and C). After segregation from the multiple locus, plants bearing the one-copy locus became GUS positive (see text).

Table 3. Epistatic interaction between transgene loci combined by sexual crossing. Selected *35S/uidA*-silenced lines (JDV83, 90, 95) were used in crosses with functional lines (JDV85, 86); GUS expression and bialaphos resistance were assayed in the resulting hybrids. Crosses were made in both directions and 2–3 hybrids were analyzed.

Hybrid[a]	GUS	BAR
JDV85 × 95	+	+
JDV86 × 95	+	+
JDV88* × 95	+	+
JDV83 × 95	–	+
JDV85 × 83	–	+
JDV86 × 83	–	+
JDV85 × 90	–	+
JDV86 × 90	–	+

[a]* = silenced JDV88 R_1 lines.

reflect different causes for silencing. For example, the presence of a multicopy locus may trigger silencing at relatively distant loci such as the super-silencer loci described by Matzke et al (1994) and Vaucheret (1993), whereas the apparently potently silenced locus (both *uidA* and *bar* are inactive) in JDV95 may have only local (*cis*) effects. It is noteworthy that the *int-uidA* gene was preferentially silenced in these experiments and it is tempting to attribute this to its being driven by the *35S* promoter, which has been implicated in many cases of transgene silencing (Iyer et al 2000).

Interestingly, a reactivation of *35S*/*int-uidA* was observed in JDV95 × JDV88. The reactivation of the silenced *35S*/*int-uidA* in JDV88 in the presence of (or by) the transgene locus in JDV95 is unique in that interaction between two silenced loci resulted in reactivation of the silenced gene. This observation implies that the interaction among the silenced loci is dependent on distance because the crosses between JDV95 and other functional loci (JDV85, JDV86) and silenced loci (JDV83, JDV90) did not show any effect on the reporter gene expression. The remote possibility exists that the activation observed represents complementation of mutations caused by transgene insertion. Transgene inactivation resulting from interacting loci has previously been shown to be chromosomal-location-dependent (Schmulling and Rohrig 1995). Addressing the physical distance between transgene loci is critical in understanding inactivation and reactivation, but remains a challenge because current FISH technology does not allow estimation of the native physical distance between two transgene loci on different chromosomes. We are examining segregating hybrid lines to determine whether the modification of the functional locus by the multicopy transgene locus is reversible.

That transgene silencing is caused by various mechanisms involving modifications of DNA, RNA, and chromatin structure has now been rigorously established. However, the transgene-silencing phenomena we have observed in rice are complicated, for example, AzaC and TSA treatment reactivated GUS expression in lines JDV90 and JDV92 but not in JDV95. Although multicopy transgene events are most frequently silenced (Flavell 1994, Garrick et al 1998, Henikoff 1998, Kumpatla and Hall 1999), silencing can occur for a single intact transgene locus (Table 1) and even for a single gene flanked by active transgenes at a single locus (JDV88); thus, there may or may not be epistatic interactions between transgene loci.

Alleviating transgene silencing

As silencing of desirable transgenes is a substantial complication in biotechnological development of novel crops, there is considerable interest in avoiding gene silencing or in overcoming its effects. Insight into the mechanisms of both TGS and PTGS is rapidly increasing (Cogoni and Macino 1999b, Fagard and Vaucheret 2000, Iyer et al 2000) and several mutations (mostly in *Arabidopsis*) have now been shown to affect gene silencing. Several groups, including our own (Kumpatla et al 1998), are exploring strategies for evading silencing by introducing sequence heterogeneity into the *RCg2* promoter. The suppression of methylation systems and the reduction of opportunities for heterochromatinization, for example, by the insertion of enhancer ele-

