AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Rice Genetics V

Edited by D.S. Brar, D.J. Mackill, and B. Hardy

RICE GENETICS V

Proceedings of the Fifth International Rice Genetics Symposium

RICE GENETICS V

Proceedings of the Fifth International Rice Genetics Symposium

The Philippines 19 – 23 November 2005

edited by

Darshan S Brar
David J Mackill
Bill Hardy

International Rice Research Institute, Philippines

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

The International Rice Research Institute (IRRI) was established in 1960 by the Ford and Rockefeller Foundations with the help and approval of the Government of the Philippines. Today, IRRI is one of the 15 nonprofit international research centers supported by the Consultative Group on International Agricultural Research (CGIAR – www.cgiar.org).

IRRI receives support from several CGIAR members, including the World Bank, European Union, Asian Development Bank, International Fund for Agricultural Development, Rockefeller Foundation, Food and Agriculture Organization of the United Nations, and agencies of the following countries: Australia, Canada, Denmark, France, Germany, India, Iran, Japan, Norway, People's Republic of China, Republic of Korea, Republic of the Philippines, Sweden, Switzerland, Thailand, United Kingdom, United States, and Vietnam.

The responsibility for this publication rests with the International Rice Research Institute.

RICE GENETICS V

Copyright International Rice Research Institute 2007

Mailing address: DAPO Box 7777, Metro Manila, Philippines
Phone: +63 (2) 580-5600
Fax: +63 (2) 580-5699
Email: irri@cgiar.org
Web: www.irri.org.
Rice Knowledge Bank: www.knowledgebank.irri.org
Courier address: Suite 1009, Security Bank Center
6776 Ayala Avenue, Makati City, Philippines
Tel. +63 (2) 891-1236, 891-1174, 891-1258, 891-1303

Suggested Citation:
Brar DS, Mackill DJ, Hardy B, editors. 2007. Rice genetics V. Proceedings of the Fifth International Rice Genetics Symposium, 19-23 November 2005, Manila, Philippines. Singapore: World Scientific Publishing and Los Baños (Philippines): International Rice Research Institute. 355 p.

ISBN-13978-971-22-0213-1

Cover design: Juan Lazaro IV
Page makeup and composition: Ariel Paelmo
Figures and illustrations: Ariel Paelmo

Printed in Singapore.

Contents

FOREWORD v

ACKNOWLEDGMENTS vii

RICE—A MODEL GENOME FOR CEREAL RESEARCH 1

Rice as a reference genome and more 3
R.L. Phillips, W.E. Odland, and A.L. Kahler

The complete rice genome sequence: a gold mine for future rice research 17
T. Sasaki, T. Matsumoto, J. Wu, and N. Namiki

Annotation of the rice genome 29
Shu Ouyang, Wei Zhu, J. Hamilton, Haining Lin, M. Campbell, Yuandan Lee, R.L. Malek, Aihui Wang, Qiaoping Yuan, B. Haas, J. Wortman, and C.R. Buell

STRUCTURAL GENOMICS AND RESOURCES 49

The Oryza map alignment project (OMAP): a new resource for comparative genomics studies within Oryza 51
R.A. Wing, H.R. Kim, J.L. Goicoechea, Y. Yu, D. Kudrna, A. Zuccolo, S.S. Ammiraju Jetty, M. Luo, W. Nelson, C. Soderlund, P. San Miguel, N. Gill, J. Walling, S. Jackson, B. Hurwitz, D. Ware, L. Stein, D. Brar, and D. Mackill

Analysis of oligo hybridization properties by high-resolution tiling microarrays in rice 65
Xiangfeng Wang, Lei Li, Viktor Stolc, Waraporn Tongprasit, Chen Chen, Jun Wang, Songgang Li, and Xing Wang Deng

Tissue culture–induced mutations and overexpression of full-length cDNAs as a tool for functional analysis of rice genes 77
H. Hirochika, A. Miyao, M. Yamazaki, A. Takahashi, G.K. Agrawal, C. Cheng, Y. Yamashita, M. Harada, H. Nakamura, M. Hakata, and H. Ichikawa

Analysis of genome sequences from the maternal and paternal parents of an elite rice hybrid 85
Jun Yu, Gane K.-S. Wong, Siqi Liu, Jian Wang, and Huanming Yang

DEVELOPMENTAL BIOLOGY AND GENE REGULATION 101

T-DNA tagging for developmental biology 103
G. An, D.-H. Jeong, S. An, and S. Park

Novel insights into the genomics of rice root adaptive development **117**
C. Périn, J. Rebouillat, A.M.C. Brasileiro, A. Diévart, P. Gantet, J.C. Breitler, A.A.T. Johnson, B. Courtois, N. Ahmadi, M. de Raissac, D. Luquet, M. Conte, D. This, P.K. Pati, Q.H. Le, D. Meynard, J.L. Verdeil, and E. Guiderdoni
Molecular signaling in disease resistance of rice **143**
K. Shimamoto, A. Nakashima, M. Fujiwara, N.T. Phuong, L. Chen, H.L. Wong, D. Miki, K. Imai, S. Maisonneuve, H. Takahashi, Y. Kawaguchi, S. Hirai, and T. Kawasaki

APPLIED GENETICS **153**
QTLs in rice breeding: examplesfor abiotic stresses **155**
D.J. Mackill, B.C.Y. Collard, C.N. Neeraja, R.M. Rodriguez, S. Heuer, and A.M. Ismail
Isolation of a QTL gene controlling grain number and QTL pyramidingto combine loci for grain number and plant height in rice **169**
M. Ashikari, S. Lin, T. Yamamoto, T. Takashi, A. Nishimura, E.R. Angeles, Q. Qian, H. Kitano, and M. Matsuoka
Genetic and molecular dissection of flowering time in rice **177**
M. Yano and T. Izawa
Understanding broad-spectrum durable resistance in rice **191**
J.E. Leach, R. Davidson, B. Liu, P. Manosalva, R. Mauleon, G. Carrillo, M. Bruce, J. Stephens, M.G. Diaz, R. Nelson, C. Vera Cruz, and H. Leung
Identification and transfer of trait-enhancing alleles from wild species **209**
S.R. McCouch, M. Sweeney, J. Li, H. Jiang, M. Thomson, E. Septiningsih, P. Moncada, J. Xiao, J. Coburn, E. Fraker, A. Garris, T. Tai, C. Martínez, J. Tohme, M. Sugiono, A. McClung, L.P. Yuan, and S.-N. Ahn
Genomics-based strategies for the development of “green super rice” **235**
Qifa Zhang
From gene to adaptation in rice **251**
K. Onishi and Y. Sano
Lessons from applying genomics to wheat and barley improvement **267**
P. Langridge
The major chromosome pairing locus (*Ph1*) in hexaploid wheat: a perspective **285**
G. Moore

FUNCTIONAL GENOMICS AND RICE IMPROVEMENT **299**
Functional genomics for genediscovery in abiotic stress response and tolerance **301**
Kazuo Shinozaki and Kazuko Yamaguchi-Shinozaki
Expression and functional analysis of rice genes involved in reproductive development and stress response **313**
A.K. Tyagi, J.P. Khurana, P. Khurana, S. Kapoor, V.P. Singh, A.K. Singh, J. K. Thakur, V. Gupta, S. Anand, S. Vij, M. Jain, S. Ray, P. Agarwal, R. Arora, P. Sharma, S. Mukherjee, A. Nijhawan, J. Giri, and R. Khurana
Designing and constructing novel gene promoters to generate stress-tolerant plants without yield penalty **331**
Tuan-hua David Ho, Chwan-Yang Hong, Ming-Tsair Chan, and Sumay Yu
Rice: an emerging model for plant systems biology **341**
A. von Zychlinski, S. Baginsky, and W. Gruissem

AN OVERVIEW: FIVE INTERNATIONAL RICE GENETICS SYMPOSIA (1985-2005) **353**
G.S. Khush

Foreword

Rice is the principal food of nearly half of humankind and more than 90% of it is grown in developing countries, where problems of food security are more acute. From being a poor cousin to maize, wheat, and tomato for genetic knowledge, as recently as the 1980s, rice has become a model plant and reference genome for molecular genetic research. During the last few decades, major progress has been made in increasing rice productivity. World rice production has more than doubled, from 257 million tons in 1966 to 600 million tons in 2006. This has mainly been achieved by applying principles of Mendelian genetics and conventional plant breeding methods. The current world population of 6.5 billion is likely to reach to 8.0 billion by 2030. To meet the growing food need and overcome malnutrition, rice varieties with higher yield potential, multiple resistance to stresses, and improved nutritional quality are needed. Recent advances in genetics featured in this symposium offer new opportunities to achieve these objectives.

The Fifth International Rice Genetics Symposium (IRGS-V) continues 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 a strong point of the third symposium. In the fourth symposium, progress on international efforts on sequencing the rice genome and developing novel genetic resources for structural and functional genomics were among the many highlights.

IRGS-V had 710 registered participants from 38 countries and featured 26 plenary lectures in six sessions, 54 contributory papers in eight concurrent sessions, and 380 poster presentations on different aspects of rice genetics. Renowned geneticists delivered plenary lectures covering a wide range of topics from classical genetics to the most advanced cutting-edge 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, genetic diversity, molecular mapping of genes/QTLs for biotic and abiotic stresses, single nucleotide polymorphisms and novel molecular markers, applied genetics, transformation, genome organization, gene isolation, regulation of gene expression, and functional genomics. The symposium also featured four workshops: on temperate rice, reproductive biology, *Oryza* map alignment and alien introgression, and genetics of insect resistance.

I would like to thank the organizing committee members (D.J. Mackill, D.S. Brar, H. Leung, J. Bennett, D. Macintosh, and B. Hardy), who devoted a great deal of time to organizing this symposium. In addition, IRRI would especially like to acknowledge the Rockefeller Foundation for its financial support for this symposium.

ROBERT S. ZEIGLER
Director General
International Rice Research Institute

Acknowledgments

We would like to extend our special thanks to Robert Zeigler, Ren Wang, and William Padolina for their scientific and financial support. We also thank the secretarial staff, particularly Elma Nicolas, Marlyn Rala, Yolanda Aranguren, Emily Alcantara, Minerva Bandian, Crisanta Culala, Fe Danglay, Nelie Delos Reyes, Leni Nazarea, Cecille Salonga, and Diane Martinez for their assistance. We appreciate the valuable help provided by the staff of Communication and Publications Services and Visitors and Information Services. Special thanks also go to R.K. Singh and his team of Plant Breeding, Genetics, and Biotechnology Division researchers, the Experiment Station, Safety and Security Services, and Transport Services for their valuable help in organizing the field tour at IRRI. We would also like to acknowledge the services provided by Marsman Travel.

Rice—a model genome for cereal research

Rice as a reference genome and more

R.L. Phillips, W.E. Odland, and A.L. Kahler

The rice (*Oryza sativa* L.) genome has become the reference genome to which others are compared. Part of the reason for this is that rice has the lowest DNA content of the common cereals and its gene content and gene order are found in other grass species used for food. Having the genome sequence of rice, both japonica and indica, allows comparisons with regard to genomic structure, gene constitution, and gene expression. Map locations for single-copy genes, families of genes, and quantitative trait loci (QTLs) are often compared among species, usually with rice as the reference. Specialized databases have been developed to facilitate cross-species homology relationships relative to genome and EST sequencing, protein structure, gene function, and other useful aspects. The evolutionary relationship of rice and several other cereals such as maize (*Zea mays* L.) and sorghum is clearly observed when highlighting syntenic regions. The colinearity of rice and American wildrice (*Zizania palustris*) has been exploited to develop a molecular genetic map and to locate QTLs in wildrice. The goal of this paper is to illustrate the value of rice for comparative genome referencing.

Keywords: comparative, duplication, map, polyploidy, synteny, QTLs, wildrice

Information on rice molecular genetics will help in attaining the 700 million tons of rice needed to feed the expected 650 million additional rice consumers in the next 20 years (Asian Biotechnology Development Review at www.ris.org.in/abdr.html). Determining the complete sequence of rice may be the most significant scientific milestone that has occurred in the last few decades. The high quality of sequence information for rice sets the standard for the sequencing of many other crop species—ushering in a new horizon in the plant sciences that will have a large impact on world food production. The sequencing of rice has not only established rice as a model for the grasses, but the knowledge being learned from related genomes also directly improves understanding of the rice genome. Studies of different organisms are revealing beneficial genes and pathways responsible for specialized characteristics.

Transgenic rice has been generated using genes from other species. Golden rice, for example, was engineered to produce beta-carotene (pro-vitamin A) with genes from the daffodil and *Erwinia* (Ye et al 2000). More recently, the beta-carotene content of golden rice has been substantially increased by using a maize phytoene synthase gene (*psy*) (Paine et al 2005).

Rice as a reference

"Reference allele" is a term used in maize genetics to denote the mutation by which the gene was identified, for example, *bz1-Ref*, and this allele becomes the one against which other alleles are compared (www.maizedb.org). Likewise, the genome of rice (*Oryza sativa* L.)—which is the first major crop as well as the first monocot to be completely sequenced (japonica: Goff et al 2002, International Rice Genome Sequencing Project 2005; indica: Yu et al 2002)—has become the one (i.e., the "reference genome") to which other plant genomes are compared, especially those of grasses. One reason for this is that rice has the lowest amount of DNA among the cereal species, with around 389 Mb. Other reasons are that the genetic behavior of rice is largely diploid, and a rich history exists of mutant discovery, linkage maps, and cytogenetics. Despite this extensive heritage of genetic information, much more is being learned about the rice genome structure and evolution and how it relates to other members of the grass (Poaceae) family.

The rice genome was presumed to be a simple grass genome because of its small size. Analysis of rice DNA sequences, however, revealed that it has a complex genome containing ancient segmental duplications and/or polyploidy. An early evaluation of rice sequencing information revealed that approximately 15% of the known rice genes were duplicated and resided in homologous colinear (syntenic) regions within the rice genome (Vandepoele et al 2003). Following the release of a more complete rice genomic sequence as well as coding sequences inferred by computational approaches, the amount of homoeologous duplication in the genome was found to be much higher at around 50% of the coding sequences (Wang et al 2005, Paterson et al 2003). A dating of the time period for the occurrence of the majority of these duplications has been estimated at about 70 million years ago. Wang et al (2005) also noted that a more recent segmental duplication occurred for chromosomes 11 and 12 about 5 million years ago, further illustrating the complexity of rice genome evolution.

The comparative maps of rice and maize presented by Ahn and Tanksley (1993) made it quite clear that these species share much of their genomic information. A restriction fragment length polymorphism (RFLP) marker-based comparison between a number of species by Moore et al (1995) showed that rice has a gene content and order similar to those of many other grass species. The circle diagram of chromosomes from several grass species emphatically illustrated relationships at the DNA level. Out of these comparisons, rice arose as the model species to which other species were referenced. The circle diagram has since been updated by Devos (2005). Although there are limitations, we now know that much can be learned through comparative genomics.

Rice and maize diverged from a common ancestor about 50 million years ago; comparisons between their genomes allow for insight into how a genome can evolve. Since the time of divergence, there has been a rapid change in gene content between rice and maize. For example, 22% of the unigenes identified in maize endosperm were not found in the rice genome (Lai et al 2004). In addition, an estimated 50% of the duplicated genes from the two progenitor species of maize have been lost in today's maize (Lai et al 2004). In a study of the Orp 1 and Orp 2 regions of maize and the orthologous regions of rice and sorghum, Ma et al (2005) found that only 40% of the genes are in the same order and orientation between sorghum and rice. The physical size of a region can be greatly variable, as seen in the orthologous Sh2/A1 region of wheat and barley that is approximately fourfold longer than the equivalent region in rice (Li and Gill 2002). At the Lr10 wheat and rice orthologous loci, differences in microsynteny could be interpreted as resulting from transposition, amplification, deletion, and inversion (Guyot et al 2004). A 300-kb sequence of barley was compared to the colinear region in rice (Caldwell et al 2004); although five orthologous genes were found, extensive transposon insertions, a translocation, and several gene duplications existed. Even though differences exist between genomes, a comparison of one or more species against each other to gain further information has been informative.

Comparisons beyond the grasses can provide additional evolutionary relationships. The latest japonica sequence draft has been compared with the sequence of *Arabidopsis* and it was found that about 71% of the genes are similar (International Rice Genome Sequencing Project 2005). Apparently because of an estimated 200 million years since a common ancestor (Wikström et al 2001), no large-scale synteny exists between rice and *Arabidopsis* (*A. thaliana*). Close analysis of the coding sequences in rice and *Arabidopsis* has revealed microsynteny to still be present between the two genomes, ranging from 4 to 11 homologous gene pairs (Vandepoele et al 2002). Interestingly, more than 2,800 rice genes could not be found in *Arabidopsis* (International Rice Genome Sequencing Project 2005). Comparison between two subspecies of rice, japonica and indica, identified polymorphism at 80,127 sites. This amount of polymorphism indicates that the two subspecies of rice are about 20 times more likely to differ than are two ecotypes (Columbia and Landsberg) of *Arabidopsis* (International Rice Genome Sequencing Project 2005). Since the sequencing of many of the large genomes may only be in the gene-rich regions (Martienssen et al 2004), the complete sequence of rice will be useful as a standard comparison, that is, a reference genome.

Finding orthologous genes and quantitative trait loci (QTLs)

The observed synteny between rice and several other grasses based on molecular genetic markers leads to the expectation that comparable genomic regions control related traits (Bennetzen and Ma 2003). One example is that homologous genes of *Aegilops tauschii* and rice controlling isoamylase have been located to syntenic regions (Rahman et al 2003). Several other examples of genes located in syntenic regions control related traits, including those related to domestication (Paterson et

al 1995). Of the estimated 37,544 protein-encoding genes in rice (International Rice Genome Sequencing Project 2005), it will be fascinating to learn what proportions are in similar genomic locations in the various species. Learning the function of each of these genes and the resulting phenotypes will be even more interesting. Currently, 1,488 genes from rice have associated phenotypes (www.gramene.org).

Genes controlling quantitative traits also can be expected to be located in orthologous regions among related species. Orthologous sequences to the rice heading-date gene (*Hd1*) were identified in perennial ryegrass (*Lolium perenne* L.) and meadow fescue (*Festuca pratensis* Huds.) by screening an *F. pratensis* BAC library with a marker physically near *Hd1* in rice (Armstead et al 2005). Candidate sequences were mapped to chromosome 7 in *L. perenne* and *F. pratensis*, a region syntenic with the *Hd1* on rice chromosome 6. Another example concerns a QTL (*Gnla*) in rice that apparently encodes the enzyme cytokinin oxidase/dehydrogenase (OsCKX2), which degrades the phytohormone cytokinin. This mutation on chromosome 1 results in reduced expression of OsCKX2 and increases the number of reproductive organs, resulting in enhanced grain yield (Ashikari et al 2005); transgenic tests supported this conclusion. By standard breeding, Ashikari et al (2005) combined the gene for greater grain number with the *sd1 (semi-dwarf)* gene to enhance the chances of achieving yield increases. Rice chromosome 1 shows regions of relatedness with chromosomes 3, 6, and 8 in maize, where several QTLs for grain yield have been mapped (Veldboom and Lee 1996, Austin et al 2000, Ho et al 2002; www.maizegdb.org). Some of the inflorescence architecture genes identified in maize map to the related locations as genes in millet species controlling primary branch number, branch density, and spikelet number (Doust et al 2004).

Finding published QTLs in orthologous regions will be much easier in the future with the Gramene database. As stated on the Gramene Web site (www.gramene.org), "Gramene is a curated, open-source, Web-accessible data resource for comparative genome analysis in the grasses. Our goal is to facilitate the study of cross-species homology relationships using information derived from public projects involved in genomic and EST sequencing, protein structure and function analysis, genetic and physical mapping, interpretation of biochemical pathways, gene and QTL localization, and descriptions of phenotypic characters and mutations." The Gramene Comparative Map Viewer (CMap) allows one to view various types of maps (sequence, genetic, QTL, and Fingerprint Contig) and make comparisons among them. More than 8,400 QTLs are available for comparison in the Gramene database. For example, the various fusarium head-blight QTLs reported in barley and wheat along with their map positions are presented in a way that comparisons can be made to the rice genome.

The massive amounts of data generated through the rice genome sequencing project necessitated the development of bioinformatics approaches and tools. The Institute for Genome Research (TIGR) has started a bioinformatics approach to decipher and annotate the gene sequences in rice. To date, 7,226 coding sequences (www.TIGR.org) have been identified in the rice genome. Historically, researchers working on a particular species have assigned a unique name to a gene or locus of interest. As comparative genomics projects have attempted to use the rice data, it has become ap-

parent that discussing sequence and functional homology between species can become confusing. The history of using different names for the same characteristic makes such comparisons quite difficult; a standard ontology is needed in order to make useful comparisons across species. Fortunately, Gramene will include a common ontology across grass species. Information is provided on cDNAs, proteins, various genetic and physical maps, mutant phenotypes, and QTLs (Yamazaki and Jaiswal 2005). Another resource for rice-based comparative genomics is the BGI-RIS (Beijing Genomics Institute-Rice Information System), which makes available sequence information on indica rice, japonica rice, and other cereal species (Zhao et al 2004).

Value of rice for cloning genes underlying QTLs

Only a few QTLs have been cloned in plants: *Arabidopsis* (4 QTLs), maize (1), rice (5), and tomato (3) (see Salvi and Tuberosa 2005, Ashikari et al 2005). Because of the large cost to clone a gene underlying a QTL, verification of major QTLs across species would help to identify those QTLs that may be the most important. Then using rice to clone the pan-species QTLs would seem to be the most efficient approach now. As a cautionary note, several populations will likely need to be examined in any one species, since a QTL will be detected only in those crosses with different alleles at the locus. This is the strategy suggested by Tuberosa et al (2003) for root traits: "Our long-term goal is to identify major QTLs controlling root traits in maize and clone the genes underlying such QTLs using rice as a model species." As a first step toward cloning, they placed emphasis on QTLs for root traits in maize and rice that appeared to be in syntenic regions. Major QTLs for root traits in maize were found on five chromosomes. The maize QTL data were then compared to more than 400 QTLs for root traits in seven rice populations. Near-isogenic lines of maize are being developed for one of the major pan-species QTLs. The rice genomic information will indicate possible candidate genes and nearby markers. We propose that a useful term for this approach would be "comparative genome referencing." Methodology for more efficiently determining QTLs or candidate genes is becoming available, such as from breeding information (Yu et al 2005), association mapping (Buckler and Thornsberry 2002), or expression profiling (DeCook et al 2006).

To effectively use pan-species QTLs and other genetic information, the relationships of the various genomes must be defined. Using rice as a reference genome to itself has led to an understanding of genome duplications and their evolution. The duplications of rice have lost 30–65% of the homologous genes between its macro-duplications (Wang et al 2005). Knowing that genomic duplicates diverge, it is important to define all related regions to properly search for homologous genes or related traits. Assuming that the duplications within a genome are related and that syntenic regions between genomes are also related, homologies can be defined among the duplicates of a syntenic region (Fig. 1). Understanding the complexity of duplications will improve candidate gene searches and the understanding of how genomes evolve. Rice is playing a pivotal role in understanding pan-species relationships.