ments (Francastel et al 1999), are potential targets for reducing the incidence of gene silencing. Chromosomal location is important and the formation of protected local chromatin domains by the inclusion of matrix attachment region (MAR) elements flanking the gene to be introduced has been proposed in order to minimize position effect (Kumpatla et al 1998, Mlynárová et al 1994, Spiker and Thompson 1996). Transgenic rice plants containing a *uidA* reporter construct flanked by MARs from tobacco (Rb7) or yeast (ARS1) were reported to have significantly higher transgene expression levels than the controls, and GUS activity increased in proportion to transgene copy number up to 20 copies (Vain et al 1999). However, the data show that the constructs were extensively rearranged in the transgenic rice plants and the modest twofold overall increase in expression may reflect the ability of AT-rich enhancer elements present in the MAR sequences to function as independent enhancers, rather than any chromatin effect (van der Geest and Hall 1997). Thus, additional experiments with more discretely characterized transgene inserts and direct analysis of chromatin domain effects are needed to evaluate the contribution of MAR elements to the fidelity of transgene expression.

Although sequence similarity is closely associated with methylation-based silencing, many endogenous genes are present in multiple copies. The small sequence differences present in such genes may be very important in avoiding detection by genome surveillance processes and the introduction of promoter and coding sequence diversity between transgenes and resident genes has been proposed (Iyer et al 2000, Kumpatla et al 1998). These reviews also include consideration of other characteristics of transgenes that make them visible to genome surveillance processes, such as the disruption of normal gene functions and the possib ility that prokaryotic vector sequences have signature sequences or structures that mark them as alien to the eukaryotic genome (Bernardi 1995, Karlin et al 1998).

Recently, the use of *Arabidopsis* as a model system has given remarkable insight into gene assemblies that are involved in genome-silencing surveillance processes. Table 4 provides a synopsis of these genes. The *mom* gene (Amedeo et al 2000), which has yet to be cloned, is especially exciting in that it appears to be able to overcome PTGS for several genes without developmental penalties. It will be important to determine whether expression of it, or its orthologs, can alleviate silencing in rice and other crop plants. Similarly, the *som* mutants (Elmayan et al 1998) appear to be effective against TGS in *Arabidopsis* and tests of their efficacy in crop plants are eagerly awaited.

Table 4. Suppressors of silencing found in *Arabidopsis*.

Mutant and complete name	Derived by[a]	P/TGS[b]	Effect on methylation	Function in wild-type form	Phenotype[c]	References
ddm1 *d*ecreased *D*NA *m*ethylation	EMS	TGS	Decreased	SWI2/SNF2 required for normal methylation	Aberrant development	Vongs et al (1993) Kakutani et al (1999) Jeddeloh et al (1998) Jeddeloh et al (1999) Furner et al (1998) Mittelsten Scheid et al (1998) Hirochika et al (2000)
hog1 *ho*mology-dependent *g*ene silencing 1	EMS, X-ray	TGS	Decreased	Maintain methylation status of transgene, rDNA	ns	Furner et al (1998)
mom1 *Mo*rpheus *m*olecule	T-DNA	TGS	None	Putative, chromatin remodeling	Normal	Amedeo et al (2000) Steimer et al (2000)
met1 *met*hyltransferase	Antisense	TGS	Decreased	DNA methyltransferase	Aberrant flowers	Ronemus et al (1996) Genger et al (1999)
sgs1, sgs2 *s*uppressor of *g*ene *s*ilencing	EMS	PTGS	Decreased	Maintains methylation status of transgene	ns	Elmayan et al (1998)
sil1, sil2 *sil*encing	EMS, X-ray	TGS	None	Putative, recognizes methylated sequence	ns	Furner et al (1998)
som1-8 *som*niferous	EMS, FNR	TGS	Decreased	Maintains methylation status of transgene and centrimeric repeats	Aberrant in F_4	Mittelsten Scheid et al (1998)

[a]EMS = ethyl methane sulfonate, FNR = fast neutron irradiation. [b]TGS = transcriptional gene silencing, PTGS = posttranscriptional gene silencing. [c]ns = not stated.

References

Amedeo P, Habu Y, Afsar K, Scheid OM, Paszkowski J. 2000. Disruption of the plant gene *MOM* releases transcriptional silencing of methylated genes. Nature 405:203-206.

Assaad FF, Tucker KL, Signer ER. 1993. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. Plant Mol. Biol. 22:1067-1085.

Battraw MJ, Hall TC. 1991. Stable transformation of *Sorghum bicolor* protoplasts with chimeric neomycin phosphotransferase II and β-glucuronidase genes. Theor. Appl. Genet. 82:161-168.

Baulcombe D. 1999. Viruses and gene silencing in plants. Arch. Virol. Suppl. 15:189-201.

Bernardi G. 1995. The human genome: organization and evolutionary history. Annu. Rev. Genet. 29:445-476.

Bestor TH, Tycko B. 1996. Creation of genomic methylation patterns. Nat. Genet. 12:363-367.

Bird AP, Wolffe AP. 1999. Methylation-induced repression—belts, braces, and chromatin. Cell 99:451-454.

Buchholz WG, Teng W, Wallace D, Ambler JR, Hall TC. 1998. Production of transgenic rice (*Oryza sativa* subspecies *japonica* cv. Taipei 309). Methods Mol. Biol. 81:383-396.

Cogoni C, Macino G. 1999a. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. Nat. 399:166-169.

Cogoni C, Macino G. 1999b. Homology-dependent gene silencing in plants and fungi: a number of variations on the same theme. Curr. Opin. Microbiol. 2:657-662.

Dong JJ, Teng WM, Buchholz WG, Hall TC. 1996. *Agrobacterium*-mediated transformation of javanica rice. Mol. Breed. 2:267-276.

Dorer DR, Henikoff S. 1994. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. Cell 77:993-1002.

Elmayan T, Balzergue S, Beon F, Bourdon V, Daubremet J, Guenet Y, Mourrain P, Palauqui JC, Vernhettes S, Vialle T, Wostrikoff K, Vaucheret H. 1998. *Arabidopsis* mutants impaired in cosuppression. Plant Cell 10:1747-1758.

English JJ, Mueller E, Baulcombe DC. 1996. Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes. Plant Cell 8:179-188.

Fagard M, Vaucheret H. 2000. (Trans)gene silencing in plants: how many mechanisms? Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:167-194.

Fire A. 1999. RNA-triggered gene silencing. Trends Genet. 15:358-363.

Flavell RB. 1994. Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proc. Natl. Acad. Sci. USA 91:3490-3496.

Francastel C, Walters MC, Groudine M, Martin DI. 1999. A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. Cell 99:259-269.

Fu X, Duc LT, Fontana S, Bong BB, Tinjuangjun P, Sudhakar D, Twyman RM, Christou P, Kohli A. 2000a. Linear transgene constructs lacking vector backbone sequences generate low-copy-number transgenic plants with simple integration patterns. Transgenic Res. 9:11-19.

Fu X, Kohli A, Twyman RM, Christou P. 2000b. Alternative silencing effects involve distinct types of non-spreading cytosine methylation at a three-gene, single-copy transgenic locus in rice. Mol. Gen. Genet. 263:106-118.

Furner IJ, Sheikh MA, Collett CE. 1998. Gene silencing and homology-dependent gene silencing in *Arabidopsis*: genetic modifiers and DNA methylation. Genetics 149:651-662.