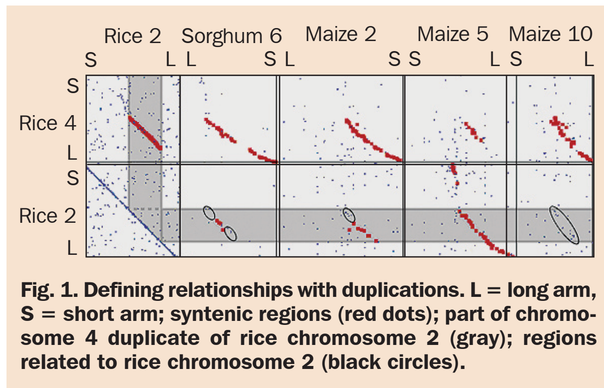


Fig. 1. Defining relationships with duplications. L = long arm, S = short arm; syntenic regions (red dots); part of chromosome 4 duplicate of rice chromosome 2 (gray); regions related to rice chromosome 2 (black circles).

Evolutionary changes can mask relationships between genomic sections. Depicted in Figure 1 are homologous sequences within and between rice (Osa1 pseudomolecules), sorghum (Bowers et al 2003), and maize (IBM2 2004 neighbors genetic map) that have an E value of less than 1 e^{-10} by BLAST analysis (Altschul et al 1990). Syntenic regions, resolved by DiagHunter (Cannon et al 2003), are shown with enlarged red dots. The rice portion of chromosome 4 that is a duplicate of rice chromosome 2 is highlighted in gray. Black circles draw attention to regions related to rice chromosome 2 defined by synteny to its duplicate copy on rice chromosome 4. The location of pan-specific orthologous genes can be predicted from such analyses.

American wildrice as a case study

The beginning of the comparative genomics era was rooted in the fact that RFLP markers developed in one species were found to be useful across various grass species (Moore et al 1995). As cDNA libraries, and later RFLP markers from the libraries, were produced from various grasses, comparative genetic maps revealed highly conserved genome regions. A set of RFLP markers chosen for their ability to be mapped across the grass species was identified using the rice genome as a reference (Van Deynze et al 1998). This set of "anchor markers" became a significant tool for assisting other grass mapping projects. The high level of colinearity between grass species as illustrated by the circle diagram of Gale and Devos (1998) has led to the use of rice genetics information in newly established grass molecular genetics projects.

Molecular markers have proven useful as a tool for plant improvement, for example, in marker-assisted backcrossing (Chen et al 2000). Rice genetics programs were among the first to embrace the use of RFLP markers in germplasm improvement and breeding programs (McCouch et al 1988). RFLP markers were initially the markers of choice due to their codominant expression, which allows the identification of heterozygous individuals within a population. The availability of the rice genome

sequence has advanced the development and use of molecular markers (McCouch et al 2002). PCR-based molecular markers can be efficiently designed from rice DNA. American wildrice (*Zizania palustris*) is one example of a species that is poised to take advantage of the availability of the rice genome sequence and other molecular tools. This aquatic plant is a diploid (2n=2x=30) (Elliot 1980) that was once thought to be closely related to maize based on its architecture. This conclusion was based largely on phenotypic observation that wildrice has separate pistillate and staminate flowers like maize rather than perfect flowers like rice.

The first wildrice linkage map was constructed in the late 1990s (Kennard et al 2000). American wildrice is unusual in that 99% of the markers on the RFLP molecular map were previously mapped in *Oryza sativa*. Therefore, this first molecular map is also a comparative genetic map. Early research in wildrice genetics showed *O. sativa* and *Z. palustris* to be highly similar based on genomic DNA hybridization (Kennard et al 2000). The results of the mapping study showed that the two species have colinearity with 80–90% of the mapped markers. This comparison of genetic data between rice and wildrice gives insight into the evolution of the species. Wildrice has three additional chromosome pairs compared with rice, reflecting a possible duplication of three rice chromosomes or chromosome segments or polyploidization followed by chromosome loss (Kennard et al 2000).

The availability of mapped marker loci in wildrice led to a QTL project to find and map important agronomic trait loci (Kennard et al 2002). One result from the study was the discovery of three QTL markers for the shattering trait (Fig. 2) and the development of an inheritance model. The major QTL marker accounted for approximately 39% of the shattering variation in an F_2 mapping population. Table 1 (adapted from Kennard et al 2000) lists traits in wildrice for which a marker interval was identified as a QTL and a known rice QTL mapped in a nearby colinear region.

The marker associated with the major shattering QTL in wildrice was a cDNA probe (UMC305) from maize (Fig. 3). PCR primers designed from the UMC305 maize sequence (www.ncbi.nlm.nih.gov) did not amplify wildrice DNA. As rice sequence data became available, the region homologous to UMC305 was identified in rice. This *in silico* work has allowed for the development of PCR primers, using rice DNA sequence, that amplify wildrice DNA. The chromosome 2 region of wildrice containing the major shattering QTL was compared with the syntenous region of rice chromosome 2 (Fig. 3). Colinear markers were used to align the two maps and identify rice simple sequence repeat (SSR) markers to test in wildrice.

A mapped rice SSR marker has been identified in the rice genomic region that is colinear to the major wildrice QTL. While the SSR marker did not segregate in a wildrice F_2 mapping population, it did show polymorphism corresponding to the nonshattering trait in a wildrice breeding population.

A set of PCR-based SSR markers from rice has been adapted for use in wildrice. Approximately 2,700 SSR markers have been designed in rice (McCouch et al 2002). A subset of 330 markers from across the rice genome was used to determine the usefulness of the rice markers in wildrice. In total, 89 of the 330 markers tested were polymorphic and fit the expected segregation ratio. To date, 40 of these 89 poly-

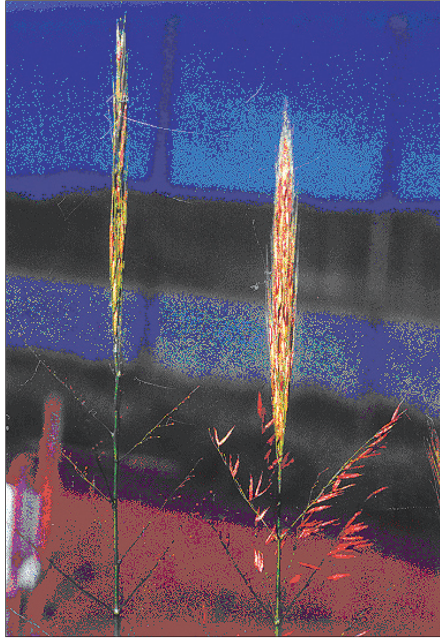


Fig. 2. American wildrice panicles illustrating the shattering phenotype (left) with loss of seed and staminate flowers (left, bottom portion) and nonshattering (right) with retention of seed and lower staminate flowers.

morphic markers have been added to the existing wildrice map (Kahler, unpublished). The successful use of rice sequence data as a reference for wildrice opens the door to advancing wildrice molecular genetics and its application relatively quickly and without exorbitant expense. Marker-assisted selection of wildrice for nonshattering is a goal of the wildrice molecular genetics project. With the use of rice genome information, the wildrice genetics program has been able to develop and implement a molecular marker program. This is a tremendous advance for a niche crop with a small research group. The wildrice molecular genetics program is an example of the usefulness of the rice genome sequence and molecular genetic and bioinformatics tools.

References

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

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410

Table 1. Mapped wildrice QTLs and putative orthologous rice loci. Traits in wildrice for which a marker interval was identified as a QTL and a known rice QTL mapped in a nearby colinear region.

Trait	Wildrice QTL marker	Locus in rice colinear region
Seed shattering	UMC305	Consensus QTL (Paterson et al 1995)
	CDO244	*Sh-3* (Matsuo et al 1997), Shattering 3 (Yoshimura et al 1997), qSHT-4 (Cai and Morishima 2000)
Heading date	CDO78	*Se-1* (Laurie 1997)
Plant height	CDO328	*d-18* (Yoshimura et al 1997)
	RZ730b	*d-10* (Yoshimura et al 1997), *sd-1* (Matsuo et al 1997)
	RG139b	*d-5* (Yoshimura et al 1997), *d-3* (Yoshimura et al 1997)
Tiller number	RZ237	*d-10* (Matsuo et al 1997)
Stem color	C122	*A* (Matsuo et al 1997)
	CDO920	*A* (Matsuo et al 1997, Yoshimura et al 1997)
Flower color	RZ475b	*A* (Yoshimura et al 1997)
Plant habit	RZ448	*dl* (Matsuo et al 1997, Yoshimura et al 1997)
Seed length	RZ244b	Consensus QTL (Paterson et al 1995)
Seed number	RZ2b	*Dn-1* (Matsuo et al 1997, Yoshimura et al 1997)
Panicle length	RZ557	*sp* (Matsuo et al 1997, Yoshimura et al 1997)

Source: adapted from Kennard et al (2000).

Armstead IP, Skot L,Turner LB, Skot K, Donnison IS, Humphreys MO, King IP. 2005. Identification of perennial ryegrass (*Lolium perenne* L.) and meadow fescue (*Festuca pratensis* Huds.) candidate orthologous sequences to the rice *Hd1 (Swe1)* and barley *HvCO 1 CONSTANS*-like genes through comparative mapping and microsynteny. New Phytol. 167:239-247.

Ashikari M., Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles E, Qian Q, Kitano H, Matsuoka M. 2005. Cytokinin oxidase regulates rice grain production. Science 309:741-745.

Austin DF, Lee M, Veldboom LR, Hallauer AR. 2000. Genetic mapping in maize with hybrid progeny across testers and generations. Crop Sci. 40:30-39.

Bennetzen JL, Ma JX. 2003. The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr. Opin. Plant Biol. 6:128-133.

Bowers J, Abbey C, Anderson S, Chang C, Draye X, Hoppe A, Jessup R, Lemke C, Lennington J, Li Z, et al. 2003. A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367-386.

Buckler ES IV, Thornsberry J. 2002. Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5:107-111.

Cai HW, Morishima H. 2000. Genomic regions affecting seed shattering and seed dormancy in rice. Theor. Appl. Genet. 100:840-846.

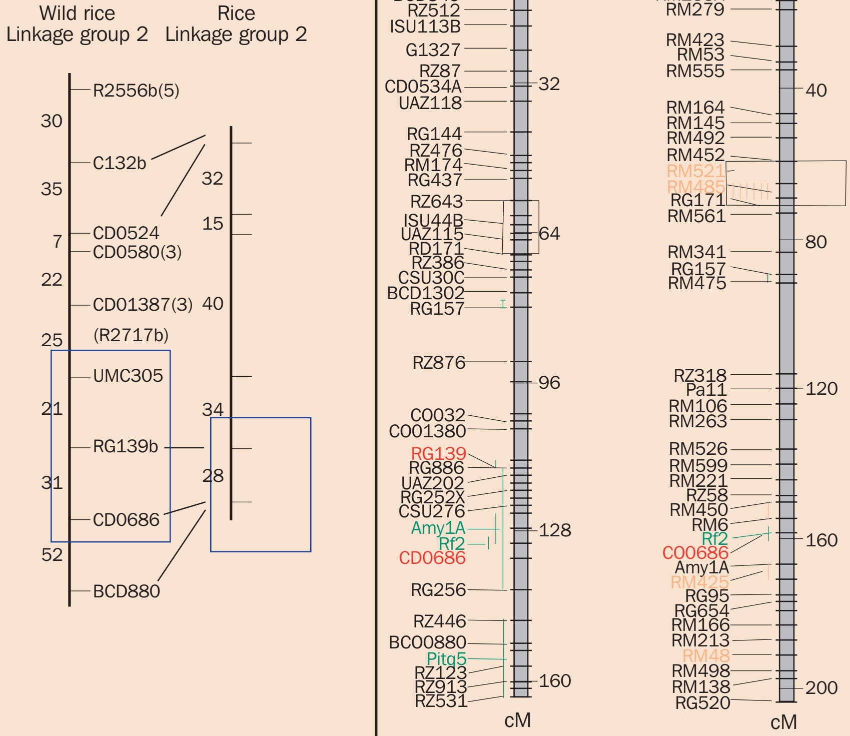


Fig. 3. CMap tool (www.gramene.org) used to identify potentially useful rice SSR markers using the wildrice comparative genetic map. The chromosome 2 region of wildrice containing the major shattering QTL (Kennard et al 2002) was compared with the syntenous region of rice chromosome 2. Colinear markers were used to align the two maps and identify rice SSR markers to test in wildrice.

Caldwell KS, Langridge P, Powell W. 2004. Comparative sequence analysis of the region harboring the hardness locus in barley and its colinear region in rice. Plant Physiol. 136:3177-3190.
Cannon SB, Kozik A, Chan B, Michelmore R,Young ND. 2003. DiagHunter and GenoPix2D: programs for genomic comparisons, large-scale homology discovery and visualization. Genome Biol. 4:R68.
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.
DeCook R, Lall S, Nettleton D, Howell S. 2006. Genetic regulation of gene expression during shoot development in *Arabidopsis*. Genetics 172:1155-1164.
Devos KM. 2005. Updating the 'crop circle'. Curr. Opin. Plant Biol. 8:155-162.
Doust AN, Devos KM, Gadberry MD, Gale MD, Kellogg EA. 2004. Genetic control of branching in foxtail millet. Proc. Natl. Acad. Sci. USA 101:9045-9050.
Elliot WA. 1980. Wild rice. In: Fehr WR, Hadley HH, editors. Hybridization of crop plants. Madison, Wis. (USA): American Society of Agronomy. p 721-731.
Gale MD, Devos KM. 1998. Plant comparative genetics after 10 years. Science 282:656-659.
Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100.
Guyot R, Yahiaoui N, Feuillet C, Keller B. 2004. *In silico* comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1 AS and rice chromosome 5S. Funct. Integr. Genomics 4:47-58.
Ho JC, McCouch SR, Smith ME. 2002. Improvements of hybrid yield by advanced backcross QTL analysis in elite maize. Theor. Appl. Genet. 105:440-448.
International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793-800.
Kennard WC, Phillips RL, Porter RA, Grombacher AW. 2000. A comparative map of wild rice (*Zizania palustris* L. 2n=2x=30). Theor. Appl. Genet. 101:677-684.
Kennard W, Phillips R, Porter R. 2002. Genetic dissection of seed shattering, agronomic, and color traits in American wildrice (*Zizania palustris* var. interior L.) with a comparative genetic map. Theor. Appl. Genet. 105:1075-1086.
Lai J, Dey N, Kim CS, Bharti AK, Rudd S, Mayer KFX, Larkins BA, Becraft P, Messing J. 2004. Characterization of the maize endosperm transcriptome and its comparison to the rice genome. Genome Res. 14:1932-1937.
Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, Llaca V, Tanyolac B, Park Y-J, Jeong O-Y, Bennetzen JL, Messing J. 2004. Gene loss and movement in the maize genome. Genome Res. 14:1924-1931.
Laurie DA. 1997. Comparative genetics of flowering time. Plant Mol. Biol. 35:167-177.
Li W, Gill BS. 2002. The colinearity of the Sh2/A1 orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the Triticeae. Genetics 160:1153-1162.
Ma JX, San Miguel P, Lai JS, Messing J, Bennetzen JL. 2005. DNA rearrangement in orthologous Orp regions of the maize, rice and sorghum genomes. Genetics 170:1209-1220.
Martienssen RA, Rabinowicz, PD, O'Shaughnessy A, McCombie WR. 2004. Sequencing the corn genome. Curr. Opin. Plant Biol. 7:102-107.
Matsuo T, Yuzo F, Fumio K, Hikoyuki Y. 1997. Science of the rice plant. Vol. III. Genetics. Tokyo (Japan): Food and Agriculture Research Center.

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.

McCouch S, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, et al. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res. 9:193-207.

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

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R. 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nature Biotechnol. 23:482-487.

Paterson AH, Bowers JE, Peterson DG, Estill JC, Chapman BA. 2003. Structure and evolution of cereal genomes. Curr. Opin. Genet. Dev. 13:644-650.

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714-1718.

Rahman S, Nakamura Y, Li Z, Clarke B, Fujita N, Mukai Y, Yamamoto M, Regina A, Tan Z, Kawasaki S, Morell M. 2003. The sugary-type isoamylase gene from rice and *Aegilops tauschii*: characterization and comparison with maize and *Arabidopsis*. Genome 46:496-506.

Salvi S, Tuberosa R. 2005. To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci. 10:297-304.

Tuberosa R, Salvi S, Giuliani S, Bellotti M, Sanguineti MC, Stefanelli S, Landi PI. 2003. A comparative analysis of QTLs for root traits in maize and rice: results and perspectives for QTL cloning using rice as a model. Am. Soc. Plant Biol. Abstr. 923, p 58.

Vandepoele K, Saeys Y, Simillion C, Raes J, Van De Peer Y. 2002. The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between *Arabidopsis* and rice. Genome Res. 12:1792-1801.

Vandepoele K, Simillion C, Van De Peer Y. 2003. Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15(9):2192-2202.

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.

Veldboom LR, Lee M. 1996. Genetic mapping of quantitative trait loci in maize in stress and non-stress environments. 1. Grain yield and yield components. Crop Sci. 36:1310-1319.

Wang X, Shi X, Hao B, Ge S, Luo J. 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol. 165:937-946.

Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. Proc. R. Soc. Lond. B Biol. Sci. 268:2211-2220.

Yamazaki Y, Jaiswal P. 2005. Biological ontologies in rice databases: an introduction to the activities in Gramene and Oryzabase. Plant Cell Physiol. 46:63-68.

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

Yoshimura A, Ideta O, Iwata N. 1997. Linkage map of phenotype and RFLP markers in rice. Plant Mol. Biol. 35:49-60.

Yu J, Hu S, Wang J, Wong G, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79-92.

Yu J, Pressoir G, Holland JB, Kresovich S, Buckler ES. 2005. Bridging molecular diversity and functional diversity: complex traits dissection in maize. Crop Sci. Soc. Am. Abstr. 256-257.

Zhao W, Wang J, He X, Huang X, Jiao Y, Dai M, Wei S, Fu J, Chen Y, Ren X, et al. 2004. BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. Nucl. Acids Res. 1:32.

Notes

Authors' address: Department of Agronomy and Plant Genetics and Microbial and Plant Genomics Institute, University of Minnesota, St. Paul, MN 55108.

The complete rice genome sequence: a gold mine for future rice research

T. Sasaki, T. Matsumoto, J. Wu, and N. Namiki

The map-based complete rice genome sequence is now freely available to researchers worldwide, providing the most fundamental tool that should further accelerate efforts to improve the staple crop that feeds more than half the world's population. The finished-quality sequence covers 95% of the 389-Mb genome, including virtually all of the euchromatin and two complete centromeres. A total of 37,544 non-transposable-element-related protein-coding genes were identified. The complete genetic information on rice will serve as a gold mine for genomic research in rice and other cereal species. It will facilitate the identification of many important genes by both forward and reverse genetics strategies, and clarify the relationships between sequence variation and phenotypes. The genome sequence derived from *Oryza sativa* subspecies japonica can be used as a reference sequence for comparative analysis among *Oryza* species that will help in understanding the major factors involved in speciation and searching for useful genetic resources. Furthermore, the completed sequence will also serve as a standard for cereal genome comparison and identification of rice orthologous genes in other cereal crops, thereby providing a platform for establishing the genomics of each cereal species.

Keywords: rice, genome sequence, physical maps, tandem-arrayed genes, transposable elements

Rice breeding has a long historical background although the exact time and origin of the start of rice cultivation are still uncertain. It is believed that domestication of rice took place independently in several Asian countries. Some of the oldest evidence was discovered from archaeological artifacts in southern China, which date back to more than 10,000 years ago. The present cultivated rice belongs to two major subspecies of *Oryza sativa*: japonica and indica. In the beginning, early rice growers must have been simply interested in cultivating rice plants with large grains and high yield. Eventually, the needs of local people for varieties with good quality and the demand to continuously increase yield must have also driven farmers to make continuous efforts to improve rice

plants. Furthermore, the wide range of geographic and climate conditions and cultural practices associated with rice cultivation may have contributed to the development of many varieties carrying tolerance of biotic and abiotic stresses, as well as varieties well suited to regional cooking and eating habits of the local people. Therefore, the basic principle of breeding that involves crossing of two strains with complementary favorable characteristics, followed by selection for progenies with more favorable phenotypes, must have been widely practiced even before the rule of inheritance was established based on Gregor Mendel's experiments almost 140 years ago.

Although well-experienced breeders and farmers can easily distinguish improved agronomic traits from among many progenies, the frequency of successfully finding a desired trait is oftentimes rare. This is partly because we have very limited knowledge of the relationship between the phenotype and gene. Although most traits are simply controlled by a single gene, many traits are also controlled by multigenes and, in special cases, some traits are not even controlled by the transcript itself but by the promoter sequence. These examples have been reported in the model weedy plant, *Arabidopsis thaliana*, upon completion of the genome sequence (Arabidopsis Genome Initiative 2000). This reflects the remarkable development of forward and reverse genetics relying on genome sequence information. For rice improvement, undoubtedly, we also need detailed genome sequence information to analyze the relationship between phenotype and gene, particularly because most of the important agronomic traits are controlled by quantitative trait loci (McMullen 2003). The difficulty in improvement of quantitative traits resides not only in the complexity of inheritance but also in the mode of interaction of the component genes in QTLs, which remains unknown. The use of genome sequence information in conjunction with strategies for genetic analysis will ensure success in finding favorable progenies under a strategic crossing and selection program.

Systematic genetic analysis of rice was launched in the early 1990s in several countries such as Japan, China, the United States, and partly in collaboration with the International Rice Research Institute (IRRI). A series of molecular markers based on restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and microsatellites have been detected in nucleotide sequences and successfully used to construct genetic maps based on recombination frequency. The co-dominant RFLP markers were most commonly used for mapping and especially most trusted to give reproducible data. The high-density genetic map of rice has a remarkable impact not only for understanding the genome structure of rice but also for elucidating the structure of other cereal genomes (Harushima et al 1998). The segmentally conserved order of DNA markers has been elucidated among cereal crops and the utility of the rice genome as a reference for other cereals has become widely recognized (Devos 2005).

The great potential for rice as an agronomic and model plant has encouraged rice researchers to pursue genome analysis beyond genetic and physical mapping. A workshop was held in Singapore in September 1997 to discuss the feasibility of an international collaborative project. The International Rice Genome Sequencing Project (IRGSP) was subsequently organized with the aim of completely decoding the entire

rice genome. Although funding for participating countries was not yet committed at that time, the participants were optimistic because rice is not only an important plant for scientific research but also an economically valuable crop and main food source for almost half of the world population. As expected, the IRGSP was supported by government funds from participating countries as well as private and nonprofit organizations that had strong interest in production and research. With the active cooperation of participating groups, the IRGSP finished the complete and high-quality sequence of the entire rice genome by the end of December 2004 (http://rgp.dna.affrc.go.jp/IRGSP/).

Overview of the rice genome structure

Elucidating the genome sequence of an organism has been simplified with continuous technological advances using a high-throughput automated nucleotide sequence analyzer. A whole-genome shotgun sequencing approach has further accelerated the sequencing of entire genomes. However, if the collected sequences are not correctly assigned to their original genomic position, the utility of the sequence is not satisfactory. In the case of bacteria with a genome size of several mega base pairs, a random collection of untargeted sequences could be easily assembled using a powerful computer system to accurately reconstruct the original genome (Fleischmann et al 1995). On the other hand, higher organisms of more than 100 Mb and rich in repetitive and similar sequences such as transposable elements are very difficult to reassemble and compare to the original genome sequence. For such an important crop as rice, the genome sequence is much anticipated for practical application in rice breeding, particularly in the identification of a difference in nucleotide sequence that is reflected as the difference in phenotype among various cultivars. This difference is a relative characteristic and is identified only by comparison with the standard sequence. Thus, the IRGSP adopted the map-based clone-by-clone sequencing strategy that involves constructing a physical map based on the genetic map, developing a library of genomic clones, anchoring the clones to a specific location identified from the map, and sequencing the clones. Although this approach is a lot more expensive and time-consuming, the resulting sequence is highly accurate. The public availability of the genome information provides the scientific community with a standard sequence necessary for identification and association of each functional gene to a target phenotype.