Garrick D, Fiering S, Martin DI, Whitelaw E. 1998. Repeat-induced gene silencing in mammals. Nat. Genet. 18:56-59.
Genger RK, Kovac KA, Dennis ES, Peacock WJ, Finnegan EJ. 1999. Multiple DNA methyltransferase genes in *Arabidopsis thaliana*. Plant Mol. Biol. 41:269-278.
Henikoff S. 1998. Conspiracy of silence among repeated transgenes. Bioessays 20:532-535.
Hirochika H, Okamoto H, Kakutani T. 2000. Silencing of retrotransposons in *Arabidopsis* and reactivation by the ddm1 mutation. Plant Cell 12:357-369.
Iyer LM, Kumpatla SP, Chandrasekharan MB, Hall TC. 2000. Transgene silencing in monocots. Plant Mol. Biol. 43:323-346.
Jeddeloh JA, Bender J, Richards EJ. 1998. The DNA methylation locus DDM1 is required for maintenance of gene silencing in *Arabidopsis*. Genes Dev. 12:1714-1725.
Jeddeloh JA, Stokes TL, Richards EJ. 1999. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat. Genet. 22:94-97.
Jefferson RA, Kavanaugh TA, Bevan MW. 1987. GUS fusions β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6:3901-3908.
Kakutani T, Munakata K, Richards EJ, Hirochika H. 1999. Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of *Arabidopsis thaliana*. Genetics 151:831-838.
Karlin S, Campbell AM, Mrazek J. 1998. Comparative DNA analysis across diverse genomes. Annu. Rev. Genet. 32:185-225.
Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P. 1999. Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. Plant J. 17:591-601.
Kohli A, Leech M, Vain P, Laurie DA, Christou P. 1998. Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. Proc. Natl. Acad. Sci. USA 95:7203-7208.
Kumpatla SP, Chandrasekharan MB, Iyer LM, Li G, Hall TC. 1998. Genome intruder scanning and modulation systems and transgene silencing. Trends Plant Sci. 3:97-104.
Kumpatla SP, Hall TC. 1998a. Longevity of 5-azacytidine-mediated gene expression and reestablishment of silencing in transgenic rice. Plant Mol. Biol. 38:1113-1122.
Kumpatla SP, Hall TC. 1998b. Recurrent onset of epigenetic silencing in rice harboring a multicopy transgene. Plant J. 14:129-135.
Kumpatla SP, Hall TC. 1999. Organizational complexity of a rice transgene locus susceptible to methylation-based silencing. IUBMB Life 48:459-467.
Kumpatla SP, Teng W, Buchholz WG, Hall TC. 1997. Epigenetic transcriptional silencing and 5-azacytidine-mediated reactivation of a complex transgene in rice. Plant Physiol. 115:361-373.
Lindbo JA, Silva Rosales L, Proebsting WM, Dougherty WG. 1993. Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. Plant Cell 5:1749-1759.
Matzke AJ, Neuhuber F, Park YD, Ambros PF, Matzke MA. 1994. Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes. Mol. Gen. Genet. 244:219-229.
Matzke MA, Matzke AJM, Eggleston WB. 1996. Paramutation and transgene silencing: a common response to invasive DNA. Trends Plant Sci. 1:382-388.