The physical map constructed by the IRGSP is shown in Figure 1 with the position of physically identified centromeres and telomeres. The sizes of the remaining gaps have been measured by a fiber-FISH method except for several large gaps around the centromere. As a result, the gap sizes are revealed within the range of a small number of BAC/PAC clones. Although a total of nearly 60 times coverage of the rice genome has been used for physical mapping, some regions remain as gaps. This might be attributed to the existence of an unusual sequence that could not be harbored by the host bacteria. The genome size of japonica rice variety Nipponbare is now estimated as 388.8 Mb (International Rice Genome Sequencing Project 2005). This value is less than that previously reported by analysis with flow cytometry, which

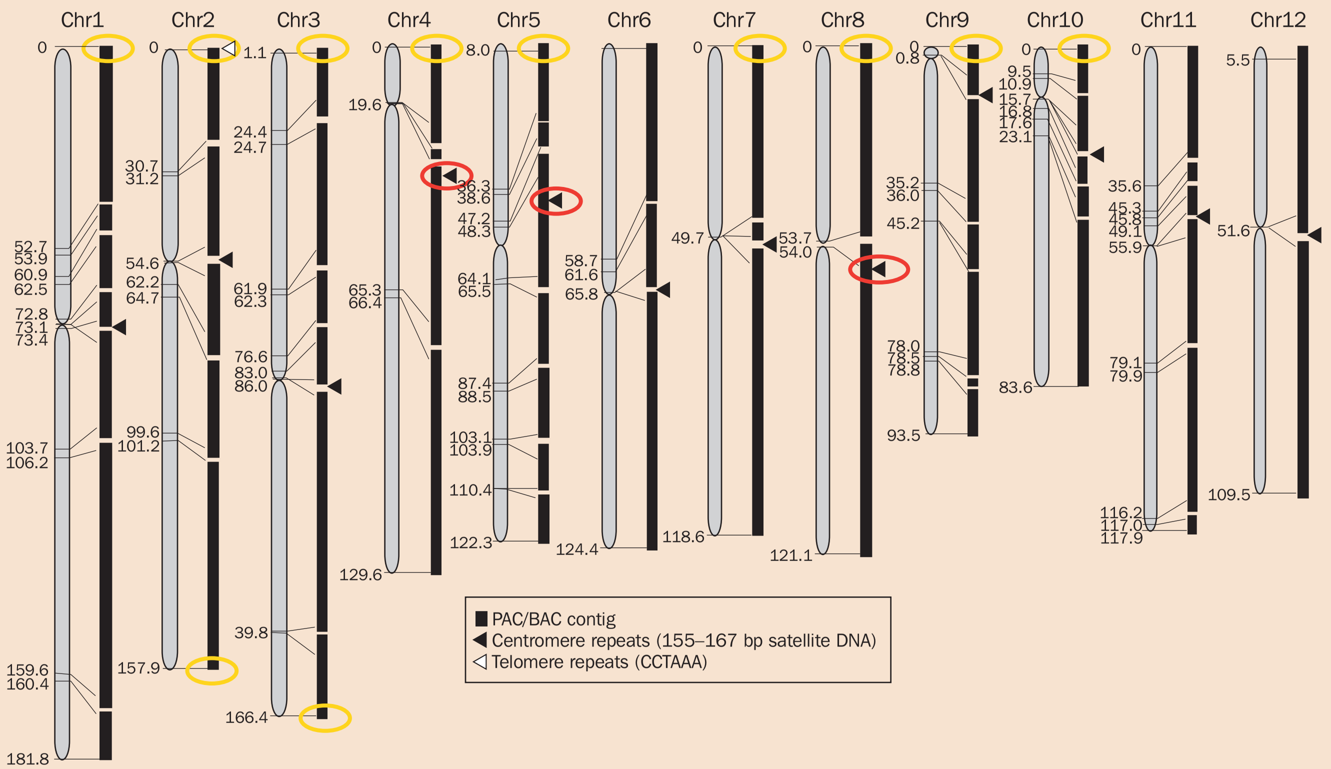


Fig. 1. Physical maps of the 12 rice chromosomes. For each chromosome, the genetic map is shown on the left and the PAC/BAC contigs on the right. The numbers of the markers flanking PAC/BAC contigs indicate the genetic distances on the genetic map. Black bars indicate the physical contigs. Red and yellow circles show the centromere and telomere regions, respectively, that have been successfully sequenced.

gives a relative value to a reference species (Arumuganathan and Earle 1991). The analysis of the map-based genome sequence provides an absolute value of the genome size of the target species.

Each of the 12 rice chromosomes has a characteristic genome structure. The rice chromosomes range in size from 45 Mb (chromosome 1) to 24 Mb (chromosome 10) and the gene density varies from 8.7 (chromosome 3) to 11.6 (chromosome 12) genes per kilobase pair. The nuclear organizing region is located at the end of the short arm of chromosome 9 and its genomic structure is elucidated in detail. Three centromeres among 12 chromosomes have been completely mapped by BAC/PAC clones and two of three centromeres (chromosomes 4 and 8) have been completely sequenced (Wu et al 2004, Zhang et al 2004). Both structures indicate a remarkable difference in distribution of rice centromere-specific and common repeats. The CentO repeat units are scattered within the centromere of chromosome 4 and clustered within the centromere of chromosome 8. The centromere of chromosome 5 remains to be sequenced and may add an important clue for completely understanding the complex structures and evolution of the 12 rice centromeres. The segmental duplication between chromosomes may also provide a resource for analysis of the process of evolution of rice chromosomes. Based on the gene models generated by the FGENESH prediction program, a parallel and an antiparallel duplicated region between chromosomes 1 and 5 were also identified in addition to the previously characterized duplicated distal ends of chromosomes 11 and 12. Although at least one region in each chromosome was distinctly duplicated in another chromosome, segmental regions in triplicate or more were not identified in any chromosomes. Moreover, segmental duplication in rice is much simpler than that in *Arabidopsis*, which has a high-frequency segmental duplication among the five chromosomes. For example, almost all of the chromosomal segments of the long arm of chromosome 4 have counterparts in the remaining four chromosomes. The difference in duplication pattern between rice and *Arabidopsis* may be associated with the pressure of selection during the breeding process that may have induced alteration in the corresponding genome structure. The relationships between the mechanism of chromosomal duplication and evolution will be facilitated by the map-based sequences of both species and other related species using the standard genome sequence.

In addition to the chromosomal segment duplication, tandem gene multiplication is frequently observed in the rice genome. A total of 47 copies of receptor-like kinase genes were observed between 0.2 Mb and 5.1 Mb from the top of the short arm of chromosome 1 (Fig. 2). Some of these multiplied genes are disrupted by transposable elements to produce pseudogenes. The frequency of multiple-copied tandem genes in rice (14%) is slightly lower than that in *Arabidopsis* (17%). The total predicted gene models in rice and *Arabidopsis* are 37,544 and 25,498, respectively. This difference does not necessarily reflect the degree of gene multiplication even if the comparison of both predicted gene models indicates the existence of 90% of rice homologous counterparts of *Arabidopsis* and 71% of *Arabidopsis* homologues of rice. When using the same criteria of tandem repeats in both species or tandem arrangement in 5-Mb intervals, 29% of the predicted rice genes (10,837) are amplified at least once

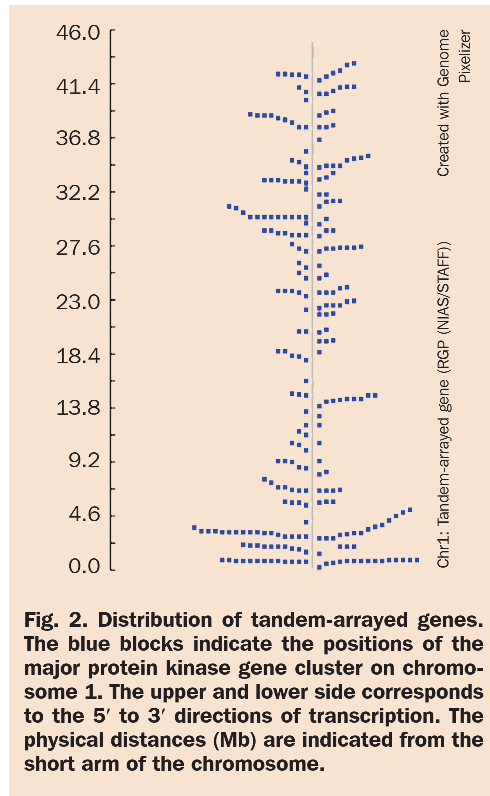


Fig. 2. Distribution of tandem-arrayed genes. The blue blocks indicate the positions of the major protein kinase gene cluster on chromosome 1. The upper and lower side corresponds to the 5′ to 3′ directions of transcription. The physical distances (Mb) are indicated from the short arm of the chromosome.

in tandem. This observation of tandem repeats of genes may provide insights into the process of evolution and the mechanism of gene duplication. In addition, analysis of the expression mode of each tandem gene could reveal the specified role of the promoter for each gene.

The rice genome contains many types of transposable elements, a major component of genome constructs that provides historical evidence of genome evolution and divergence (Vitte et al 2004). Class I transposable elements, or retro-elements such as Ty1/*copia* and Ty3/*gypsy* and non-LTR or LINEs and SINEs, occupy a total of 19.3% of the genome. Class II transposable elements, or DNA transposons such as IS5/*tourist* and IS256/*mutator*, occupy a total of 13% of the genome. When adding other types of transposons, almost 34.8% of the rice genome corresponds to transposons and retroransposons. Although the biological role of transposons is still under argument, diversification of species may be partly attributed to transposition or insertion of transposons into a new genomic position. Therefore, a thorough analysis of rice transposons will help in understanding the origin of difference in genome structure among Oryzeae and give important insight into the history of the evolution of this genus.

Another important structural component of the genome is simple sequence repeats (SSRs). SSRs can be very useful for genetic analysis of target traits if their genomic positions are clearly identified. As a co-dominant marker, SSR polymorphism can be easily detected by PCR using flanking sequences as primers. In the elucidated Nipponbare genome sequence, a total of 18,828 class 1 SSRs, or perfect repeats with more than 20 nucleotide bases in length, were detected. Analysis of the distribution of SSRs in other rice varieties may provide additional resources for genetic mapping of many agronomic traits and identification of target genes.

Single nucleotide polymorphism (SNP) is the most abundant polymorphism in the genome. SNPs could be found in both intron and exon regions that affect the translated product or amino acid sequence and induce alteration of normal function. In some cases, this may result in a reduction in enzyme activity and partial or total loss of function. This indicates that SNPs may be important in the expression of quantitative traits that are controlled by a combination of multiple genes. In the case of QTLs involved in rice heading date, SNPs have been identified to play the key roles in the adaptation of heading date to daylength at a certain latitude. Although a genome-wide survey of SNPs among rice varieties is currently unrealistic, the identification of SNPs within the target genomic region of a QTL is necessary for efficient breeding of quantitative traits. For this purpose, the National Institute of Agrobiological Sciences (NIAS) has recently launched a program aimed at elucidating the correlation between sequence variation among representative rice varieties and trait indices in rice.

Genome-wide comparison of SNPs and insertion/deletion (In/Del) is important and indispensable to clarify the difference in genome structure between japonica and indica rice. So far, accurate map-based genome sequence information is available only for japonica variety Nipponbare. However, a Chinese indica variety, 93-11, has been sequenced by a whole-genome sequencing method that could provide an important resource for sequence comparison of genic regions. Another sequence comparison was made using the BAC-end sequence data of indica variety Kasalath, which is one of the parent cultivars of the F_2 progenies used to construct the high-density genetic map of rice. Although a BAC-end sequence in itself has no positional information, BAC contigs aligned after assembly are remarkably reliable if combined with the Nipponbare genome sequence and mapped expressed sequence tags on the genetic map. The rate of SNPs and In/Del detected between Nipponbare and Kasalath by this method is 0.71% and 1.23 sites per kilobase, respectively.

Comparison of the genome structure between *Oryza sativa* and one of its wild relatives, *O. nivara*, has been performed for chromosome 3 (The Rice Chromosome 3 Sequencing Consortium 2005). A total of 3,163 paired BAC-end sequences of *O. nivara* could be assigned by comparison with the genome sequence of *O. sativa* chromosome 3. Sequence similarities of transcribed and nontranscribed regions were 98.8% and 97.7%, respectively. The fingerprint data could be assembled into 16 contigs. This analysis showed that no major rearrangement occurred between these two *Oryza* species. Based on the average distances of paired BAC-end sequences of both species, the size of chromosome 3 of *O. nivara* was estimated as 21% smaller than that of *O. sativa*. Detailed supportive information must be obtained by genome sequence

analysis of these contigs in the future. Comparative analysis of genome sequences at the intra- and interspecies level may elucidate variation in genome structure and function. It may also provide the basis for clarifying which genes and/or genomic regions were lost or redundantly gained by naturally occurring mutation and/or by breeding pressure during the domestication of rice. Such information will be useful for further improvement of cultivated rice.

Utility of rice genome sequence information

Accurate map-based rice genome sequence information is widely applicable to innovative new research. Several main targets that are closely linked are thought important at the moment, although this situation will definitely change, especially with the rapid progress in rice and plant genomics. With the completion of the genome sequence, the next immediate goal is to provide an accurate annotation of all predicted genes in the genome and a comprehensive database based on annotation. The Nipponbare genome sequence is greatly facilitated by the availability of more than 30,000 full-length cDNA nucleotide sequences derived from a similar cultivar (Kikuchi et al 2003). Mapping of these sequences to the genome sequence is the best and most reliable way of attaching biological meaning to the sequence and identifying the genomic position of each transcript and correspondence of predicted and actual genes. These mapping data should be used for identifying a responsible candidate gene to a target trait, mining a plausible promoter region of each gene, or looking for statistical data on the sequence characteristics of the starting point of transcript and exon/intron junction. Toward this goal, the Rice Annotation Project (RAP) began immediately upon completion of the rice genome sequence to facilitate annotation of the function of predicted genes matched with full-length or partial cDNAs. This initiative will continue in the next two or three years to allow periodic revision of annotation and to incorporate novel findings that may result from extensive analysis of gene functions.

As for gene identification corresponding to a phenotype, both forward and reverse genetics methods are undoubtedly receiving great benefit by using the completed rice genome sequence information. Concomitant with the progress of map-based genome sequencing, the recent increase in cases of gene identification is remarkable. For example, genes involved in rice plant height such as *sd1*, *ebisu dwarf* (*d2*), and *gid2* have been identified. In addition to these mutants discretely caused by one gene variation, genes involved in QTLs have been identified such as in heading date (*Hd1, Hd3a,* and *Hd6*) and grain number (*Gn1a*). The identification of these genes is very much facilitated by using the genome sequence information from the beginning of the preparation of chromosomal segmental lines by repeated backcross to the final identification of a candidate gene by map-based cloning (Ashikari et al 2005). The genomic information of such identified genes has been promptly transferred to a trial of homologous gene identification in other cereal crops. The most frequently used rice genes are *Hd1*, *Hd3a,* and *Hd6* because heading date is the most crucial trait for agriculture to obtain enough yield by an appropriate timing of ripening under a changing daylength and temperature. The information from these rice genes has been used

to identify corresponding genes in barley, wheat, ryegrass, and meadow fescue. This challenge is based on the presence of synteny among cereals. A straightforward approach of functional comparison of rice genes involved in rice heading date with those found in other cereals is not necessarily fruitful because of the difference in response to daylength and temperature during their growth. Searching the rice homologs and mapping to other cereal species could provide useful information for further analysis of the genes involved in heading date in major cereal crops. The opportunities for such comparative analysis would increase further as more genes are isolated in rice plants.

Another important and practical issue is how to use sequence information in breeding, that is, selection and transfer of favorable genes through genomics or by genetic engineering. The key for advanced molecular breeding includes finding a favorable gene, its introgression into an existing elite variety within a short period, and confirmation of the transfer and genomic position of the new gene. The target genes or phenotypes to be chosen for further breeding must be those regarding productivity and biotic/abiotic stress tolerance. Trials to find out and isolate these genes have been tackled for a long time by many researchers worldwide. However, these are complex traits controlled by many genes and so far they could not be identified with confidence. Currently, precise genetic analyses of these complex traits are under way relying on sequence polymorphism described above. Of course, selection of favorable genes and their accurate and reproducible evaluation determining phenotype are a prerequisite. Currently available varieties of cultivated rice might not contain the targeted favorable gene resources. In such a case, the survey of favorable genes will be extended to the wild relatives of *O. sativa*. For this purpose, introgression and embryo rescue should be applied and favorable genes could be detected by referring the genome sequence of *O. sativa* using a sequence of the genic region (Brar and Khush 1997). This idea comes from the general observation that the genic region is well conserved among a wide range of plant species. However, it is important to choose the appropriate sequence as probes, including PCR primer sets.

To increase the opportunity to obtain polymorphic markers for the selection of favorable traits, SNPs linked to a targeted phenotype must be collected. For this purpose, germplasm collection should be important for obtaining polymorphisms linked to the target phenotype by linkage disequilibrium (LD) analysis. This analysis is greatly facilitated by the standard genome sequence, which allows the association and validation of the occurrence of sequence polymorphism with relevant phenotypes. In general, the frequency of LD is expected to be high in cultivated species because of bias induced by artificial selection. However, extensive LD analysis has just started and we could expect valuable results available within the next couple of years (Semon et al 2005).

To propagate rice genome sequence information for genetic analysis not only to cereal crops but to a wide range of cultivated species that are too large in terms of genome size or too difficult for genome-wide sequencing, it is necessary to establish DNA markers promptly at a reasonable cost. It is known that, even in a distantly related plant species, amino acid sequences of gene products are well conserved. Considering

the characteristics of codon usage in monocot/dicot plants and the position of exon/intron junction identified by rice genome annotation, nucleotide polymorphisms focusing on the intron region might be obtained among varieties of target species. This idea requires many trials to obtain successful results and still be critically evaluated.

Conclusions

An accurate map-based rice genome sequence is now in the public domain. This information must improve breeding strategies from now on. Since the beginning of agriculture and rice cultivation, favorable agronomic traits have been selected based on phenotypes. And, it is quite surprising that more than 100,000 varieties of rice are now grown worldwide as a result of crossing and selection by farmers and breeders in the last 10,000 years. However, in the next 30–50 years, rice-growing regions will face three main challenges: to increase production as the world population is expected to increase 1.5-fold; to use less land, water, and labor and fewer chemicals; and to satisfy increasing demand from consumers in terms of rice quality. We could not expect to overcome these challenges by mere luck based on conventional breeding methods. Therefore, sufficient knowledge of the complete genetic code of rice is indispensable from now on to help breeders develop strains with specific characteristics such as stress tolerance, disease resistance, and nutrient-enriched or high yield much quicker than through traditional methods. For this purpose, sequence information linked to complete physical and genetic maps of the genome is required to exploit the full potential of the sequences. Now more than ever, it is necessary to address the increasingly complex challenges confronting rice research and rice production. As rice researchers, we should direct our efforts and seize the opportunity for enhancing efficient and sustainable rice development to secure sufficient food for all humans.

References

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.

Arumuganathan K, Earle ED. 1991. Nuclear content of some important plant species. Plant Mol. Biol. Rep. 9:208-219.

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M. 2005. Cytokinin oxidase regulates rice grain production. Science 309:741-745.

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

Devos K. 2005. Updating the 'crop circle'. Curr. Opin. Plant Biol. 8:155-162.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496-512.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, et al. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793-800.
Kikuchi S, Satoh K, Nagata N, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379.
McMullen M. 2003. Quantitative trait locus analysis as a gene discovery tool. Methods Mol. Biol. 236:141-154.
Semon M, Nielsen R, Jones M, McCouch S. 2005. The population structure of African cultivated rice *Oryza glaberrima* (Steud.): evidence for elevated levels of LD caused by admixture with *O. sativa* and ecological adaptation. Genetics 169:1639-1647.
The Rice Chromosome 3 Sequencing Consortium. 2005. Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. Genome Res. 15:1284-1291.
Vitte C, Ishii T, Lamy F, Brar D, Panaud O. 2004. Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol. Gen. Genomics 272:504-511.
Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakaguchi M, Yosihara R, et al. 2004. Composition and structure of the centromeric region of rice chromosome 8. Plant Cell 16:967-976.
Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y,Wing RA, Han B. 2004. Structural features of the rice chromosome 4 centromere. Nucl. Acids Res. 32:2023-2030.

Notes

Authors' addresses: T. Sasaki, T. Matsumoto, and J. Wu, National Institute of Agrobiological Sciences, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan; N. Namiki, STAFF Institute, 446-1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305-0854, Japan.

Annotation of the rice genome

Shu Ouyang, Wei Zhu, J. Hamilton, Haining Lin, M. Campbell, Yuandan Lee,
R.L. Malek, Aihui Wang, Qiaoping Yuan, B. Haas, J. Wortman, and C.R. Buell

A high-quality finished sequence of the rice genome was completed in 2005. However, to maximally use the sequences, quality annotation of the genes and genome features is necessary. The process of annotation is iterative in nature and requires the application and refinement of computational tools coupled with manual curation and evalutation. We are funded by the U.S. National Science Foundation to annotate the rice genome and have constructed pseudomolecules for the 12 *Oryza sativa* subspecies japonica var. Nipponbare chromosomes, which are publicly available through our project Web site (http://rice.tigr.org). We identified genes, gene models, and other annotation features in the rice genome. We expanded our annotation features to include a rice transcript assembly and its alignment with the rice genome, small noncoding RNAs, simple sequence repeats, as well as single nucleotide polymorphisms and insertions/deletions based on alignment with the indica subspecies. We updated our *Oryza* repeat database, which has allowed us to better quantify the repetitive sequences within the rice genome, which total 29% of the genome. To assist users in accessing the genome and our annotation, we expanded the content and functions of our Rice Genome Browser such that it supports 37 annotation tracks and data downloads of the underlying annotation data in various formats.

Keywords: TIGR, annotation, transcript assembly, repetitive sequences, indels, noncoding RNAs

A full description of our annotation process and the features that we have annotated previously in the rice genome has been published recently (Yuan et al 2005). The salient annotation features from Release 3, made available in December 2004 (Yuan et al 2005), are listed in Table 1. In this article, we focus on new annotation features not previously described and expand on some components of our overall annotation project. Thus, readers are referred to Yuan et al (2005) for more details on annotation methods and statistics of our current release (Table 1).

Table 1. Salient statistics of the TIGR Rice Annotation, Release 3.

Statistic	Number
Non-TE-related genes[a]	43,719
TE-related genes	14,196
Average gene model size (bp)	2,852
Average non-TE-related gene model size (bp)	2,672
Average TE-related gene model size (bp)	3,446
No. of FSTs mapped[b]	23,140
No. of non-TE-related proteins	46,976
No. of known proteins	20,911
No. of expressed proteins	6,607
No. of hypothetical proteins	19,458
No. of TE-related proteins	14,274
Proteins containing TM domains[c]	8,205
Proteins containing Pfam domains	29,570
Proteins containing signal peptides	12,750
Proteins with GO assignments[d]	17,169

[a]TE = transposable element. [b]Flanking sequence tags (FSTs) were mapped using 95% identity over 80% of the FST length. [c]TM = transmembrane domain. [d]GO = gene ontology.

Construction of a rice transcript assembly

The use of the TIGR Rice Gene Index (*Oryza sativa*, OsGI) is problematic for our annotation methods and for display of transcript evidence as the TIGR Gene Indices incorporate *in silico* annotated genes that have been deposited in GenBank by genome sequencing projects (Quackenbush et al 2000). Consequently, OsGI contains nonexperimentally derived sequences such as gene predictions and is a source of error propagation when used as evidence in the annotation. To overcome this problem, we constructed a "rice transcript assembly" (TA) using the same clustering/assembly process as used by the TIGR Gene Index team but by excluding nonexperimentally derived sequences. The rice transcript assembly provides nonredundant consensus assemblies of rice cDNA sequences, including expressed sequence tags (ESTs), full-length cDNAs (FL-cDNAs), and nonpredicted coding sequences (CDS) in GenBank from two subspecies of rice, japonica and indica. Construction of the rice TA is summarized in Figure 1, and is similar to the procedure used for the construction of OsGI (Quackenbush et al 2000, Lee et al 2005). Rice cDNA sequences (430,106), including 393,821 ESTs, 32,365 FL-cDNAs, and 3,920 nonpredicted CDS, were cleaned, clustered, and assembled. The result was 28,324 TA sequences and 54,257 singleton sequences.

TA and singleton sequences were aligned to the rice pseudomolecules using the gap2 program and the alignments can be viewed within the Rice Genome

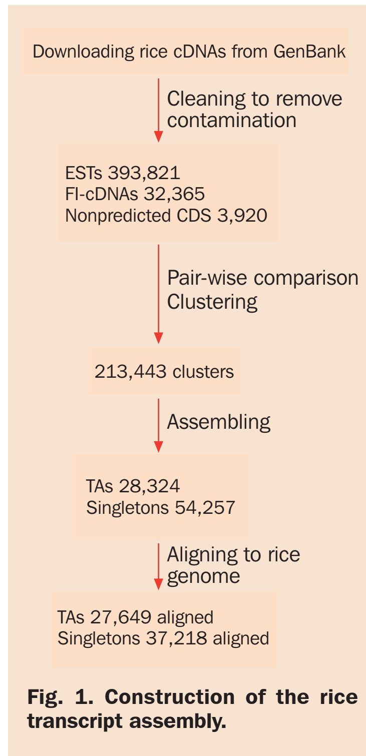


Fig. 1. Construction of the rice transcript assembly.

Browser (see below). Of the 82,581 unique sequences in the transcript assembly, we aligned 78.5% (97.6% TAs and 68.6% singletons) to the rice genome. A report page (http://rice.tigr.org/tdb/e2k1/osa1/expression/TA_search.shtml) was generated for each TA sequence containing a TA assembly diagram, TA assembly component and information table, chromosome location, gene model, and the FASTA-formatted DNA sequence (Fig. 2). The full rice transcript assembly is searchable through a BLAST search engine and can be downloaded for local use via the project ftp site.

Repetitive sequences in the rice genome

As a part of our effort to annotate rice genomic sequences, we created the TIGR *Oryza* Repeat Database (Ouyang and Buell 2004) to identify repetitive sequences within the rice genome. Using RepeatMasker (www.repeatmasker.org) with a default cut-off score of 225, the TIGR *Oryza* Repeat Database, and release 3 of our pseudomolecules, we were able to classify 28.8% of the rice genome as repetitive (Table 2).

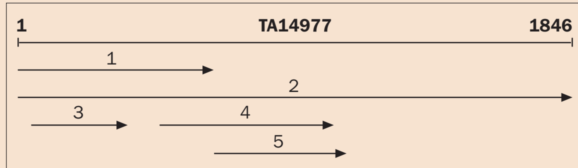
In total, transposable elements (TEs) account for 91.4% of the entire length of the repetitive sequences identified in the rice genome. Among the repetitive elements, retrotransposons dominate in length while miniature inverted-repeat transposable elements (MITEs) dominate in abundance. For the retrotransposons, the Ty3-gypsy

Rice Transcript Assembly - TA14977

Rice Transcript Assembly (OsTA release 1.0 represents a nonredundant set of rice transcripts through clustering and assembly of rice full-length cDNAs, expressed sequence tags (ESTs), and known mRNAs, excluding the predicted CDS sequences derived from rice genome sequencing projects.

Show TA14977 Sequence - Show TA14977 Genomic Alignments - Assembly Components Table

Rice Transcript TA14977 Assembly Diagram:



Rice Transcript Assembly TA14977 Assembly Components:

Seq Number	GenBank Accession	GenBank Division	Left Coord	Right Coord
1	CK040067	EST	1	626
2	AK121147	PLN	2	1846
3	CK033998	EST	50	364
4	CB653746	EST	471	1049
5	CX118996	EST	655	1097

Rice Transcript Assembly TA14977 Genomic Alignments:

Genomic Alignment	Aligned Loci
Chr6: 30464663-30467440	LOC Os06g50390

Rice Transcript Assembly TA14977 Sequence:

```
>TA14977
GCCCCACCATGCCTCACTCCTCCTACCTCGCCCTCGCCCTCGCCGTTGCGGCGGTTCTTG
CTATGGCCGCCGTCGAGGTGTCCGGCCTCGGCTTCGACCTCCACCACCGCTACTCCCCCA
TCGTCCAGCGGTGGGCGGAGGAGCGTGGCCACGCTGGCGTGTCCTGGCCTGCCGGCGCCG
AGGTGATCGGCTCGCCCGAGTACTACTCCGCGCTGTCCCGACACGACCACGCCCTCTTCG
CGCGACGCGGGCTCGCCCAGGGCGACGGCCTCGTCACCTTCGCCGACGGCAACATCACCC
TGCGGCTCGACGGATCGCTGCACTACGCGGAGGTGGCGGTGGGCACGCCGAACACGACGT
TCCTGGTGGCGCTCGACACCGGCAGCGACCTCTTCTGGGTGCCCTGCGACTGCAAGCAGT
GCGCTCCCCTCGGCAACCTCACGGCGGTGGACGGCGGCGGCGGCCCGGAGCTCAGGCAGT
ACAGCCCGAGCAAGTCGTCGACGAGCAAGACGGTCACCTGCGCCAGCAACCTCTGCGACC
AGCCCAACGCCTGCGCCACGGCCACCAGCAGCTGCCCCTACGCCGTCCGCTACGCCATGG
CCAACACCTCCTCCTCCGGCGAGCTCGTCGAGGACGTGCTCTACCTCACCAGGGAGAAGG
GCGCCGCCGCCGCCGCCGCCGGAGCCGCCGTGCGAACTCCCGTGGTGTTCGGGTGCGGGC
AGGTGCAGACGGGCTCGTTCCTCGACGGCGCCGCCGCCGACGGCCTCATGGGTCTCGGCA
TGGAGAAGGTGTCCGTGCCCAGCATCCTCGCCAGCACCGGCGTCGTCAAGTCGAACAGCT
TCTCCATGTGCTTCAGCAAAGACGGCCTCGGCCGGATCAACTTCGGCGACACCGGCAGCG
CCGACCAGTCCGAGACGCCGTTCATCGTCAAATCCACACACTCCTACTACAACATCAGCA
TCACGTCGATGAGCGTCGGGGACAAGAACTTGCCGTTGGGGTTCTACGCCATTGCTGACT
CCGGCACGTCCTTCACCTACCTCAACGACCCGGCCTACACGGCCTACACCACCAATTTCA
ACGCGCAGATCAGCGAGAGGAGGGCCAACTTCAGCGGCAGTACTCGTTCTGGTCCATTTC
CTTTCGAGTACTGCTACTCGTTGAGTCCTGATCAAACAACGGTGGAACTGCCTGTTGTGA
GCCTGACGACCAATGGAGGAGCCGTGTTCCCGGTGACCAGCCCTGTTTACCCCATCGCGG
CTCAGATGACAAACGGAGAGATTCGTATCATTGGTTACTGCCTGGCTGTCATCAAGAGCG
ATCTCCCCATCGACATCATTGGCCAGAACTTCATGACCGGGCTCAAGGTCGTCTTCAACC
GTGAGAAATCCGTCCTGGGCTGGCAAAAGTTCGACTGTTACAAGGACGAGAAGATGACAG
ACGACGGGTCGAGCGTCGGGAGCCCGAGCCCATCGCCGGGGCCGACCACTCACGTCTTCC
CGCAGCCGCAGGAGAGCGACTCGCCGGCGGGCAGAACGCCCATCCCCGGCGCGGCGCCGG
TGCCGCGGTCGAGCAGCGCGGCAGCTGGCGGCCGTGCAGGCTTCCGGCTCCTGTCGTCCA
TGGCGCTCTTGCTGCTGGCTGCTGCTGCTGCAGCTGTTTTCTGATGATTTTTTTTTTACT
TGGGGAGTTTATAGGTGATTTCATTCGTGGTGGCTGCTCTGCTGTGTACATAACACATGC
ATGCATTCGAGATGATTTTTATTCAGAAAGCGTAGATAGATCGATCGATCATTGTAGGAT
TTAGTAATTTGTTCTTGTGATGATGATTGATGTGTATATGGTTACG
```

Fig. 2. Screenshot of TA report page showing alignment of individual sequences within the assembly, coordinates of each EST/FL-cDNA within the assembly, links to the Rice Genome Browser and cognate locus, and consensus sequence.

Table 2. Repetitive sequence contents in the rice genome.[a]

Classification	Total length (kb)	Percent in repeat fraction[b]	Percent in rice genome[c]	Occurrence	Occurrence (per Mb)	Average size (bp)
Transposable element	97,425.5	91.4	26.3	210,040	566.7	463.8
Retrotransposon	49,102.3	46.1	13.3	36,710	99.1	1,337.6
Transposon	27,086.1	25.4	7.3	60,982	164.5	444.2
MITE	21,237.2	19.9	5.7	112,348	303.1	189.0
Centromere-related	2,681.0	2.5	0.7	4,756	12.8	563.7
Centromere-specific retrotransposon	1,381.0	1.3	0.4	3,116	8.4	443.2
CentO	1,127.2	1.1	0.3	1,172	3.2	961.7
Unclassified centromere sequence	172.9	0.2	0.1	468	1.3	369.3
Telomere-related	1,223.8	1.2	0.3	3,084	8.3	396.8
Telomere-associated	90.2	0.1	0.0	360	1.0	250.6
Telomere	1,133.6	1.1	0.3	2,724	7.4	416.2
rDNA	174.5	0.2	0.1	508	1.4	343.5
45S rDNA	166.7	0.2	0.0	426	1.2	391.4
5S rDNA	7.8	0.0	0.0	82	0.2	94.5
Unclassified	5,045.8	4.7	1.4	21,170	57.1	238.3
Total	106,550.6	100.0	28.8	239,558	646.4	

[a]A total of 370.6 Mb of *O. sativa* subspecies japonica var. Nipponbare genomic DNA was searched against the TIGR *Oryza* Repeat Database. A summary of the repetitive sequences identified is shown. [b]The percent of repeat fraction refers to the percent (in length) of each repeat within the repetitive fraction identified in the TIGR rice pseudomolecules. [c]The percent of rice genome is the percent (in length) of each repeat subclass in the TIGR rice pseudomolecules.

subclass accounts for 6.4% of the total length of the genomic sequences and 5.7% of the number of repeat sequences identified in the 370.6 Mb of rice genomic DNA, making Ty3-gypsy the most abundant repetitive element in the rice genome (Table 2). The En/Spm (CACTA) subclass of transposons was the second most abundant repetitive sequence with respect to length (2.7%) and the most frequently occurring transposon (9,429; 3.9% in number of repeat sequences). Despite being predominantly <500 bp in length, MITEs represent 5.7% of the total length of the rice genome. Two subclasses of MITEs, Tourist and MITE-adh type B, were extremely abundant, accounting for 0.8% and 0.9% of the total length of the rice genome, respectively. A

total of 0.7% of the surveyed genomic sequences was identified as centromere-related sequences (Table 2). Sequences (1.2 Mb) related to the rice centromere-specific tandem repeat, CentO (Cheng et al 2002), and telomere-related sequences (1.2 Mb), mostly occurring in fragments shorter than 2 kb, were identified. In addition, 174.5 kb of ribosomal RNA gene sequences were detected. The largest fragment was about 8.8 kb, approximately the size of one copy of the 45S rDNA (Nandabalan and Padayatty 1989). Only a handful of 5S rRNA gene sequences were found, with the largest fragment only 330 bp. However, small rDNA fragments could be found throughout the rice genome, suggesting that rRNA genes, some possibly of plastid origin, could exist outside of rDNA clusters. Unclassified repetitive sequences, which were distributed all over the genome, accounted for 1.4% of the total length of the rice genome.

Retrotransposon and transposon sequences were distributed throughout the rice genome. However, centromeric regions contained a higher percentage of retrotransposon and transposon sequences than other areas (Fig. 3). Moreover, these sequences occurred at a higher frequency on the short arms of chromosome 4 and 10 than at other regions, consistent with cytological data that indicate that 4S and 10S are the most heterochromatic arms of the 12 rice chromosomes (Cheng et al 2001). The majority of the CentO sequences were located in the centromeric regions (Fig. 3). Three large tracts of CentO sequences, with 18.3, 7.6, and 12.2 kb, respectively, were detected in the centromeric region of chromosome 8 (Nagaki et al 2004). As expected, centromere-related sequences could be found more frequently at centromeric regions, as demonstrated on every chromosome (Fig. 3). Not surprisingly, these sequences, which are also retrotransposon sequences, could be identified at locations other than centromeres, albeit at a lower frequency.

The overall repeat content in the rice genome has probably been underestimated. Several technical limitations result in the underrepresentation of repetitive sequences. For instance, because of the high percent of repetitive sequences, centromeric regions are difficult to finish and, in bacterial artificial chromosomes (BAC-based genome sequencing projects), these sequences are typically underrepresented. Such is the case with the sequence generated by the IRGSP (2005), which comprises the underlying sequence in our pseudomolecules. As a matter of fact, only two centromeric regions have been completely sequenced (Nagaki et al 2004, Wu et al 2004, Zhang et al 2004). In addition, because of the lack of restriction enzyme sites, telomere-related sequences are underrepresented in the BAC libraries created through restriction enzyme digestion and ligation. Difficulties in sequencing the tandem repeats of rDNAs could explain the paucity of rDNA sequences identified.

TIGR Rice Genome Browser

The Rice Genome Browser (www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice) is an implementation of the Generic Genome Browser (Stein et al 2002) developed by the Generic Model Organism Database Project (www.gmod.org). The Rice Genome Browser allows users to view the TIGR rice genome annotation and other genome annotation features in an interactive fashion using a Web browser. The browser window

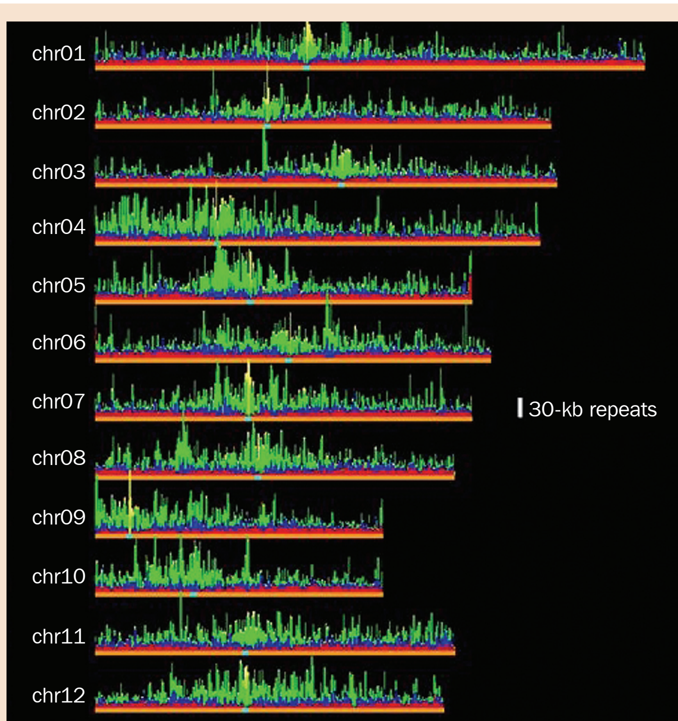


Fig. 3. Distribution of repeats in the rice genome. Superclasses/classes of repetitive sequences were quantitated in 100-kb windows after performing the searches with RepeatMasker. The amount of each repeat superclass/class within each window on a chromosome was drawn proportionally. Gold: rice chromosomes; light blue: centromeres; green: retrotransposons; blue: transposons; red: MITEs; yellow: centromere-related sequences; brown: telomere-related sequences; white: rDNAs.

is divided into three sections (Fig. 4). The top section contains a diagram of the rice chromosomes. Clicking on a chromosome loads the pseudomolecule into the browser and displays the first 100 kb of sequence and its annotation. The next section contains the viewing window, viewer controls, and search box. The search box allows the input of loci identifiers, gene models, BAC accessions, and genomic regions specified by coordinates along the pseudomolecules. The viewing window (Fig. 5) displays the annotation in tracks, with each track displaying the feature using a graphic of a distinctive shape and color and a label identifying the feature. In some cases, features within a track use several colors to convey information about the features. The features are hyperlinked and clicking on them will open a new browser window displaying information about the track. The viewing controls allow the user to move the view along the pseudomolecule and zoom in and out. The last section contains the track

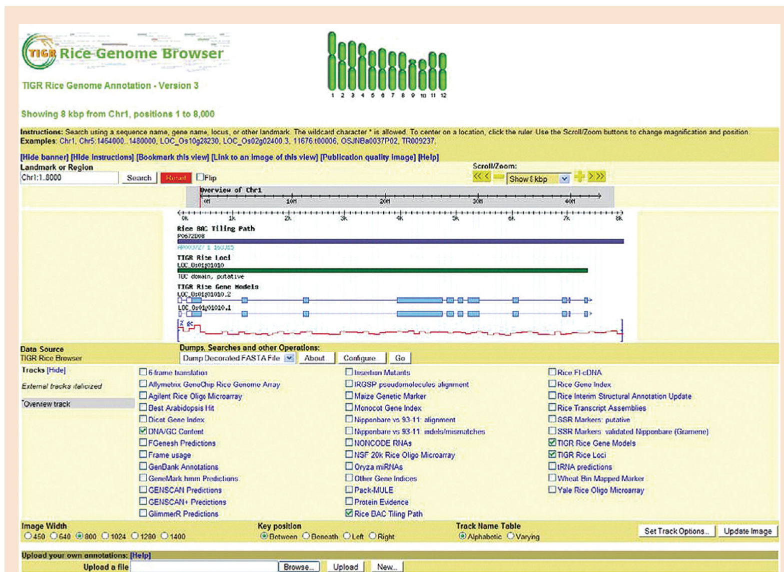


Fig. 4. A screenshot of the TIGR Rice Genome Browser. A single locus has been selected for viewing. The scale of the display can be altered and users can move up and down the pseudomolecule using the selection tools. The tracks are listed at the bottom and can be altered by users based on preference.

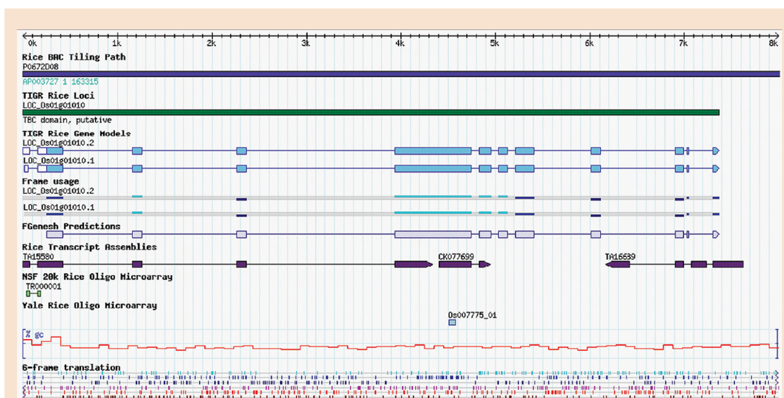


Fig. 5. A detailed view of the TIGR Rice Genome Browser annotation viewing window centered on LOC_Os01g01010. In total, 10 tracks are displayed: Rice BAC tiling path, TIGR rice loci, TIGR rice gene models, frame usage, FGenesh predictions, rice transcript assemblies, NSF 20k rice oligo microarray probes, Yale rice oligo microarray probes, GC% graph, and 6-frame translation.

selection table and plug-in controls. The track selection table allows users to display and remove annotation tracks from the viewing window. The TIGR Rice Genome Browser currently has 37 tracks available, which are listed and described in Table 3. The drop-down-box labeled "Dumps, Searches, and other Operations" provides access to the genome browser plug-ins. Several plug-ins are available, including Dump GFF File, Dump Decorated FASTA File, and Dump Sequence File. These plug-ins allow users to download the annotations and genomic sequence visible in the viewing window in a user-selected format. At the bottom of this section are input boxes that allow users to upload their own annotation and specify a remote location where their own annotation can be accessed. The user-supplied annotation is displayed in a separate track and can be modified through a built-in editing window.

Noncoding RNAs

The rice genome, as do genomes from a wide variety of organisms, ranging from prokaryotes to eukaryotes and mammals, contains several examples of noncoding RNA (ncRNA) classes. Noncoding RNAs are defined as RNAs that function without being translated into proteins. Noncoding RNAs play pivitol roles as constituents of important cellular machinery, including ribosomes, spliceosomes, and telomerase. Noncoding RNAs affect processes as diverse as DNA transcription, repair, and replication to RNA processing, modification, editing, and translation. We have mapped the location of small nucleolar (snoRNAs), small nuclear (snRNAs), micro (miRNAs), and transfer (tRNAs) RNAs onto the rice pseudomolecules and displayed these specific classes of ncRNAs as tracks on the Rice Genome Browser.

snoRNAs and snRNAs

The snoRNAs are involved in RNA processing and modification. Box C/D snoRNAs direct the 2′-*O*-ribose methylation of spliceosomal snRNAs while H/ACA snoRNAs act as guides for the pseudouridylation of rRNAs. We downloaded the set of unique ncRNAs from the integrated knowledgebase, NONCODE (Liu et al 2005, http://noncode.bioinfo.org.cn/), which contains more than 5,300 nonredundant sequences from more than 800 different organisms. The file was subsequently parsed to identify *Oryza* snoRNAs (436), snmRNAs (2), snRNAs (3), and one each of SRP-7SL, a ribozyme, and nc1. Eleven snoRNAs were identified whose described function is pseudouridylation of RNA, 423 snoRNAs were curated as being involved in 2′-*O*-ribose methylation, and 8 snoRNAs participated in RNA cleavage. It is noteworthy to mention that some snoRNAs carry out more than one modification process. As has been previously reported (Liang et al 2002, Chen et al 2003), many of the snoRNA genes occur in clusters, and many of the clusters map to intronic regions (Fig. 6). In general, the rice snoRNA genes are redundant in that two-thirds of the snoRNAs have multiple isoforms and/or have duplicated copies of the same snoRNA.

Table 3. Features displayed on the TIGR Rice Genome Browser.