Matzke MA, Neuhuber F, Matzke AJ. 1993. A variety of epistatic interactions can occur between partially homologous transgene loci brought together by sexual crossing. Mol. Gen. Genet. 236:379-386.
Matzke MA, Primig M, Trnovsky J, Matzke AJM. 1989. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. EMBO J. 8:643-649.
Meyer P. 1996. Repeat-induced gene silencing: common mechanisms in plants and fungi. Biol. Chem. Hoppe Seyler 377:87-95.
Mittelsten Scheid O, Afsar K, Paszkowski Z. 1998. Release of epigenetic gene silencing by *trans*-acting mutations in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 95:632-637.
Mlynárová L, Loonen A, Heldens J, Jansen RC, Keizer P, Stiekama WJ, Nap J-P. 1994. Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. Plant Cell 6:417-426.
Nap JP, Conner AJ, Mlynárová L, Stiekema WJ, Jansen RC. 1997. Dissection of a synthesized quantitative trait to characterize transgene interactions. Genetics 147:315-320.
Napoli C, Lemieux C, Jorgensen R. 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*. Plant Cell 2:279-289.
Neuhuber F, Park YD, Matzke AJM, Matzke MA. 1994. Susceptibility of transgene loci to homology-dependent gene silencing. Mol. Gen. Genet. 224:230-241.
Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. 1996. Demethylation-induced developmental pleiotropy in *Arabidopsis*. Science 273:654-657.
Schmulling T, Rohrig H. 1995. Gene silencing in transgenic tobacco hybrids: frequency of the event and visualization of somatic inactivation pattern. Mol. Gen. Genet. 249:375-390.
Selker EU. 1999. Gene silencing: repeats that count. Cell 97:157-160.
Spiker S, Thompson WF. 1996. Nuclear matrix attachment regions and transgene expression in plants. Plant Physiol. 110:15-21.
Steimer A, Amedeo P, Afsar K, Fransz P, Scheid OM, Paszkowski J. 2000. Endogenous targets of transcriptional gene silencing in *Arabidopsis*. Plant Cell 12:1165-1178.
Sutton DW, Havstad PK, Kemp JD. 1992. Synthetic *cryIIIA* gene from *Bacillus thuringiensis* improved for high expression in plants. Transgenic Res. 1:228-236.
Vain P, Worland B, Kohli A, Snape JW, Christou P, Allen GC, Thompson WF. 1999. Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J. 18:233-242.
van der Geest AH, Hall TC. 1997. The β-phaseolin 5′ matrix attachment region acts as an enhancer facilitator. Plant Mol. Biol. 33:553-557.
van der Krol AR, Mur LA, Beld M, Mol JNM, Stuitje AR. 1990. Flavonoid genes in *Petunia*: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell 2:291-299.
Vaucheret H. 1993. Identification of a general silencer for 19S and 35S promoters in a transgenic tobacco plant: 90 bp of homology in the promoter sequence are sufficient for *trans*-inactivation. C. R. Acad. Sci. Paris, Sciences de la vie 316:1471-1483.
Vaucheret H. 1994. Promoter-dependent *trans*-inactivation in transgenic tobacco plants: kinetic aspects of gene silencing and gene reactivation. C. R. Acad. Sci. Paris, Sciences de la vie/Life Sciences 317:310-323.
Vongs A, Kakutani T, Martienssen RA, Richards EJ. 1993. *Arabidopsis thaliana* DNA methylation mutants. Science 260:1926-1928.
Wolffe AP, Matzke MA. 1999. Epigenetics: regulation through repression. Science 286:481-486.

Xu Y, Buchholz WG, DeRose RT, Hall TC. 1995. Characterization of a rice gene family encoding root-specific proteins. Plant Mol. Biol. 27:237-248.
Ye F, Signer ER. 1996. RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration. Proc. Natl. Acad. Sci. USA 93:10881-10886.
Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101:25-33.

Notes

Authors' addresses: T.C. Hall, L. Iyer, M. Cervera, Y. Jiang, T. Wang, G. Yang, J. Dong, Institute of Developmental and Molecular Biology (IDMB) and Department of Biology, Texas A&M University, College Station, TX 77843-3155, USA; S.P. Kumpatla, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA; P. Kharb, Department of Genetics, CCS Haryana Agricultural University, Hisar 125004, India; P. Teerawanichpan, J. Narangajavana, Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand.
Acknowledgments: Aspects of the work reported here were supported by grants from the Texas Advanced Technology Program (010366-0092-1999), the National Science Foundation (MCB 99-74706), and Rhône-Poulenc Agro. We thank the Spanish Ministry of Education and Culture for postdoctoral fellowship support for M. Cervera and the Thailand Research Fund for partial predoctoral support for P. Teerawanichpan. The FISH analysis of YXB plants was supported in part by an Interdisciplinary Grant from Texas A&M University.
Citation: Khush GS, Brar DS, Hardy B, editors. 2001. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22-27 October 2000, Los Baños, Philippines. New Delhi (India): Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. 488 p.