Track name	Track description
Rice BAC Tiling Path	TIGR Rice Pseudomolecules BAC Tiling Path
TIGR Rice Loci	Genes annotated by the TIGR Rice Annotation Project
TIGR Rice Gene Models	Gene models annotated by the TIGR Rice Annotation Project
Rice Interim Structural Annotation Update	Gene models updated/created/merged using the PASA2 pipeline
Frame Usage	Depicts the reading frame of each CDS using a "musical staff" notation
GenBank Annotations	Annotation provided by IRGSP sequencing groups as deposited in GenBank
FGenesh Predictions	Gene model derived from FGenesh algorithm (monocot matrix)
GeneMark.hmm Predictions	Gene model derived from GeneMark.hmm algorithm (rice matrix)
GENSCAN Predictions	Gene model derived from GENSCAN algorithm (maize matrix)
GENSCAN+ Predictions	Gene model derived from GENSCAN algorithm (*Arabidopsis* matrix)
GlimmerR Predictions	Gene model derived from GlimmerR algorithm (rice matrix)
Rice Fl-cDNA	Full-length cDNAs for rice deposited in GenBank
Rice Gene Index	TIGR Rice Gene Index is aligned to the rice genome using the gap2 program with cut-off criteria of a minimum of 65% identity and coverage greater than 70%.
Rice Transcript Assemblies	Release 1 of the TIGR Rice Transcriptome Assembly; 430,106 ESTs, FL-cDNAs, and mRNAs were obtained from GenBank, clustered, and assembled into 82,581 unique sequences. The transcript assemblies were aligned to the rice genome using the gap2 program with a minimum identity of 90% and coverage greater than 70%.
Monocot Gene Index	TIGR Plant Gene Indices (monocots only), including barley, maize, onion, sugarcane, sorghum, and rye, are aligned to the rice genome using the gap2 program with cut-off criteria of a minimum of 65% identity and coverage greater than 70%. Rice is aligned to the genome using gap2 with a minimum identity of 90% and coverage greater than 70%.
Dicot Gene Index	TIGR Plant Gene Indices (dicots only) are aligned to the rice genome using the gap2 program with cut-off criteria of a minimum of 65% identity and coverage greater than 70%.
Other Gene Indices	TIGR Plant Gene Indices (remaining plant indices) are aligned to the rice genome using the gap2 program with cut-off criteria of a minimum of 65% identity and coverage greater than 70%.

continued on next page

Table 3 continued.

Track name	Track description
Protein Evidence	TIGR Plant Protein Database is aligned to the rice genome using the gap2 program with cut-off criteria of a minimum of 30% identity and coverage greater than 40%.
Maize Genetic Marker	Maize genetic markers mapped to rice BAC/PAC sequences
Wheat Bin Mapped Marker	Wheat genetic markers mapped to rice BAC/PAC sequences
Insertion Mutants	Flanking sequence tags (FSTs) from *Tos17*, T-DNA, and *Ac/Ds* insertion mutants
IRGSP pseudomolecules alignment	IRGSP pseudomolecules (Build 3.0) aligned to the TIGR pseudomolecules (Version 3.0)
Nipponbare vs 93-11: alignment	BGI indica pseudomolecules (downloaded on 14 Feb. 2005) aligned to the TIGR pseudomolecules (Version 3.0). This track shows high-scoring segment pairs (HSPs) of the alignment between the BGI assemblies and our pseudomolecules. The track "93-11 vs Nipponbare indels/mismatches" shows the detailed mismatches (SNP candidates) and indels.
Nipponbare vs 93-11: indels/mismatches	BGI indica pseudomolecules (downloaded on 14 Feb. 2005) aligned to the TIGR psedumolecules (Version 3.0). This track shows only mismatches (SNP candidates) and indels between the two genome sequences.
Best Arabidopsis Hit	Best match of rice protein to *Arabidopsis* proteome (Release 5). Red if match is a mutual best hit, yellow otherwise.
Pack-MULE	Pack-MULEs on chromosomes 1 and 10 as identified by Jiang et al (2004)
SSR Markers: putative	Putative SSRs in the 12 pseudomolecules identified by the Simple Sequence Repeat Indentification Tool (Temnykh et al 2001)
SSR Markers: validated Nipponbare	Known SSR markers in Nipponbare as determined by Gramene Release 16
(Gramene) tRNA predictions	tRNA genes identified using the tRNAScan-SE algorithm (Lowe and Eddy 1997)
Oryza miRNAs	Predicted stem-loop precursor miRNA transcripts and mature *Oryza sativa* were retrieved from the miRBase database at the Sanger Centre. Stem-loop precursors and then mature miRNA sequences were mapped using BLASTN.
NONCODE RNAs	Unique *Oryza* snoRNAs and snRNAs were downloaded from NONCODE (Liu et al 2005). The sequences were mapped to the rice pseudomolecules using BLASTN.
NSF 20k Rice Oligo Microarray	Probes from the NSF Rice 20k Oligo array mapped to the pseudomolecules at 100% identity over 100% of their length.

continued on next page

Table 3 continued.

Track name	Track description
Affymetrix GeneChip Rice Genome Array	Probes from the Affymetrix array mapped to the pseudomolecules at 100% identity over 100% of their length.
Yale Rice Oligo Microarray	Probes from the Yale (http://plantgenomics.biology.yale.edu/riceatlas/) oligo array mapped to the pseudomolecules at 100% identity over 100% of their length.
Agilent Rice Oligo Microarray	Probes from the Agilent array mapped to the pseudomolecules at 100% identity over 100% of their length.
DNA/GC Content	Displays a C/G% graph and low zoom values and the pseudomolecule sequence at high zoom values.
6-frame translation	Displays the 6-frame translation on the pseudomolecule sequence.

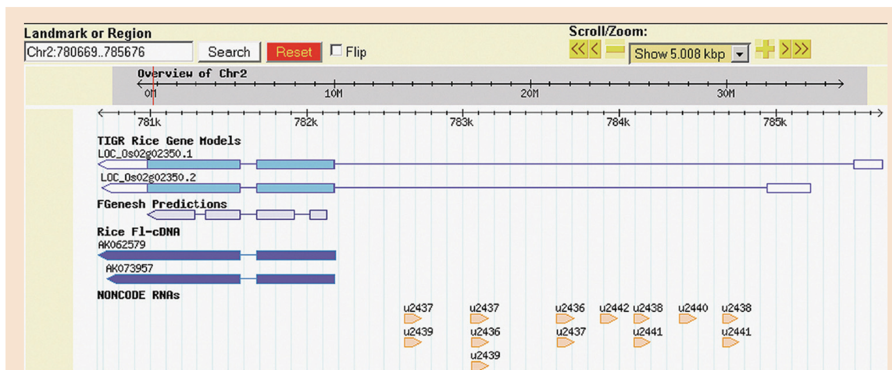


Fig. 6. Redundancy and intronic location of snoRNAs in a cluster. The first intron in the rice gene model LOC_Os02g02350 contains snoRNA genes having 1–3 copies. The NONCODE database unique identifiers that are displayed are clickable. The ncRNAs shown belong to the Z118a-h family and function in RNA methylation.

miRNAs

MicroRNAs are small, 20–24-nucleotide endogenous RNAs that regulate post-transcriptional repression of mRNA transcripts (Bartel 2004). In plants, miRNA directed repression is primarily via miRNA-directed cleavage of the mRNA transcript. However, repression mediated by inhibition of translation also occurs (Schwab et al 2005, Tang et al 2003). Our current suite of gene prediction programs (FGenesh, GlimmerR, GeneScan, GeneMark, PASA, etc.) is not trained to predict the unusually small transcripts of the precursor miRNA transcripts, which are generally <100 nucleotides in length. Although computational strategies exist to predict miRNA genes and their targets in rice (Bonnet et al 2004, Wang et al 2004a,c, Jones-Rhoades and Bartel 2004), we have restricted our Genome Browser presentation to those miRNAs contained in miRBase (http://microrna.sanger.ac.uk/), formerly the miRNA Registry, maintained at the Sanger Institute. Predicted stem-loop precursor miRNA transcripts and mature *Oryza sativa* miRNA sequences were retrieved and stem-loop precursor transcripts were mapped to rice pseudomolecules using BLASTN. Mature miRNA sequences were then overlaid onto the precursor transcripts and displayed on the Browser (Fig. 7).

Transfer RNAs

The tRNA genes comprise one of the largest families and a typical eukaryotic genome contains hundreds of tRNA genes. As the rice mitochondrial and chloroplast genomes have been described previously (Notsu et al 2002, Hiratsuka et al 1989), tRNA predictions were restricted to the nuclear genome. We searched the rice pseudomolecules using RNA detection algorithms described by Lowe and Eddy (1997). Three tRNA prediction algorithms are combined in the program, tRNAScan-SE, which makes use of covariance models and generates secondary structure predictions. Predicted tRNAs are displayed as a separate track on the Genome Browser and can be searched

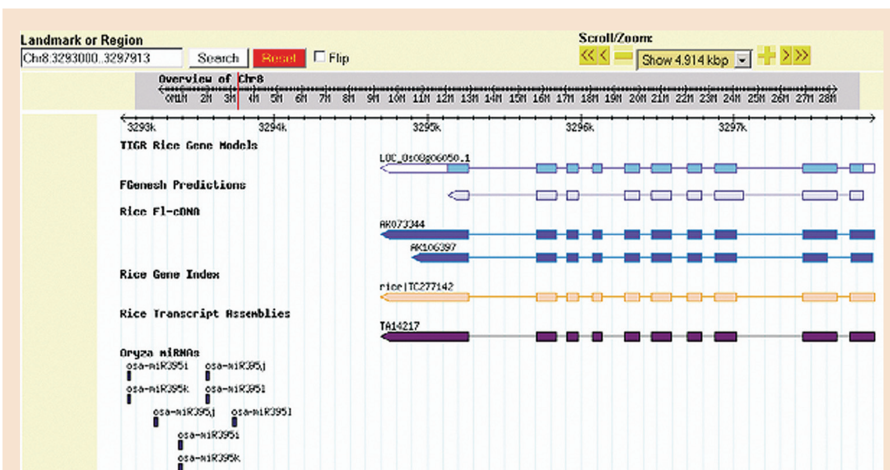


Fig. 7. MicroRNA cluster containing miR395 family members. Note the intergenic location of this family, upstream of LOC_08g06050, and the lack of EST (Rice Gene Index, transcript assemblies), FL-cDNA, and gene prediction (FGenesh) evidence. Mature miRNA sequences are shown.

Table 4. The total number of predicted tRNA genes (724) in the rice genome separated by chromosome.

Item	Chromosome											
	1	2	3	4	5	6	7	8	9	10	11	12
No. of genes	98	76	79	81	58	38	42	49	37	63	33	70

by chromosome number from the Whole Genome Annotation Database (www.tigr.org/tdb/e2k1/osa1/irgsp.shtml). We identified 724 predicted tRNA genes, 28% more than a recent rice genome survey estimate (Wang et al 2004c; Tables 4 and 5). The increased number of genes identified may be related to the greater sequence coverage contained in our current pseudomolecules. TIGR Rice Feature (TRF) identifiers have been assigned to each predicted tRNA.

Mapping of genetic markers and microsatellite repeats

Three sets of rice genetic markers have been aligned to the TIGR rice pseudomolecules. The first set is the RFLP markers used to create the high-density genetic map derived from an intersubspecies cross of rice (Harushima et al 1998) and was downloaded from the Rice Genome Program (RGP) Web site (http://rgp.dna.affrc.go.jp). The

Table 5. Transfer RNA genes and the 52 anticodons represented in the combined pseudomolecule sequences.

Amino acid	Anticodon		Amino acid	Anticodon	
Ala	AGC	20	Leu	AAG	15
	CGC	12		CAA	19
	GGC	1		CAG	9
	TGC	10		TAA	4
	Total	43		TAG	10
Arg	ACG	23		Total	57
	CCG	8	Lys	CTT	20
	CCT	11		TTT	12
	TCG	4		Total	32
	TCT	12	Met	CAT	54
	Total	58		Total	54
Asn	GTT	29	Phe	GAA	20
	Total	29		Total	20
Asp	GTC	30	Pro	AGG	14
	ATC	1		CGG	10
	Total	31		TGG	13
Cys	ACA	1		Total	37
	GCA	16	Ser	AGA	12
	Total	17		CGA	8
Gln	CTG	10		GGA	4
	TTG	20		GCT	19
	Total	30		TGA	16
Glu	CTC	26		Total	59
	TTC	15	Thr	AGT	11
	Total	41		CGT	5
Gly	CCC	11		GGT	7
	GCC	25		TGT	15
	TCC	10		Total	38
	Total	46	Trp	CCA	18
His	GTG	26		Total	18
	Total	26	Tyr	ATA	2
Ile	AAT	18		GTA	18
	TAT	5		Total	20
	Total	23	Val	AAC	17
				CAC	10
				GAC	14
				TAC	4
				Total	45

second set of markers was derived from PCR-based screening of tiled yeast artificial chromosome clones using unique EST primer sets from the RGP (http://rgp.dna.affrc.go.jp). A combined total of 13,319 marker sequences from the RGP were used for physical/genetic mapping. The RG and RZ RFLP markers from Cornell University comprise the third set of markers and 646 marker sequences were obtained from Gramene (www.gramene.org). FASTA-formatted marker sequences were obtained from GenBank using the unique GenBank accessions. The alignment program GMAP (Genomic Mapping and Alignment Program) was used to map each marker sequence to the genome (Wu and Watanabe 2005). GMAP is optimized for rapid mapping and aligning of cDNA sequences to a genome and was chosen for use because most of the markers were derived from ESTs or cDNAs. All validated marker alignments exceeded a cut-off criterion of 95% identity over 90% of the length of the marker sequence.

From a total of 13,965 marker sequences, 11,608 were aligned to the 12 rice pseudomolecules. A total of 11,172 markers (or 96.2% of the total) had a syntenous position between the genetic and physical maps, whereas 436 were nonsyntenous, suggesting that the mapped physical chromosomal position was in conflict with the genetic map chromosomal position. This high degree of synteny reflects the high quality of the rice genetic maps. Detailed information for each marker and its physical position on the rice pseudomolecules is available at the TIGR Web site (http://rice.tigr.org). In addition, the genetic markers are displayed as a separate track on the TIGR Rice Genome Browser.

Putative simple sequence repeats (SSRs) were identifed using the Simple Sequence Repeat identification tool (Temnykh et al 2001). A total of 28,148 perfect SSRs (13,700 dinucleotide SSRs, 11,877 trinucleotide SSRs, and 2,571 tetranucleotide SSRs) were identified on the TIGR Rice Pseudomolecules (v3.0). The putative SSRs can be viewed on the TIGR Rice Genome Browser and a Web-based search tool is available (www.tigr.org/tdb/e2k1/osa1/putative_ssr.shtml) that supports searches for putative SSRs by type or motif. Gramene-validated SSR markers for Nipponbare have also been mapped to the TIGR Rice Pseudomolecules using e-PCR (Schuler 1997) to map the flanking primer pairs with a total of 2,812 SSR markers successfully mapped. The SSR markers are displayed in the TIGR Rice Genome Browser and are hyperlinked to display information about the SSR markers from Gramene.

SNPs and indel identification between indica and japonica

We developed a pipeline to identify, store, and display polymorphisms among rice genomes and have identified all polymorphisms between *O. sativa* subspecies japonica Nipponbare and *O. sativa* subspecies indica 93-11. These polymorphisms are publicly available through the project Web pages. The indica 93-11 pseudomolecules (downloaded from the Beijing Genomics Institute, ftp://ftp.genomics.org.cn/pub/ricedb/RGPVs9311/9311/Sequence/Chromosome/ on 14 Feb. 2005) were aligned to the TIGR japonica pseudomolecules (Version 3.0), using the MUMMER3 package (Delcher et al 2002; http://sourceforge.net/projects/mummer). In total, about two million single nucleotide polymorphisms (SNPs) were identified, in addition to 406,848 indels (Table

Table 6. Mismatches and indels between the *Oryza sativa* subspecies japonica and indica genomes.

Chromosome	No. of SNPs	No. of insertions	No. of deletions
1	255,431	27,746	27,539
2	229,475	23,268	23,376
3	210,724	22,839	22,433
4	141,071	15,762	16,084
5	148,010	15,920	15,909
6	165,131	16,233	16,462
7	164,075	15,583	15,648
8	152,060	15,786	15,751
9	136,224	13,163	13,131
10	132,473	12,829	12,984
11	121,915	11,983	11,823
12	123,037	12,398	12,198
Total	1,979,626	203,510	203,338

6). These polymorphisms can be viewed through the TIGR Rice Genome Browser in separate tracks (Fig. 8).

Conclusions

We are able to provide a rich resource for rice and cereal biologists to data-mine the rice genome. In addition to the identification of genes, we have added new layers of functional annotation to the rice genome. The expansion of the *Oryza* Repeat Database is essential for the accurate identification of TE-related gene models that do not warrant extensive manual curation efforts. The presentation of ncRNAs provides a new resource for researchers investigating gene regulation in rice. The localization of genetic markers aligned to the rice genome, in the context of our gene annotation, provides a robust resource for researchers positionally cloning genes in rice. These annotations, coupled with the expansion and improved functionality of our Genome Browser, will allow scientists to data-mine the rice genome for genes of importance and will facilitate the identification of monocot gene function.

References

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281-297.

Bonnet E, Wuyts J, Rouze P, Van de Peer Y. 2004. Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. Proc. Natl. Acad. Sci. USA 101:11511-11516.

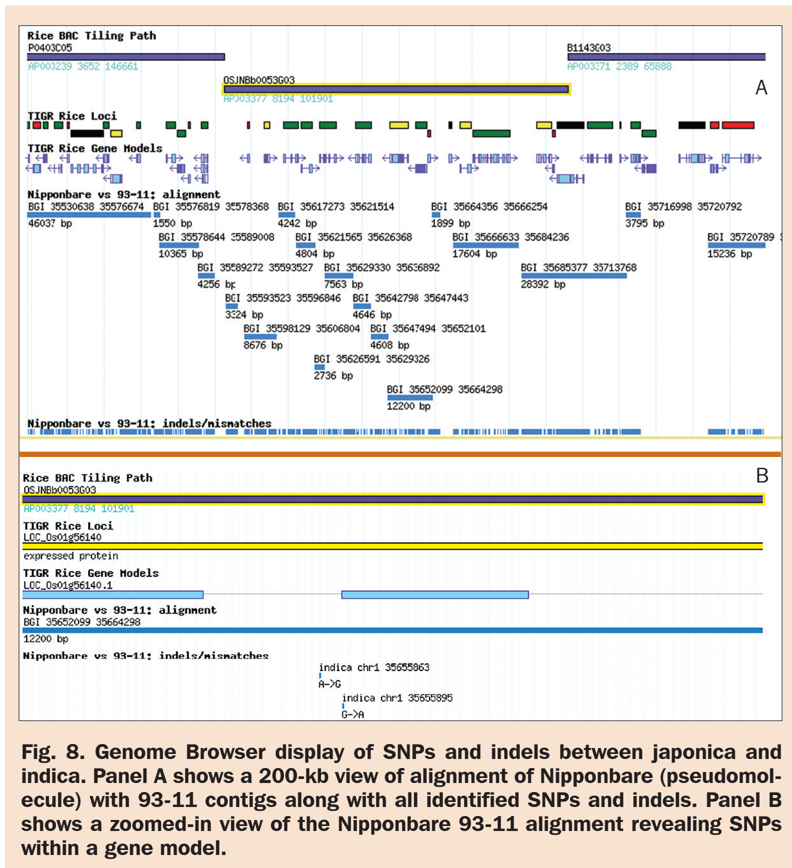


Fig. 8. Genome Browser display of SNPs and indels between japonica and indica. Panel A shows a 200-kb view of alignment of Nipponbare (pseudomolecule) with 93-11 contigs along with all identified SNPs and indels. Panel B shows a zoomed-in view of the Nipponbare 93-11 alignment revealing SNPs within a gene model.

Chen CL, Liang D, Zhou H, Zhuo M, Chen YQ, Qu LH. 2003. The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. Nucl. Acids Res. 31:2601-2613.

Cheng Z, Buell CR, Wing RA, Gu M, Jiang J. 2001. Toward a cytological characterization of the rice genome. Genome Res. 11:2133-2141.

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J. 2002. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691-1704.

Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparision. Nucl. Acids Res. 30:2478-2483.

Harushima Y, Yano M, Shomura A, Sato M, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, et al. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, et al. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. 217:185-194.

IRGSP (International Rice Genome Sequencing Project). 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Jiang N, Bao ZR, Zhang X, Eddy SR, Wessler SR. 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569-573.

Jones-Rhoades MW, Bartel DP. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell 14:787-799.

Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, Pertea G, Antonescu V, Chan A, Cheung F, Quackenbush J, et al. 2005. The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucl. Acids Res. 33:D71-74.

Liang D, Zhou H, Zhang P, Chen YQ, Chen X, Chen CL, Qu LH. 2002. A novel gene organization: intronic snoRNA gene clusters from *Oryza sativa*. Nucl. Acids Res. 30:3262-3272.

Liu C, Bai B, Skogerbo G, Cai L, Deng W, Zhang Y, Bu D, Zhao Y, Chen R. 2005. NONCODE: an integrated knowledge database of non-coding RNAs. Nucl. Acids Res. 33:D112-115.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl. Acids Res. 25:955-964.

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J. 2004. Sequencing of a rice centromere uncovers active genes. Nat. Genet. 36:138-145.

Nandabalan K, Padayatty JD. 1989. A minor 9.8 kb rDNA unit of rice variety IR-20. Indian J. Biochem. Biophys. 26:289-292.

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol. Genet. Genomics 268:434-445.

Ouyang S, Buell CR. 2004. The TIGR Plant Repeat Databases: a collective resource for identification of repetitive sequences in plants. Nucl. Acids Res. 32 Database Issue: D360-363.

Quackenbush J, Liang F, Holt I, Pertea G, Upton J. 2000. The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucl. Acids Res. 28:141-145.

Schuler GD. 1997. Sequence mapping by electronic-PCR. Genome Res. 7:541-550.

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. 2005. Specific effects of microRNAs on the plant transcriptome. Dev. Cell 8:517-527.

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S. 2002. The Generic Genome Browser: a building block for a model organism system database. Genome Res. 12:1599-1610.

Tang G, Reinhart BJ, Bartel DP, Zamore PD. 2003. A biochemical framework for RNA silencing in plants. Genes Dev. 17:49-63.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11:1441-1452.

Wang JF, Zhou H, Chen YQ, Luo QJ, Qu LH. 2004a. Identification of 20 microRNAs from *Oryza sativa*. Nucl. Acids Res. 32:1688-1695.

Wang XJ, Reyes JL, Chua NH, Gaasterland T. 2004b. Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol. 5:R65.

Wang XY, Shi XL, Hao BL. 2004c. The tRNA and rRNA genes in the *Oryza sativa* genome. Yi Chuan Xue Bao 31:871-877.

Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakaguchi M, Yoshihara R et al. 2004. Composition and structure of the centromeric region of rice chromosome 8. Plant Cell 16:967-976.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859-1875.

Yuan Q, Ouyang S, Wang A, Zhu W, Maiti R, Lin H, Hamilton J, Haas B, Sultana R, Cheung F, Wortman J, Buell CR, et al. 2005. The Institute for Genomic Research Osa1 rice genome annotation database. Plant Physiol. 138:18-26.

Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y, Wing R, Han B, et al. 2004. Structural features of the rice chromosome 4 centromere. Nucl. Acids Res. 32:2023-2030.

Notes

Authors' address: The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, phone: 301-795-7558, fax: 301-838-0208, e-mail: rbuell@tigr.org.

Acknowledgments: The efforts of the TIGR Bioinformatics Department in generating a suite of tools and resources for eukaryotic sequence and annotation are appreciated. Work on rice genome annotation is funded by grants to C.R.B. from the National Science Foundation (DBI-0321538) and from the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service (2003-35317-13173).

Structural genomics and resources

The *Oryza* map alignment project (OMAP): a new resource for comparative genomics studies within *Oryza*

R.A. Wing, H.R. Kim, J.L. Goicoechea, Y. Yu, D. Kudrna, A. Zuccolo, S.S. Ammiraju Jetty, M. Luo, W. Nelson, C. Soderlund, P. San Miguel, N. Gill, J. Walling, S. Jackson, B. Hurwitz, D. Ware, L. Stein, D. Brar, and D. Mackill

With the completion of a finished genome sequence, we must now functionally characterize the rice genome by a variety of methods, including comparative genomic analysis between cereal species and within the genus *Oryza*. *Oryza* contains two cultivated and 22 wild species that represent 10 distinct genome types. The wild species, in particular, contain an essentially untapped reservoir of agriculturally important genes that must be harnessed to enhance and sustain crop productivity.

OMAP was established two years ago to generate a comprehensive set of genomics resources to investigate genome evolution and enhance positional cloning efforts in the genus *Oryza*. To date, we have generated (1) 12 high-quality BAC libraries that encompass the 10 genome types of *Oryza*, (2) approximately 1,000 Mb of BAC end sequence from these libraries, and (3) SNaPshot fingerprint databases for 10 of the 12 libraries. All of these resources are publicly available through the AGI BAC/EST Resource Center, GenBank, or at www.OMAP.org. The fingerprints and end sequences have been combined to develop 10 phase I physical maps. Six of these physical maps, *O. nivara* (AA), *O. rufipogon* (AA), *O. glaberrima* (AA), *O. punctata* (BB), *O. officinalis* (CC), and *O. brachyantha* (FF), have been heavily manually edited (HME) and aligned to the reference rice genome sequence. These alignments have revealed a large array of genome rearrangements relative to the japonica (Nipponbare) genome and have allowed us to begin drawing a more complete picture of *Oryza* genome evolution. We present the current status of OMAP and discuss recent analysis of the HME maps and comparative sequence analysis of select loci across the *Oryza* AA genome diploids.

Keywords: *Oryza*, wild species, BAC libraries, map alignment, end sequences

Rice (*Oryza sativa* L.) is the most important human food crop in the world. The agronomic importance of rice, its shared evolutionary history with major cereal crops, and small genome size have led to the generation of a high-quality finished genome sequence by the International Rice Genome Sequencing Project (2005). The highly

accurate and public IRGSP sequence now serves as a unifying research platform for a complete functional characterization of the rice genome. Such an analysis will investigate the rice transcriptome, proteome, and metabolome, with the goal of understanding the biological function of all rice genes (35,000–40,000) and applying that information to improve rice production and quality. This comprehensive analysis will use a variety of techniques and resources from expression and genome tiling arrays to collections of tagged mutant populations developed in elite cultivars grown around the world.

Comparative genomics between the cereal genomes and within the genus *Oryza* will also play a critical role in our understanding of the rice genome (Ahn et al 1993, Ahn and Tanksley 1993, Bennetzen and Ma 2003, Han and Xue 2003, Huang and Kochert 1994, Jena et al 1994, Ma and Bennetzen 2004). By comparing genome organization, genes, and intergenic regions between cereal species, one can identify regions of the genome that are highly conserved or rapidly evolving. Such regions are expected to yield key insights into genome evolution, speciation, and domestication. The study of conserved noncoding sequences (CNSs) between cereal genomes will also increase our ability to understand and isolate *cis*-regulatory elements required for precise developmental and temporal gene expression (Kaplinsky et al 2002).

The genus *Oryza* is composed of two cultivated (*O. sativa* and *O. glaberrima*) and 22 wild species (Khush 1997, Vaughan et al 2003). Based on recent phylogenetic data, Ge et al (1999) proposed that *Porteresia coarctata* should also be included in the genus *Oryza*. Cultivated rice (2n=24) is classified as an AA genome and it has six wild AA genome relatives. The remaining 15 wild species are classified into nine other genome types that include both diploid and tetraploid species. Figure 1 shows a proposed phylogenetic tree of the genus *Oryza* as described by Ge et al (1999) based on the analysis of two nuclear genes and one chloroplast gene. Wild species of *Oryza* are a reservoir of useful genes for resistance to major biotic and abiotic stresses, including yield-enhancing loci (Brar and Khush 2006). These species offer a largely untapped resource of agriculturally important genes that have the potential to solve many problems in rice production that we face today such as yield, drought and salt tolerance, and disease and insect resistance.

To better understand the wild species of rice and take advantage of the IRGSP genome sequence, we have embarked on an ambitious comparative genomics program entitled the *Oryza* Map Alignment Project (OMAP). The long-term objective of OMAP is to create a genome-level closed experimental system for the genus *Oryza* that can be used as a research platform to study evolution, development, genome organization, polyploidy, domestication, gene regulatory networks, and crop improvement. The specific objectives of OMAP are to (1) construct deep-coverage large-insert BAC libraries from 11 wild species of *Oryza* and 1 cultivated African rice species (*O. glaberrima*), (2) fingerprint and end-sequence the clones from all 12 BAC libraries, (3) construct physical maps for all 12 *Oryza* species and align them to the IRGSP genome sequence, and (4) perform a detailed reconstruction of rice chromosomes 1, 3, and 10 across all 12 *Oryza* species.

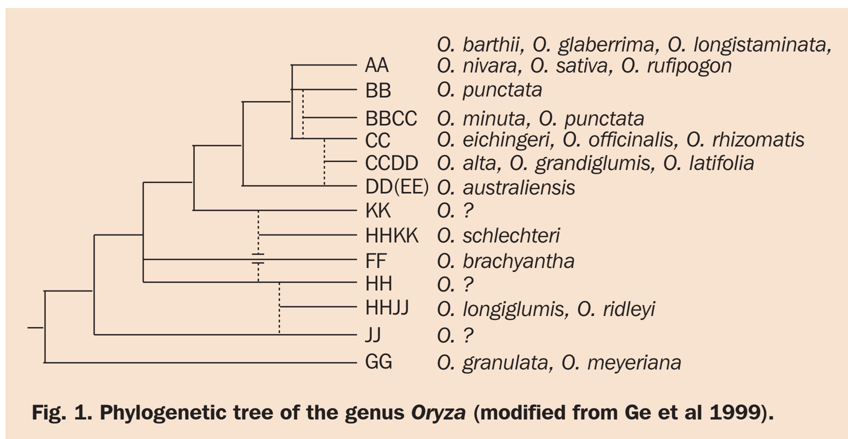


Fig. 1. Phylogenetic tree of the genus *Oryza* (modified from Ge et al 1999).

We report on our current progress in OMAP and give some early glimpses into the results we are finding.

Results

Development of the OMAP BAC library resource

Wild species were obtained from (1) the International Rice Research Institute (IRRI), Los Baños, Philippines; (2) the National Institute of Genetics, Mishima, Japan; and (3) Cornell University, Ithaca, New York (Table 1). Our major criteria for the selection of these wild rice accessions were that each one was robust and sufficient seed was available for distribution to the community, and that each contained potentially useful agronomic traits.

High-molecular-weight DNA was obtained from young seedlings for the AA genome species *O. nivara*, *O. rufipogon*, and *O. glaberrima*. In contrast, because no inbred single-seed descent material was available for the remaining wild species, we prepared DNA from single plants that were clonally propagated at IRRI. Efforts are now under way to generate inbred seed from these wild species and seed should be available from the IRRI seed bank within 2–3 years.

Deep-coverage large-insert BAC libraries were developed for all 12 OMAP species using standard procedures developed in our laboratory over the past 10 years (Table 1) (Luo and Wing 2003, Wing et al 2005). All libraries were quality tested for insert size and depth of coverage and were found to represent at least 10 genome equivalents, with average insert sizes ranging from 123 kb (*O. coarctata*) to 161 kb (*O. nivara*) (Ammiraju et al 2006). All OMAP BAC libraries were deposited in the Arizona Genomic Institute's BAC/EST Resource Center for public distribution (www.genome.arizona.edu).

Table 1. OMAP BAC library summary.[a]

Oryza species	Genome type	Accession number	Accession source	# Clones	# Plates in the library	Av insert size (kb)	Old genome size (Mb)	New genome size (Mb)	# Plates to run for OMAP 10x
O. glaberrima	AA	96717	Africa	55,296	144	140	809	354	96
O. nivara	AA	W0106	India	55,296	144	161	760	448	144
O. rufipogon	AA	105491	Malaysia	64,512	168	134	760	439	96
O. punctata	BB	105690	Africa	36,864	96	142	539	423	96
O. officinalis	CC	100896	Thailand	92,160	240	141	1,201	653	144
O. minuta	BBCC	101141	Philippines	129,024	336	125	1,691	1124	240
O. alta	CCDD	105143	S. America	92,160	240	133	1,000	772	192
O. australiensis	EE	100882	Australia	92,160	240	153	1,054	960	192
O. brachyantha	FF	101232	Africa	36,864	96	131	343	338	96
O. granulata	GG	102118	Thailand	73,728	192	134	907	862	192
O. ridleyi	HHJJ	100821	Thailand	129,024	336	127	1,568	1,278	288
O. coarctata	HHKK	104502	Bangladesh	147,456	384	123	1,568	ND[b]	336

[a]BAC libraries and hybridization filters can be ordered from the AGI BAC/EST Resource Center (www.genome.arizona.edu). “Old” genome sizes were retrieved from the Angiosperm DNA C-value database, release 3.1, September 2001 (Bennet and Leitch 2001) or re-determined experimentally as “new” genome sizes using flow cytometry. [b]ND = not done.

Development of wild species FPC/STC physical maps

After BAC library construction, the libraries are then BAC end-sequenced and fingerprinted. The fingerprints are assembled into contigs based on shared bands between clones using FPC software (Soderlund et al 2000). FPC contigs can then be aligned to the IRGSP genome sequence using the associated BAC end sequences.

BAC end sequencing

As shown in Table 2, all the BAC libraries have been end-sequenced except *O. granulata* and *O. ridleyi*, which are in progress. Using *O. nivara* as an example, AGI attempted to sequence 110,589 BAC ends and successfully sequenced 106,104 BAC ends. The average high-quality sequence read length was 665 bases. All BAC end sequences have been submitted to the GSS section of GenBank.

BAC fingerprinting

To fingerprint a BAC library, we use a modification of the SNaPshot fingerprinting method described by Luo and Wing (2003). Briefly, BAC DNA is isolated using a semiautomated 96-well alkaline lysis protocol (Kim HR and Wing RA, unpublished), and then digested with five restriction enzymes, of which four generate 5′ overhangs. The corresponding 3′ OH ends are then extended using a single fluorescently labeled ddNTP and DNA polymerase. The reaction products are then separated on ABI3730XL capillary electrophoresis sequencers and the labeled fragments are identified using ABI fragment analysis software.

As shown in Table 3, 10 of the 12 BAC libraries have been fingerprinted and assembled into phase I physical maps. Again, using *O. nivara* as an example, AGI attempted to fingerprint 51,056 clones and achieved a 91% success rate. These fingerprints were then assembled into a phase I physical map composed of 456 contigs and 2,356 singletons.

These physical maps are now being refined as "heavily manually edited maps (HME)" using a variety of assembly parameters followed by end merging and contig alignment to the IRGSP reference sequence. Figures 2, 3, and 4 show SYMAP alignments of HME maps for *O. rufipogon* (AA), *O. punctata* (BB), and *O. brachyantha* (FF). As expected, one can see that there is a tremendous amount of colinearity between the wild species and the reference IRGSP sequence. However, several regions of structural variation can be observed, even between the two AA genome species.

All OMAP FPC maps are available on the Internet using webFPC (www.genome.arizona) and BAC end sequence alignments on Gramene (www.gramene.org) (Ware et al 2002).

Analysis of structural variation between *O. sativa* and three A genome wild species accessions used in OMAP

The *O. sativa* subsp. japonica cv. Nipponbare chromosome 3 was recently finished by the U.S. Rice Chromosome 3 Sequencing Consortia (2005). Cytologically, chromosome 3 is the second largest rice chromosome, measuring 56.41 μm (or approx. 52.4 Mb), and is one of the most euchromatic (Cheng et al 2001). Genetically, chromosome

Table 2. Summary of OMAP BAC end sequencing.

Species	Genome	# Reads	% Success after trim	# GenBank submissions	Av length after trim (in GenBank) (bp)	Mb sequenced (in GenBank)	Genome coverage (%)
O. nivara	AA	110,589	96	106,124	665	71	16
O. rufipogon	AA	73,716	96	70,982	742	53	12
O. glaberrima	AA	73,344	91	66,821	590	39	11
O. punctata	BB	68,567	93	68,384	720	49	12
O. officinalis	CC	110,592	95	103,251	717	74	11
O. minuta	BBCC	184,224	92	169,651	560	95	8
O. alta	CCDD	146,400	90	128,732	586	75	10
O. australiensis	EE	147,456	93	137,530	676	93	10
O. brachyantha	FF	71,350	94	67,364	672	45	13
O. coarctata	HHKK	208,510	94	195,285	661	129	ND[a] (>10)
Total/av		1,194,748	93	1,114,124	659	724	11

[a]ND = not done.

Table 3. Summary of OMAP SnaPshot/FPC fingerprinting.

Species	Genome	# of FP clones	% Success	# of contigs		# of singletons
				Initial FPC	HME FPC	
O. nivara	AA	51,056	91	456	343	2,356
O. rufipogon	AA	33,023	91	637	327	1,305
O. glaberrima	AA	33,065	85	905	167	2,172
O. punctata	BB	34,224	93	490	210	1,483
O. officinalis	CC	47,493	85	703	ND[a]	1,957
O. minuta	BBCC	86,861	90	3,962	ND	9,576
O. alta	CCDD	63,860	85	2,270	ND	2,894
O. australiensis	EE	58,594	86	1,409	ND	2,163
O. brachyantha	FF	25,216	78	428	225	1,801
O. coarctata	HHKK	92,522	91	1,250	ND	1,810

[a]ND = not done.

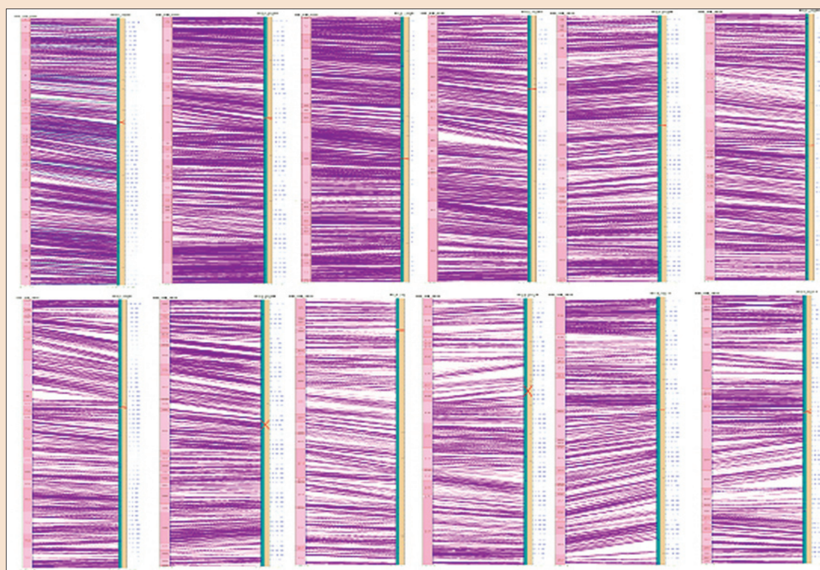


Fig. 2. SyMAP alignment of *O. rufipogon* BAC FPC/STC contig map (left columns) with the 12 IRGSP reference genome pseudomolecules (right columns). Horizontal lines represent BES alignments to the IRGSP reference sequence.

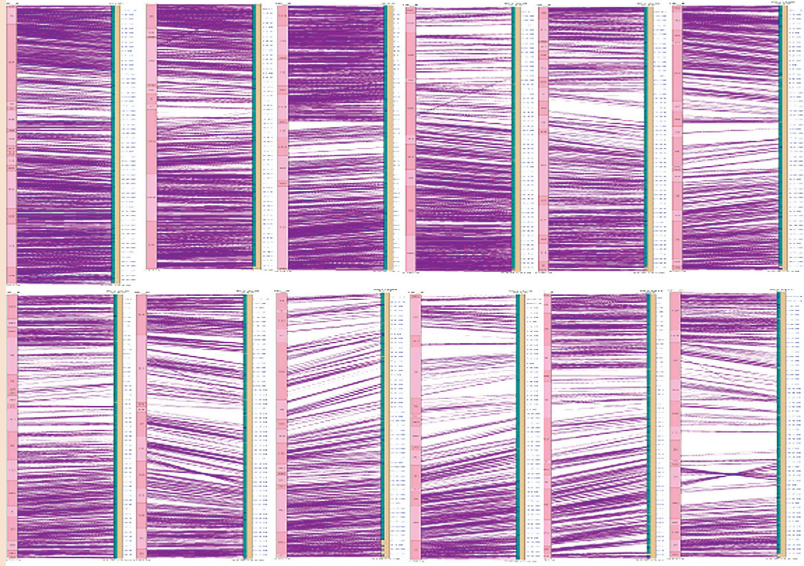


Fig. 3. SyMAP alignment of *O. punctata* BAC FPC/STC contig map (left columns) with the 12 IRGSP reference genome pseudomolecules (right columns). Horizontal lines represent BES alignments to the IRGSP reference sequence.

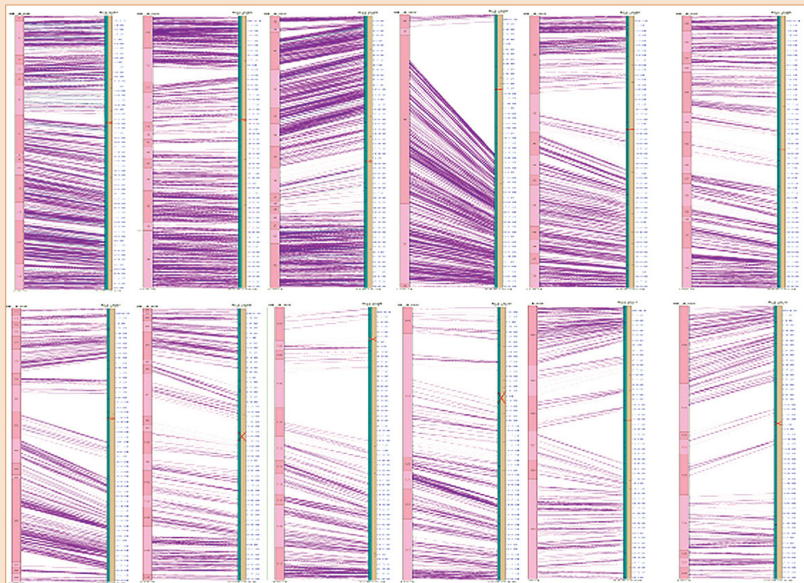


Fig. 4. SyMAP alignment of *O. brachyantha* BAC FPC/STC contig map (left columns) with the 12 IRGSP reference genome pseudomolecules (right columns). Horizontal lines represent BES alignments to the IRGSP reference sequence.

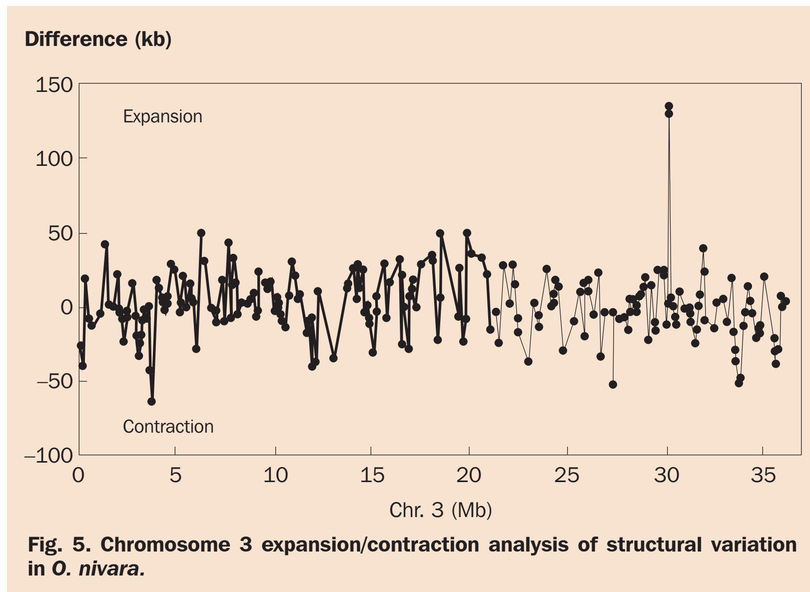


Fig. 5. Chromosome 3 expansion/contraction analysis of structural variation in *O. nivara*.

3 is 170 cM in length (Harushima et al 1998) and has 27 morphological mutants. In addition, over 133 agronomic genes/traits/QTLs (Oryzabase.org.) and 963 cDNAs (Wu et al 2002) have been found associated with chromosome 3 (Oryzabase.org). The consortia sequenced approximately 36.1 Mb of chromosome 3 and identified 6,237 new genes (Rice Chromosome 3 Sequencing Consortia 2005).

Although *Oryza* separated from maize and sorghum about 50 million years ago (MYA) and from wheat and barley about 40 MYA, their common evolutionary history can be traced by the colinear order of genetic markers across their chromosomes (Moore et al 1995). This is particularly true for the short arm of chromosome 3, which shows large stretches of genetic marker colinearity with maize chromosomes 1 and 9, sorghum linkage group L, and barley and wheat chromosome 4L. Such conserved synteny across the cereals suggests that rice chromosome 3 will be a good model to study chromosome evolution.

Insertions and deletions in genomes play a critical role in evolution. To obtain a more in-depth understanding of the role insertions and deletions are playing in reshaping the genomes of *Oryza*, we generated minimum tiling paths of BAC clones across the entire length of chromosome 3 from the HME maps of *O. nivara* (AA), *O. rufipogon* (AA), and *O. glaberrima* (AA). Then, the predicted size of each BAC clone, based on the alignment of paired BAC end sequences on the IRGSP reference genome sequence, was compared with the empirically determined BAC size as determined by pulsed-field gel electrophoresis. Figures 5, 6, and 7 show graphical representations of this analysis and reveal that all three rice chromosomes are undergoing chromosome-wide expansion and contraction relative to the IRGSP reference sequence.