Workshop reports

Rice molecular breeding workshop

This workshop was held on the International Rice Molecular Breeding Program (IRMBP), which is a comprehensive breeding strategy to enhance rice yield and stability by integrating DNA marker technology with large-scale backcross breeding activities through highly organized international collaboration. Zhikang Li highlighted the project objectives and activities. The project involves rice breeding institutions from 14 Asian countries (10 from China, 6 from India, 3 from Korea, 2 from Vietnam, 2 from Pakistan, and 1 each from Malaysia, Thailand, Bangladesh, Sri Lanka, Iran, Indonesia, Egypt, Nepal, and the Philippines). This project uses 63 elite commercial cultivars and 140 unique landraces and breeding lines as parental lines to achieve three main objectives: (1) to develop many introgression line sets in elite genetic backgrounds through large-scale backcross breeding, (2) to identify valuable genes/QTLs for important traits using introgression lines and DNA markers, and (3) to develop new cultivars with enhanced yield potential and stability by marker-aided pyramiding of desirable QTLs. The workshop was organized to discuss the current status of the IRMBP, to report on the activities and progress of the project, to report the results of molecular characterization of parental gene pools, and to discuss technical issues related to the project.

The workshop was chaired by Qifa Zhang. Gurdev Khush stressed the need to integrate DNA marker technology with conventional breeding. Workshop participants discussed backcross and phenotyping strategy, information and database establishment, and sources of funding and human resource development. A general consensus was reached for achieving the following objectives and goals of the IRMBP:

- Formulating a strategy for developing standard uniform phenotyping procedures, especially in relation to selection criteria for target traits in backcross progenies, such as using the Standard Evaluation System developed by IRRI.
- Establishing standard and common databases for parental lines and introgression lines by collecting data generated from participating institutions.
- Generating funds from local governments by submitting proposals from participating institutions with technical help from the central program at IRRI.
- Coordinating training workshops by IRRI on gene/QTL discovery and cultivar development using DNA markers and introgression lines for the participating institutions in national agricultural research and extension systems.

Functional genomics workshop

The objective of this workshop was to develop details of collaboration on three activities: (1) developing an information node, (2) promoting the sharing of genetic stocks, and (3) developing microarray resources.

Mike Gale indicated the need for more international collaboration in rice functional genomics because of the uniqueness of rice as a food crop and as a model system. Furthermore, because of the scale and costs of many genomics technologies and the emerging need for phenotyping, the research community has many incentives to share resources and combine expertise.

Web site information node

Hei Leung reported on the creation of a Web site (http://www.cgiar.org/irri/genomics/) to facilitate communication and called for active participation. Research groups with interest in participating in the working group were requested to post a summary of their research interest on the Web site and provide links to their own Web sites. Active information flow and database linkages will improve the usefulness of the site.

Genetic stocks

The representatives of research laboratories from Asia, Australia, Europe, and the United States reported on the development of mutant stocks for functional genomics research. With the growing wealth of genetic stocks (mutant lines, recombinant inbred lines, introgression lines, dihaploids, etc.), the group believed that it would be useful to have a general inventory of these resources posted on the Web site.

The group also identified several factors that potentially limit the sharing and distribution of genetic materials, particularly for transgenic insertion mutants. Several research group representatives (Gynheung An of Pohang University of Science and Technology, Korea; Andy Pereira of Plant Research International, The Netherlands; and Andrzej Kilian of CAMBIA, Australia) pointed out obstacles to the distribution of the insertion mutants: limited seed, distribution costs, special conditions needed to propagate transgenic materials, and regulatory and quarantine requirements for the movement of rice seeds between countries. These concerns represent a significant burden on the scientists to share genetic stocks on a large scale.

Hei Leung proposed two steps to initiate the process of sharing. First, collaborating laboratories can consider exchanging a small amount of seed with known characteristics. For example, insertion lines with known flanking sequences at the insertion sites can be shared among collaborators to conduct selective phenotyping. Second,

the working group can develop a network to propagate and distribute selective stocks. Gurdev Khush indicated that, because IRRI routinely distributes rice germplasm and has considerable experience in the safe movement of rice seed, IRRI can explore extending this service to mutant stocks. However, for transgenic mutants, IRRI must also assess its capacity to handle a large quantity of materials.

Development of microarray resources

Several research groups reported on the development and applications of microarray analysis in rice at the IRGS. Hans Bohnert of the University of Arizona described the progress made in using microarray to analyze gene expression under abiotic stresses. Shoshi Kikuchi of the National Institute of Agrobiological Resources, Japan, discussed the microarray project under the Japan Rice Genome Program. Novartis Agriculture Discovery Institute Inc. announced that a rice gene chip carrying 23,000 unique genes would be available for research by January 2001.

The group discussed the timeliness and benefits of developing microarray facilities that would be publicly available to the research community. The three main points were

- Microarray technology is still far from routine and technically demanding. It may be too early for an institute like IRRI to invest now unless strong technical support is available.
- The technology has multiple applications and significant technical improvements have been made over the past year. It would be a worthy investment to broaden access to this technology.
- Microarray technology requires an abundant supply of cDNA and expressed sequence tags. Thus, it is important for the research community to collaborate to make this resource available.

To address the issues of access, an ad hoc discussion group—Hans Bohnert, Andrzej Kilian, Qifa Zhang, Susan McCouch, and Akhilesh Tyagi—was formed to further analyze the costs and benefits of developing a publicly accessible microarray facility and how to promote sharing of needed resources. Hans Bohnert will continue to lead the group discussion and report back to the working group with recommendations. This will be particularly useful to IRRI as it is considering investing in a microarray facility with the goal of providing research support to many collaborating partners in Asia.

Since the International Working Group began, there has been a growing interest and commitment of the public and private research communities toward collaboration. Throughout the IRGS, many research groups expressed interest in participating in the working group. The three activities have provided the focal points for the working group to concentrate its efforts. Discussions at the IRGS workshop have provided important guidance toward sharing genetic stocks and microarray resources.

Bioinformatics workshop

The Bioinformatics Workshop was arranged to promote contacts with the rice community attending the IRGS. Graham McLaren, Richard Bruskiewich, and Michael Jackson focused on three general questions during the workshop:

1. What are our current objectives for bioinformatics development within the rice research community?
2. What additional formal collaborative framework do we need to establish within the rice bioinformatics research community to achieve those objectives?
3. On what specific technical areas should the rice bioinformatics community focus and who is willing to assume responsibility to coordinate development in those specific areas?

There was a general agreement on the goal for bioinformatics of seamlessly linking information from genome to phenotype. The initial objective of linking islands of information from different disciplines was agreed upon, and the need to develop user-friendly specialist interfaces for different users was discussed. There was a call for on-line compilation and curation of rice genetics literature and a comprehensive database on publicly available rice expressed sequence tags.

Other areas identified for development were the inclusion of more image data in databases and for graphical user interfaces. The need to develop tools to locate, manage, and display the anticipated large quantities of natural variation in sequence data was also identified. This was seen as essential for the process of allele mining in germplasm collections.

There was an agreement that the rice bioinformatics community will have to collaborate to develop standard vocabularies and data exchange protocols.

The idea of federated curation of databases was discussed, but the technical and organizational issues of enabling fully queriable, distributed databases may constrain this approach for the near future. Database links or tunnels were considered the most practical short-term mechanism of linking.

For the third point, the need to improve and standardize rice genome annotation was briefly discussed, as was the need to canvass opinions among the rice community. To facilitate this process, IRRI has set up a rice bioinformatics Web page and E-mail list at http://www.cgiar.org/irri/bioinformatics/.

Rice genetics cooperative

A meeting of the Rice Genetics Cooperative (RGC) was held and the following members of the coordinating committee and various standing committees were nominated for 2001-05:

Coordinating committee

G.S. Khush, Chairman
A. Hirai, Co-Chairman
T. Sasaki, Secretary
D.S. Brar, Treasurer
M. Yano (Japan)
R. Wing (USA)
R.S. Paroda (India)
H.P. Moon (Korea)
Qifa Zhang (China)
M.D. Gale (United Kingdom)

Committee on genetic stocks

H. Satoh (Japan), Convener
Y. Sano (Japan)
M.T. Jackson (IRRI)
B.S. Dhillon (India)

Committee on Rice Genetics Newsletter

J. Bennett (IRRI), Convener
N. Kurata, Editor
R. Wu (USA)