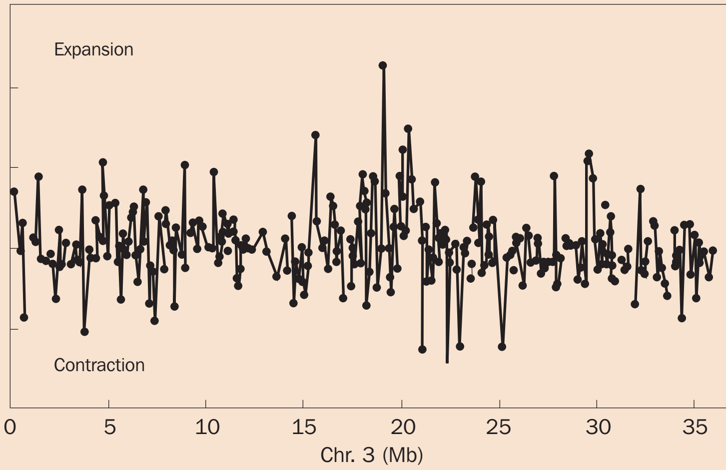


Fig. 6. Chromosome 3 expansion/contraction analysis of structural variation in *O. rufipogon.*

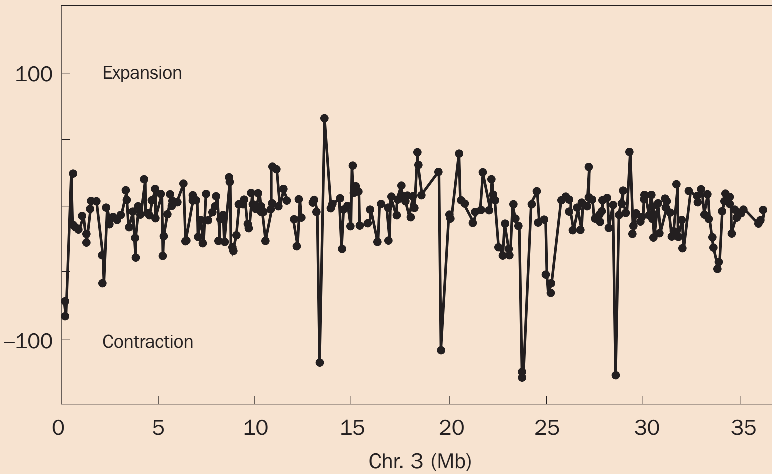
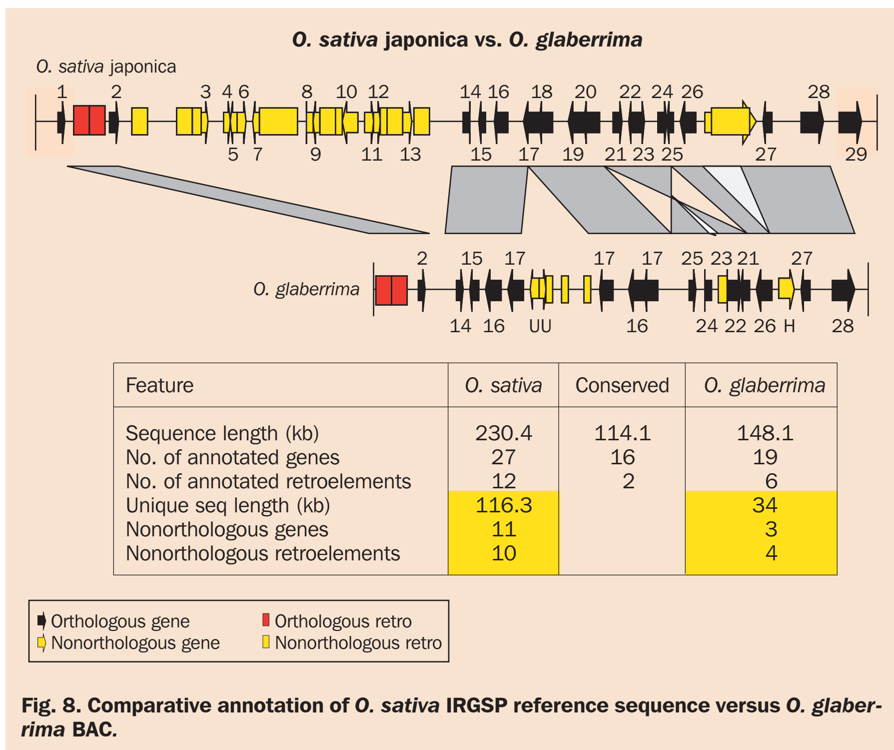


Fig. 7. Chromosome 3 expansion/contraction analysis of structural variation in *O. glaberrima.*

Table 4. Summary of annotated genes and complete retrotransposons and solo LTRs of three fully sequenced BACs in comparison to the IRGST reference sequences.

	O. sativa		×	*O. nivara*	
	Number	kb		Number	kb
Unmatched block	5	52		3	9
Genes in unmatched block	5	12		0	0
Retro/solo LTRs in unmatched block	7	39		1 + unannotated	9
	O. sativa		×	*O. rufipogon*	
	Number	kb		Number	kb
Unmatched block	6	60		4	16
Genes in unmatched block	4	12		0	0
Retro/solo LTRs in unmatched block	7	40		4	15
	O. sativa		×	*O. glaberrima*	
	Number	kb		Number	kb
Unmatched block	5	113		5	33
Genes in unmatched block	10	26		3	6
Retro/solo LTRs in unmatched block	9	50		3 + unannotated	16

To obtain a detailed sample of genome expansion in *O. sativa* relative to the other three AA genome species, we sequenced, finished, and annotated a single BAC from each of these species that mapped near the top of the short arm of chromosome 3. The selected *O. nivara* BAC had a predicted size of 220 kb based on paired BAC end sequence alignment but was found to be 178 kb after sequencing, indicating that *O. sativa* expanded 42 kb in this region or *O. nivara* contracted by the same amount (or a combination of both, resulting in an overall difference of 42 kb). The predicted sizes of the *O. rufipogon* and *O. glaberrima* BACs were 169 and 230 kb; however, their actual sequenced sizes were 126 and 148 kb, thus representing overall variation of 43 and 82 kb, respectively. Table 4 summarizes the annotation analysis and shows that the insertion of transposable elements is responsible for the relative expansion in *O. sativa* and for the presence of new genes. The most dramatic example can be seen in the sequence comparison between *O. sativa* and *O. glaberrima* as pictured in Figure 8. Here, the overlapping region in *O. sativa* spans 230.4 kb but only 114.1 kb is conserved with *O. glaberrima* and contains 16 annotated genes and two annotated retroelements. The remaining 116.3 kb of unique *O. sativa* sequence contains 10 nonorthologous genes and 11 nonorthologous retroelements. For the *O. glaberrima*



Feature	*O. sativa*	Conserved	*O. glaberrima*
Sequence length (kb)	230.4	114.1	148.1
No. of annotated genes	27	16	19
No. of annotated retroelements	12	2	6
Unique seq length (kb)	116.3		34
Nonorthologous genes	11		3
Nonorthologous retroelements	10		4

Fig. 8. Comparative annotation of *O. sativa* IRGSP reference sequence versus *O. glaberrima* BAC.

BAC, 34 kb of the 148.1-kb sequence is unique and contains three nonorthologous genes and four nonorthologous retroelements.

Summary, conclusions, and future research

The domestication of rice some 10,000 years ago has severely limited the gene pool that breeders can use to improve rice. The wild species of the genus *Oryza* contain a wealth of genetic diversity that must therefore be uncovered if we are to meet the challenges of feeding the world in the 21st century.

The *Oryza* Map Alignment Project was designed to conduct a detailed characterization of a single representative of each of the 10 genome types of wild species of rice. The alignment of these genomes to the IRGSP reference sequence will provide a comprehensive physical framework whereby numerous genome-wide applied and basic research projects can be launched to unlock the genetic potential of these wild genomes and provide breeders with new candidate genes and QTLs for rice improvement.

All of the production objectives for OMAP will be completed by the end of 2006. Over the remaining two years, we plan to accomplish the remaining OMAP objectives, including a re-evaluation of the phylogenetic tree of *Oryza,* establishment

of genome-wide SSR/SNP maps of all the AA genome species, elucidation of *Oryza* transposon dynamics and their effect on genome size variation, and development of a comprehensive "*Oryza* Rearrangement Index" to determine the majority of expansion, contraction, inversion, and translocation events with respect to the IRGSP reference sequence.

Because of the massive amount of information and resources that OMAP has and will produce, it is obvious that our consortia alone will not be able to realize the full potential of this systems biology project without collaboration and cooperation with the broader research community. We therefore propose to establish an International *Oryza* Map Alignment Project (I-OMAP) in a similar vein as the IRGSP to help coordinate research activities using this new resource. Such an I-OMAP could include the development of advanced backcross (ABC) populations and chromosome segmental substitution lines (CSSLs) using the AA genome OMAP accessions. Further, as sequencing technologies become more cost effective, I-OMAP could propose 6X whole-genome draft sequences of all 12 OMAP species. Having such powerful genetic and sequence resources available would lead to much more complete understanding of the world's most important food crop.

References

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.

Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B, et al. 2006. The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. Genome Res. 16:140-147.

Bennetzen JL, Ma J. 2003. The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr. Opin. Plant Biol. 6:128-133.

Brar DS, Khush GS. 2006. Cytogenetic manipulation and germplasm enhancement of rice (*Oryza sativa* L.). In: Singh RJ, Jauhar P, editors. Genetic resources, chromosome engineering and crop improvement. Vol. 2. Cereals. Boca Raton, Fla. (USA): CRC Press. p 115-158.

Cheng Z, Buell CR, Wing RA, Gu M, Jiang J. 2001. Toward a cytological characterization of the rice genome. Genome Res 11:2133-2141.

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.

Han B, Xue Y. 2003. Genome-wide intraspecific DNA-sequence variations in rice. Curr. Opin. Plant Biol. 6:134-138.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, et al. 1998. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494.

Huang H, Kochert G. 1994. Comparative RFLP mapping of an allotetraploid wild rice species (*Oryza latifolia*) and cultivated rice (*O. sativa*). Plant Mol. Biol. 25:633-648.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Jena KK, Khush GS, Kochert G. 1994. Comparative RFLP mapping of a wild rice, *Oryza officinalis*, and cultivated rice, *O. sativa*. Genome 37:382-389.
Kaplinsky NJ, Braun DM, Penterman J, Goff SA, Freeling M. 2002. Utility and distribution of conserved noncoding sequences in the grasses. Proc. Natl. Acad. Sci. USA 99:6147-6151.
Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. 35:25-34.
Luo M, Wing RA. 2003. An improved method for plant BAC library construction. Methods Mol. Biol. 236:3-20.
Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. USA 101:12404-12410.
Moore G, Devos KM, Wang Z, Gale MD. 1995. Cereal genome evolution: grasses, line up and form a circle. Curr. Biol. 5:737-739.
Rice Chromosome 3 Sequencing Consortia. 2005. Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. Genome Res. 15:1284-1291.
Soderlund C, Humphray S, Dunham A, French L. 2000. Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10:1772-1787.
Vaughan DA, Morishima H, Kadowaki K. 2003. Diversity in the *Oryza* genus. Curr. Opin. Plant Biol. 6:139-146.
Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S, Stein L. 2002. Gramene: a resource for comparative grass genomics. Nucl. Acids Res. 30:103-105.
Wing RA, Ammiraju JS, Luo M, Kim H, Yu Y, Kudrna D, Goicoechea JL, Wang W, Nelson W, Rao K, et al. 2005. The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. Plant Mol. Biol. 59:53-62.
Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, et al. 2002. A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525-535.

Notes

Authors' addresses: R.A. Wing, H.R. Kim, J.L. Goicoechea, Y. Yu, D. Kudrna, A. Zuccolo, S.S. Ammiraju Jetty, and M. Luo, Arizona Genomics Institute, University of Arizona, Tucson, Ariz.; W. Nelson and C. Soderlund, Arizona Genomics Computational Laboratory, University of Arizona, Tucson, Ariz.; P. San Miguel, N. Gill, J. Walling, and S. Jackson, Department of Agronomy and Genomics Core Facility, Purdue University, West Lafayette, Ind.; B. Hurwitz, D. Ware, and L. Stein, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; D. Brar and D. Mackill, International Rice Research Institute, Los Baños, Philippines.

Analysis of oligo hybridization properties by high-resolution tiling microarrays in rice

Xiangfeng Wang, Lei Li, Viktor Stolc, Waraporn Tongprasit, Chen Chen, Jun Wang, Songgang Li, and Xing Wang Deng

Rice genome sequencing and computational annotation provide a static map for understanding this model of *Gramineae* species. With the development of *in situ* oligonucleotide synthesis technology, tiling-path microarrays have become a dynamic and efficient way for monitoring large-scale transcriptional activities and detecting novel transcribed elements missed by software. Unlike conventional cDNA or oligonucleotide arrays, tiling-path platforms employ the full extent of oligos covering given genomic regions, and thus offer excellent experimental conditions in which to assay the properties of oligos in terms of their specificity and efficiency of hybridization to their corresponding targets. Here, we report a tiling-path microarray analysis of a 1-Mb region (10 to 11 Mb) in japonica rice chromosome 10, which was tiled by a 36-mer oligo set at a resolution of 5 bp. Our analysis focused on three major factors of oligo hybridization properties, including GC content, melting temperature (Tm), and the repetitiveness of oligo sequences.

Keywords: Rice, genomics, tiling-path microarrays, transcriptome, hybridization

Whole-genome tiling-path microarrays have been used in several sequenced model organisms as a dynamic and efficient way to facilitate *in silico* genome annotations. Traditional cDNA or oligomer microarrays use relatively few probes for each gene and are biased toward known and predicted gene structures (Mockler et al 2005). In contrast, tiling-path platforms provide more biologically relevant information beyond the measurement of mRNA levels. This includes genome-wide transcriptional activity monitoring, gene structure definition, the identification of novel transcribed elements, and broader uses in organ-specific alternatively spliced isoforms, chromatin-immunoprecipitation-chip studies, and other extended applications, according to different innovatively customized designs (Bertone et al 2005).

Characteristics of the rice genome and its annotation

Rice is one of the most important crops in the world, providing staple food for about half of the human population (Hoshikawa 1993). Determination of the rice genome sequence has been considered a milestone in the field of *Gramineae* species, and a valuable resource for scientists to investigate its functional elements. The genome-scale comparison of the two subspecies of rice, indica and japonica, respectively, sequenced by the Beijing Institute of Genomics and the International Rice Genome Sequencing Project, provides us with important clues in deducing the evolutionary history of grass species. Studies of single nucleotide polymorphisms between indica and japonica and their related wild species are equally important for understanding heterosis, or hybrid vigor, in rice.

Genomics analyses in rice discovered that many of rice's own unique features were different from those of other sequenced organisms. For instance, unlike most other grasses, rice has a relatively small genome of about 440 Mb (Bennetzen et al 2002), but the initial estimate of 55,000 to 60,000 rice genes is nearly twice as many as the gene content of mammalian genomes, and rice genes are usually clustered together into islands, separated by highly repetitive sequences. However, recent published estimates based on newly improved genome data indicate that the nontransposon gene count is at least 38,000 to 40,000 (Yu et al 2005). In the IRGSP newly released finished quality sequence of japonica, 37,544 protein-coding genes were identified, of which 71% had a putative homolog in *Arabidopsis*. In a reciprocal analysis, 90% of the *Arabidopsis* proteins had a putative homolog in the predicted rice proteome (International Rice Genome Sequencing Project 2005). It is possible that most of the low-homology genes might be the remnants of ancient gene duplications or fragments resulting from the transposition of transposable elements. The effort to identify pseudogenes in rice needs more evidence of expression and comparative genomics methodology.

One of the unique characteristics of the rice gene is the gradient change in GC content along the transcription direction (Wong et al 2002). This feature and the unusual high-GC content represent the major unique factors that might affect the accuracy of microarray experiments in rice. The other crucial factor that might affect these experiments is the repetitive sequences presenting the transcripts. In rice, the repetitive sequences account for nearly 45% of the genome, composed of thousands of retrotransposons, and numerous miniature inverted-repeat transposable element copies. In a recent publication on the improved rice indica genome sequence, Yu et al (2005) reported an ancient whole-genome duplication covering 65.7% of the current genome that can be dated back to a common time before the divergence of grasses. They also identified 18 distinct segmental duplication pairs and massive ongoing individual gene duplications that provide a never-ending source of raw material for gene genesis, and are major contributors to the differences between members of the grass family (Yu et al 2005).

Tiling-path microarray analysis in rice

We have been taking bioinformatics approaches and experimental methods centered on tiling-path microarray to facilitate and improve the rice genome annotation. The tiling-path strategy was first used on rice japonica chromosome 4, in which a total of 15,242 nonredundant polymerase chain reaction (PCR)-amplified genomic fragments were printed on slides as probes and hybridized with RNA samples pooled from six representative rice organs. Our analysis revealed a chromatin-level regulation of both protein-coding genes and transposable element–related genes, in hetero- and euchromatin regions at different developmental stages (Jiao et al 2005).

We also used the Maskless Array Synthesizer (MAS) platform to design higher-resolution tiling experiments, by using 36-mer oligos to cover all the nonrepetitive sequences of the newly assembled indica and japonica genomes. The implementation of this design generated about 6.5 and 6.1 million probe pairs for tiling approximately 60% of the regions (nonrepetitive portions) of the indica and japonica genome, leading to the resolution of each 36-mer oligo interrogating every 60-bp genomic sequence on average. The two sets of 34 and 32 arrays were hybridized with mixed cDNA targets derived from four tissues: seedling root, seedling shoot, panicle, and suspension-cultured cells.

To examine rice tiling-array hybridization conditions, and to establish a data-processing procedure, the pilot analysis was first conducted on two sets of hybridization data representing indica and japonica chromosome 10, both consisting of nearly 3,000 nontransposon gene models. More than 80% of the annotated gene models were verified and nearly 500 novel intergenic transcribed regions were detected by our tiling-array platform. The overview of global signals along chromosome 10 indicates that expression of transcriptome can be related to chromosomal architecture (Li et al 2005).

Using tiling microarrays to experimentally analyze probe hybridization properties

Tiling-path microarray provides an unbiased observation of genome-wide transcription activity and is a powerful approach complementary to computer-based annotation methods. Unlike the traditional cDNA or oligomer microarray, in which probes are preselected to adapt to the calculated optimal parameters and specificities, tiling-path platforms employ the full extent of oligos covering given genomic regions, and thus offer excellent experimental conditions in which to assay the properties of oligos in terms of their specificity and efficiency of hybridization to their corresponding targets. In particular, the following questions can be addressed:

1. The diverse properties of probes produce unequal hybridization intensities despite being located in the same transcribed region. For instance, oligos with high GC content are more easily hybridized with targets than those with low GC content, but they also might introduce false signals if GC composition is beyond a certain percentage. Likewise, oligos within a same gene might vary a lot in hybridization efficiency because of their different

melting temperatures. Therefore, the diversity of probe property is one of the reasonable interpretations for a large proportion of gene models lacking or having unequal hybridization intensities.

2. In rice, duplication events are common phenomena and many remain in the current genome. Although duplicated pairs diverged in both function and sequence during evolution, they still share a similar DNA composition. Cross-hybridization is therefore unavoidable in most tiling-array experiments and biases the accuracy of determining whether a gene is expressed or not.
3. As a common procedure in oligo design, RepeatMasker is generally used to mask highly repeated sequences to avoid redundant oligos being selected, but this procedure inevitably excludes numerous representative oligos, usually from the same members of a gene family, leading to a lack of information for matching the tiling-array signals with gene structure.
4. Positional effect is also a crucial factor that needs to be taken into account because the 3′ end of cDNA is more readily labeled than the 5′ end during reverse transcription.

Experimental design of the high-resolution rice tiling microarray

To determine the impact of GC composition, melting temperature, and sequence repetitiveness of oligo probes from a given gene on their hybridization strength, we selected a 1-Mb portion (from 10 to 11 Mb) from rice japonica chromosome 10 to design a high-resolution tiling microarray. In this design, 36-mer oligos in 5-bp step were used to tile both strands of the given genomic region (Fig. 1A). Some 194,497 oligo pairs were generated in total. Importantly, as an absolutely unbiased survey, all oligos were retained and synthesized. All the oligos were arranged on slides in a "chessboard" layout, which means that every positive feature (interrogating oligo) was surrounded by four negative features (blank position) for background noise subtraction in further data processing. Hybridization targets were prepared from four major rice tissues—seedling root, seedling shoot, panicle, and suspension-cultured cells—labeled with Cy3 dye. Hybridized slides were scanned by a GenePix 4000B scanner (Axon, Foster City, Calif.) using the 532-nm channel and visualized with GenePix Pro 3 image analysis software (Axon).

Tiling-array raw data processing workflow

The raw data were processed in the following order: (1) Positive and negative intensities were extracted separately, and deposited in a uniform format. (2) To smooth the variance between positive and negative intensities, arising from targets' different affinity to DNA and blank linker, a global normalization was assigned to scale positive and negative distribution to a consistent baseline. (3) To screen the background noise, the average intensity of four surrounding negative intensities was calculated and then subtracted from the centric positive intensity. (4) After filtering the negative value, the remaining noise intensities were corrected by using a hypothesis test to

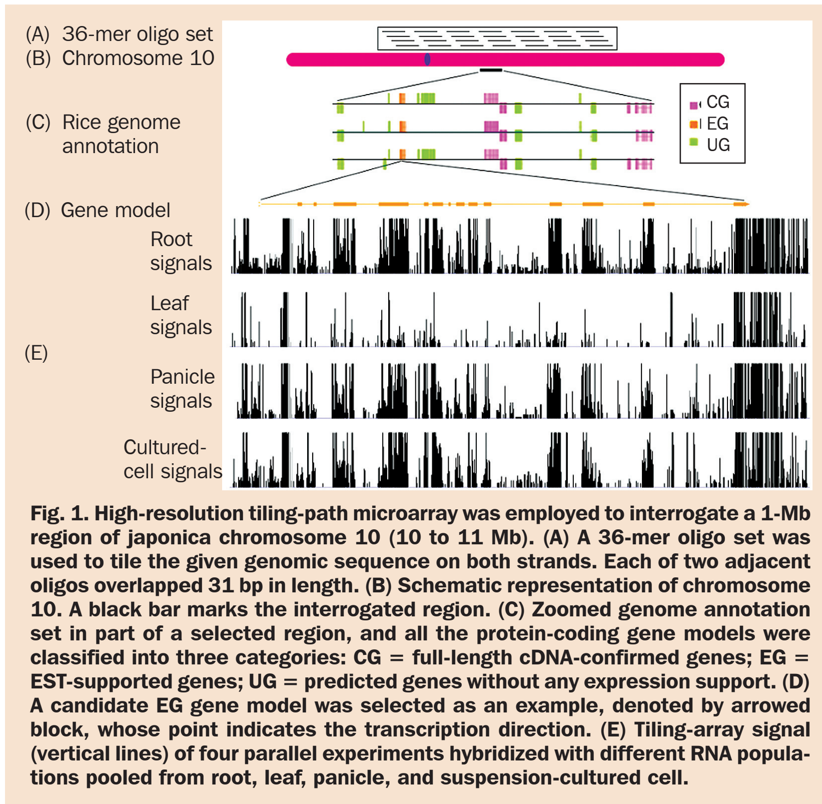


Fig. 1. High-resolution tiling-path microarray was employed to interrogate a 1-Mb region of japonica chromosome 10 (10 to 11 Mb). (A) A 36-mer oligo set was used to tile the given genomic sequence on both strands. Each of two adjacent oligos overlapped 31 bp in length. (B) Schematic representation of chromosome 10. A black bar marks the interrogated region. (C) Zoomed genome annotation set in part of a selected region, and all the protein-coding gene models were classified into three categories: CG = full-length cDNA-confirmed genes; EG = EST-supported genes; UG = predicted genes without any expression support. (D) A candidate EG gene model was selected as an example, denoted by arrowed block, whose point indicates the transcription direction. (E) Tiling-array signal (vertical lines) of four parallel experiments hybridized with different RNA populations pooled from root, leaf, panicle, and suspension-cultured cell.

determine whether a probe represented noise or a signal. The outputs of mature data were then transferred to advanced analysis for mapping signal probes to annotated gene models, scanning for novel transcription units, and further analyzing oligo hybridization properties.

Mapping tiling-array signals to annotated gene models

To measure the detection efficiency of tiling experiments, both annotated gene models and oligos (with their hybridization intensities) were mapped back to the interrogated genomic region according to their chromosomal coordinates. A visualization platform was developed under the PERL environment plus SVG package, to directly assess the matching of gene models and tiling-array signals. In total, 101 BGI annotated protein-coding gene models were localized in the given region, including 36 full-length cDNA- and EST-supported gene models. A majority of the annotated models showed agreement to signal clusters to a certain extent (Fig. 1), and an obvious difference can be observed in that the probe signals in exonic regions were more compactly clustered than in intronic regions and intergenic regions, indicating that a well-designed algo-

rithm could effectively distinguish exons and introns by recognizing characterized signal clusters.

We used the same algorithm to calculate the detection rate of 101 annotated gene models both by a medium-resolution tiling experiment (36-mer oligo with 10-bp interval) and high-resolution tiling experiment (36-mer oligo with 5-bp step). The latter experiment confirmed slightly more gene models than our former analysis (Li et al 2005).

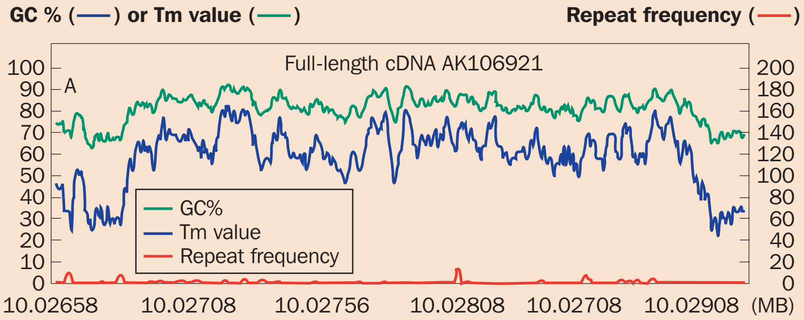
Oligo hybridization properties analysis

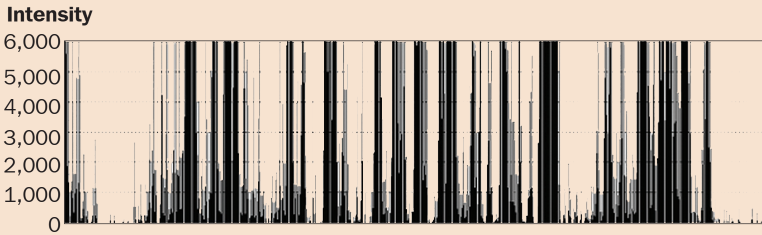
In our oligo hybridization properties analysis, we primarily focused on three fundamental factors that influence hybridization experiments according to prior knowledge and experience in analyzing microarray data: oligo GC composition, melting temperature, and repetitiveness. Two single-exon full-length cDNA verified gene models were selected in our initial analysis. As shown in Figure 2, AK106921 has an open reading frame of 2,635 bp in length. Average melting temperature and GC content of the probes tiling AK106921 are 81.7 ^{o}C and 59.7%, respectively. The abundant signal clusters show a high relationship with GC content and Tm value along the transcribed region, except for the two low-GC-content portions in the 3′ and 5′ end (nearly 30%), causing a shortage of hybridization signals. Another full-length cDNA, AK106577 (3,014 bp), shows a low GC composition (37.57% on average), which leads to an overall scarcity of hybridization signals. In contrast, some highly repeated regions, indicated by red peaks (Fig. 2), show a high correlation with signal clusters. These could be false signals caused by cross-hybridization with other transcripts in the genome.

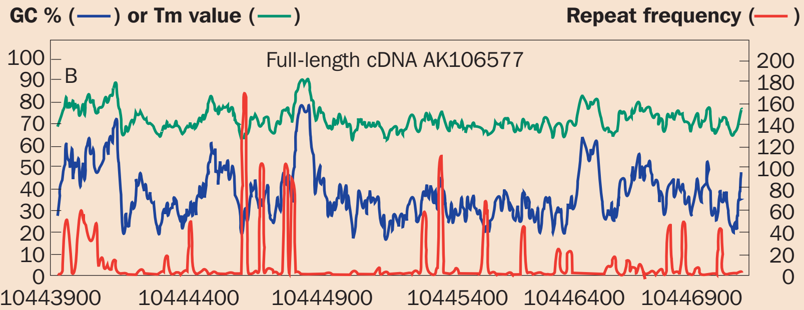
Relationship between GC composition and hybridization strength

One unique property of *Gramineae* genes is compositional gradients in GC content, codon usage, and amino acid usage, along the direction of transcription, beginning at the junction of 5′-UTR and the coding region (Wong et al 2002). As Figure 3A shows, the distribution of oligo GC content sampled from exonic regions has two peaks, with the low GC peak at about 40% and the high GC peak at about 70%. In contrast, the GC composition in intronic regions apparently has no gradient change, and has a single peak at slightly lower than 40%.

High GC content causes unexpected nonspecific hybridization because of increased oligo melting temperature. To estimate this influence, the average hybridization intensity within every GC content range was calculated and plotted (Fig. 3B). From this curve, we can see that hybridization strength was enhanced linearly with the increase in GC content, when GC content is below 63.4%. From 63.4% to 80.5%, hybridization strength reaches a relatively stable status. After this region, from 80.5% to 90.3%, the sharply increased intensities indicate the existence of nonspecific cross-hybridization that affects the hybridization signals. We selected a multiexon full-length cDNA gene model, AK072682, to illustrate the influence of partially high GC composition within genes. Figure 4A shows the trend of GC compositional gradient change in the given







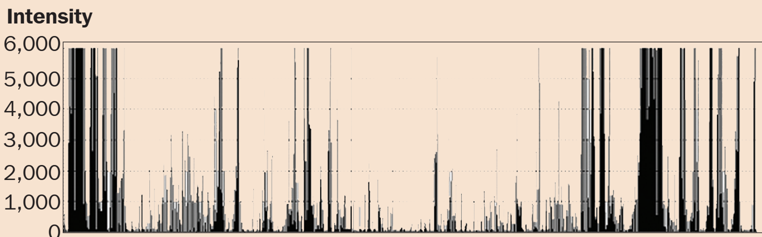


Fig. 2. Overview of three crucial factors that potentially influence the specificity and sensitivity of hybridization experiments: including melting temperature (green curve), GC composition (blue curve), and repetitiveness (red curve) of oligos. Two single-exon full-length cDNA genes were selected to study the relationship among GC%, Tm, repeat, and hybridization strength. (A) AK106921, whose average Tm value and GC% are 81.7 ºC and 59.7%, respectively, has abundant signals except for the two portions of the 3´ end and 5´ end region because of low GC content (30% on average). (B) AK106577, whose average Tm value and GC% are 71.2 ºC and 37.6%, is scarce of hybridization, but some regions show an obvious relationship to a highly repeated feature, as the sharp red peaks indicate.

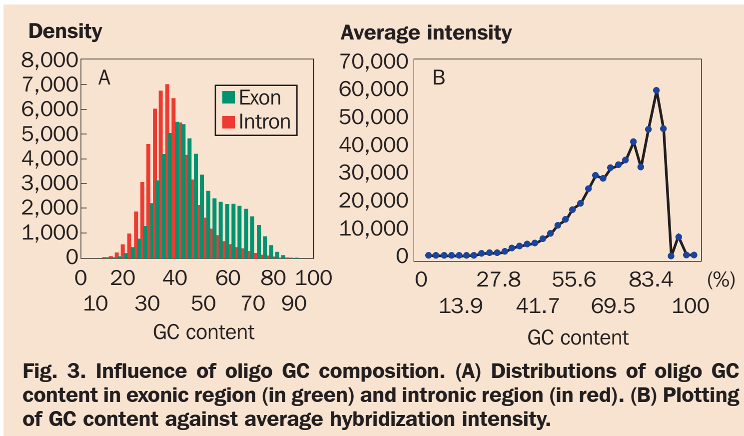


Fig. 3. Influence of oligo GC composition. (A) Distributions of oligo GC content in exonic region (in green) and intronic region (in red). (B) Plotting of GC content against average hybridization intensity.

gene model from transcription start, which is consistent with changes in hybridization intensities (Fig. 4B).

Relationship between oligo melting temperature and hybridization strength

The melting temperature of each 36-mer oligonucleotide was calculated by DAN, a program of the EMBOSS package, based on the nearest-neighbor thermodynamics algorithm. In fact, determination of an appropriate Tm range is crucial for the outcome of hybridization experiments. The Tm value of exonic oligos shows a similar distribution with that of the GC content of exonic oligos, indicating that the potential impact of melting temperature might take effect in the same way as oligo GC composition. In our further study of the impact of melting temperature on hybridization, average hybridization intensity within each Tm region was calculated and plotted, as Figure 5 shows. From this curve, we easily recognized a relatively stable intensity area that corresponds to a Tm range of 82 to 88 ^{o}C, leading us to the conclusion that 85 ±3 ^{o}C might be the optimal temperature for 36-mer oligo-based microarray experiments.

Impact of oligo repetitiveness on hybridization experiments

To evaluate the impact of oligo cross-hybridization, we determined the minimal aligned length and copy number of each probe against all annotated genes in the rice genome. To this end, we blasted all 388,954 oligos against the 45,797 cDNA sequences of rice protein-coding genes released by TIGR annotation. The number of aligned sequences with length ranging from 14 to 35 bp for each oligo provides an estimation of the nonspecific matching with gene models located in other regions of the genome. Surprisingly, the short aligned sequences with lengths of about 15 bp have a relatively high occurrence frequency, which is supposed to be some kind of common repeat motif in a number of genes.

To determine the minimal similarity length that might influence the reliability of hybridization, the average intensity of every similarity length was calculated and

Fig. 4. An example of FLcDNA gene AK072682 to illustrate the positional effect in hybridization experiments, caused by the property of GC compositional gradient in the rice gene. (A) GC content and Tm value gradient change along the transcription direction. (B) The high-GC feature of oligos in the neighborhood of transcription start introduces strongly nonspecific hybridization.

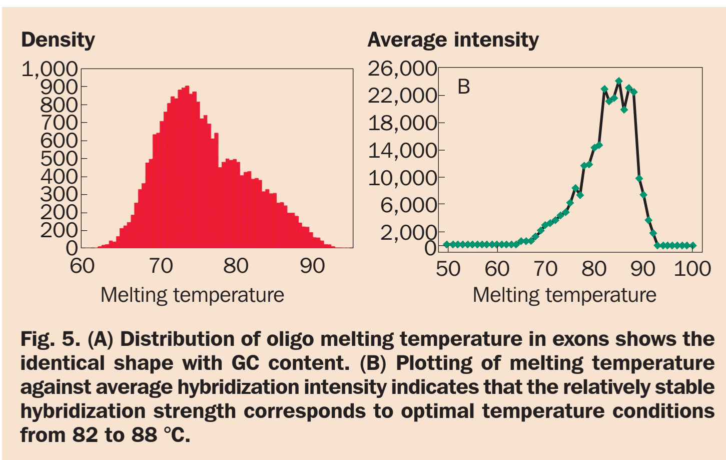


Fig. 5. (A) Distribution of oligo melting temperature in exons shows the identical shape with GC content. (B) Plotting of melting temperature against average hybridization intensity indicates that the relatively stable hybridization strength corresponds to optimal temperature conditions from 82 to 88 °C.

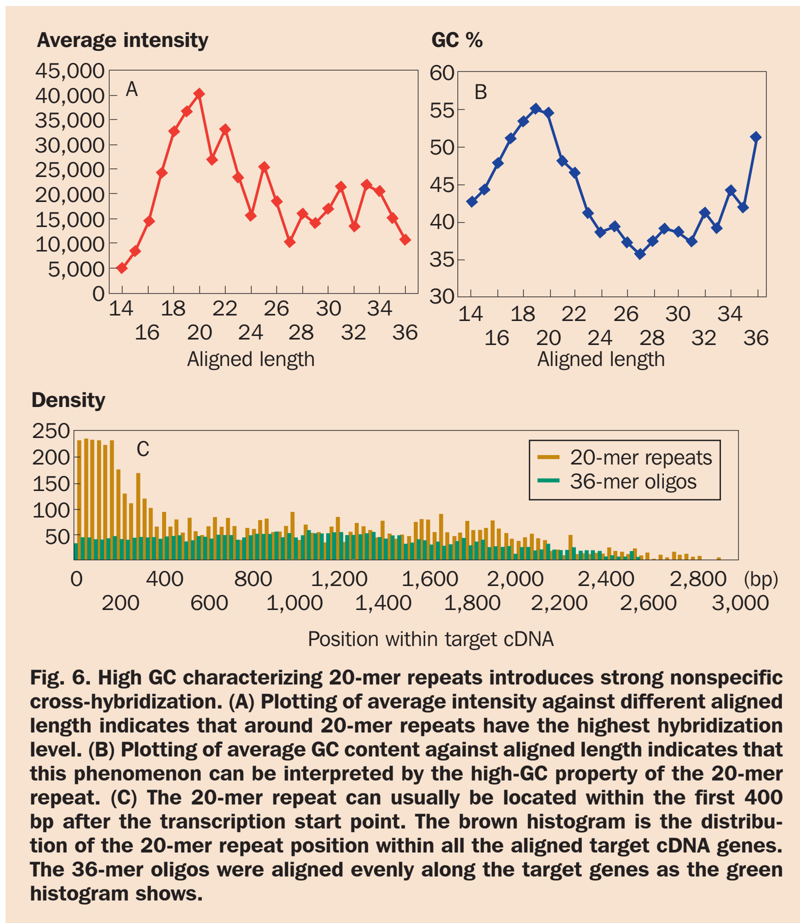


Fig. 6. High GC characterizing 20-mer repeats introduces strong nonspecific cross-hybridization. (A) Plotting of average intensity against different aligned length indicates that around 20-mer repeats have the highest hybridization level. (B) Plotting of average GC content against aligned length indicates that this phenomenon can be interpreted by the high-GC property of the 20-mer repeat. (C) The 20-mer repeat can usually be located within the first 400 bp after the transcription start point. The brown histogram is the distribution of the 20-mer repeat position within all the aligned target cDNA genes. The 36-mer oligos were aligned evenly along the target genes as the green histogram shows.

plotted against them, as Figure 6 shows. Contrary to our expectation, the oligos of around 20-mer similarity length have significantly higher hybridization strength than even 36-mer perfectly matched oligos, designed originally from the interrogated gene models. A further study to survey the relationship of GC composition and oligo similarity can interpret this phenomenon, which is illustrated in Figure 6B. In this plotting of oligo similarity length against average GC content, the 20-mer repeat motif seems to bias toward a higher GC composition than longer repeat motifs above 24 bp. By locating this kind of 20-mer repeat inside the aligned genes, we found that they are more abundant in the 5′ portion of genes than in the middle and 3′ portion of genes, as Figure 6C shows. Therefore, we could conclude that the high GC featured 20-mer repeats and the GC-biased usage of codons in the neighborhood of a gene's transcription start influence the specificity of hybridization experiments and generate most of the false positive signals that partially interfere with verification of gene structure, mostly around the transcription start point.

Conclusions and prospects

The development of tiling-path microarrays provides an efficient way to detect transcriptional activities on a genome scale. In our analysis of a high-resolution rice-tiling array, designed in a 1-Mb region on japonica chromosome 10, we made a series of analyses on three major factors that influence the specificity and efficiency of hybridization: oligo GC content, melting temperature, and repetitiveness. The feature of GC content gradient change from the 5′ end to 3′ end of the rice gene seems to be one of the crucial factors that cause nonspecific and unstable hybridization. Extremely high GC content (above 80%) causes drastically increased cross-hybridization. However, as Tm value ranges become narrower along with an increase in GC content, we identified a relatively stable melting temperature range, from 82 to 88 ^{o}C, for 36-mer oligo-based microarray design. Further analysis of the repetitiveness of all oligos revealed unexpected short repeat motifs around 20 bp in length, having relatively high GC content. These short repetitive motifs could influence the specificity of hybridization and the determination of gene expression level.

References

Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W. 2004. Consistent over-estimation of gene number in complex plant genomes. Curr. Opin. Plant Biol. 7:732-736.

Bertone P, Gerstein M, Snyder M. 2005. Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. Chromosome Res. 13:259-274.

Hoshikawa K. 1993. Science of the rice plant. Vol 1. Morphology. Tokyo (Japan): Nobunkyo. p 133-186.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Jiao Y, Jia P, Wang X, Su N, Yu S, Zhang D, Ma L, Feng Q, Jin Z, Li L, et al. 2005. A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. Plant Cell 17:1641-1657.

Li L, Wang X, Xia M, Stolc V, Su N, Peng Z, Tongprasit W, Li S, Wang J, Wang X, Deng XW. 2005. Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. Genome Biol. 6:R52.

Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. 2005. Applications of DNA tiling arrays for whole-genome analysis. Genomics 85:1-15.

Wong GK, Wang J, Tao L, Tan J, Zhang J, Passey DA, Yu J. 2002. Compositional gradients in Gramineae genes. Genome Res. 12:851-856.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, et al. 2005. The genomes of *Oryza sativa*: a history of duplications. PLoS Biol. 3:e38.

Notes

Authors' addresses: X. Wang, L. Li, and X.W. Deng, National Institute of Biological Sciences, Zhongguancun Life Science Park, Beijing 102206, China; L. Li, V. Stolc, and X.W. Deng, Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, e-mail: xingwang.deng@yale.edu; X. Wang and S. Li, Peking-Yale Joint Research Center of Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China; X. Wang, C. Chen, J. Wang, and S. Li, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 101300, China; V. Stolc, Genome Research Facility, NASA Ames Research Center, MS 239-11, Moffett Field, CA 94035; W. Tongprasit, Eloret Corporation, Sunnyvale, CA 94087.

Tissue culture–induced mutations and overexpression of full-length cDNAs as a tool for functional analysis of rice genes

H. Hirochika, A. Miyao, M. Yamazaki, A. Takahashi, G.K. Agrawal, C. Cheng, Y. Yamashita, M. Harada, H. Nakamura, M. Hakata, and H. Ichikawa

A collection of 50,000 *Tos17*-induced mutant rice lines carrying about 250,000 independent insertions was generated. DNA pools derived from 50,000 lines have been produced for polymerase chain reaction (PCR)–based reverse genetics screening. For *in silico* screening of mutants of genes of interest, a large-scale analysis of the mutants by sequencing the genomic DNA sequence flanking *Tos17* insertions is in progress. To facilitate the functional analysis, the database on phenotypes covering all the mutant lines has been developed. About half of the mutant lines exhibited at least one phenotype. About 5–10% of the mutations were shown to be caused by insertion of *Tos17*, whereas the rest of the mutations were deletions, possibly caused by double-strand break repair and point mutations. These deletion mutations can be detected by the PCR-based screening method, providing a new resource for functional analysis of genes. Considering gene redundancy in rice and the availability of a large number of full-length cDNAs, we have begun producing a new type of activation tagged lines in which 15,000 independent normalized full-length cDNAs are overexpressed under the control of the ubiquitin promoter.

Keywords: insertional mutagenesis, *Tos17*, flanking sequence tag, activation tagging, full-length cDNA, forward genetics, reverse genetics

With the completion of genomic sequencing of rice (International Rice Genome Sequencing Project 2005), rice has established a position as a model organism for both basic and applied research. The next important challenge is to uncover the functions of genes discovered by sequence analysis. Insertional mutagenesis is most suitable for a systematic functional analysis of a large number of genes in the context of whole plants because many mutant lines can be produced at the same time and induced mutations can be easily detected by simple molecular biological methods. In *Arabidopsis*, whose entire genomic sequencing has been completed (Arabidopsis Genome Initiative 2000), mutant populations covering almost all the genes have been produced by using insertion elements, such as T-DNA, *Ac/Ds*, and *En/Spm*. Although these mutant lines are still important resources for forward genetics studies of gene

function, their applications in reverse genetics become more important with the available *Arabidopsis* genome sequence. Two approaches for reverse genetics analysis have been established. One is polymerase chain reaction (PCR) screening of DNA pools from mutants using primers corresponding to a target sequence. The other is flanking sequence tag (FST) analysis, a strategy in which sequences flanking the insertion elements are determined in each mutant line to develop a database of knockout genes that can be searched electronically. Because of the labor- and time-intensive nature of PCR screening, FST database development is now the preferred strategy. In several *Arabidopsis* laboratories (http://signal.salk.edu/tabout.html, http://atidb.cshl.org, and http://genoplante-info.infobiogen.fr), a large collection of FST data, totaling more than 100,000 tags, has been established and opened to the public. Mutant lines linked with FST data are made available through several *Arabidopsis* seed stock centers.

Large-scale insertional mutagenesis has also begun in rice using T-DNA, *Ac/Ds*, *Spm/dSpm*, and the endogenous retrotransposon *Tos17*. The current status of rice mutant resources has been recently summarized by Hirochika et al (2004). More recently, T-DNA insertion mutant lines for forward and reverse genetics studies in rice have been reviewed (An et al 2005). Here, we summarize recent progress in insertional mutagenesis in rice using *Tos17* and its application to forward and reverse genetics studies. In addition, we also discuss our new challenge, the development of activation tagged lines using full-length cDNA.

Forward and reverse genetics using *Tos17*

The unique features of *Tos17* (Hirochika et al 1996, Hirochika 2001, Miyao et al 2003, International Rice Genome Sequencing Project 2005) that make it a powerful genetic tool for forward and reverse genetics 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 ten loci on average are disrupted in each plant regenerated from 5-month-old culture.
3. Integration target loci were widely distributed over the chromosomes, so that random insertion for saturation mutagenesis is feasible. As we have reported, *Tos17* insertions are not evenly distributed on chromosome 1 (Miyao et al 2003). The density of *Tos17* insertions is high in the distal regions of the chromosome, but low in the pericentromeric region. Overall, a similar distribution pattern was observed in all the chromosomes (International Rice Genome Sequencing Project 2005). On chromosome 4, the density of *Tos17* insertions is low in the entire short arm and a part of the long arm, which correspond to heterochromatic regions. These results indicate that *Tos17* prefers the euchromatic region rather than the heterochromatic region. This feature of *Tos17* is suitable for efficient gene disruption. A similar target preference

has also been reported for T-DNA insertion in *Arabidopsis* (Alonso et al 2003) and rice (An et al 2005).

4. Induced mutations are stable and germinally transmitted in the next generation.
5. The original copy number is quite low, one to five, depending on varieties, so that it is easy to identify the transposed copy responsible for the specific mutation.
6. Rearrangements are very rare at junctions between *Tos17* ends and flanking host sequences, so that screening of mutants by PCR and analysis of disrupted genes can be carried out with high efficiency.
7. The transposon is endogenous, so that screening and characterization of mutants in the field are possible without any environmental concern.

However, some features might also be disadvantageous. For example, *Tos17*-insertion targets tend to cluster, whereas rice genes are distributed throughout the chromosomes (Miyao et al 2003). This means that more mutants are required for saturation than expected from random transposition. Transposition of *Tos17* via a copy-and-paste mode means that no revertants can be obtained, although revertants are useful for confirming that a gene is tagged and can be obtained from the mutants induced by class II elements such as *Ac/Ds*.

Forward genetics

Traditional transposon tagging is still an important method for cloning important genes for functional analysis. R_1 (M_2) generations of regenerated rice (= *Tos17*-insertion mutant lines) were screened for mutants based on the phenotypes in the field. About 50% of the regenerated lines examined showed many kinds of visible mutant phenotypes, such as dwarf, sterile, yellow, albino, virescent, viviparous, brittle, and spotted leaf. In addition to screening in the field, screening *in vitro* was also conducted to isolate genes involved in salt-stress tolerance, and root growth and development. 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 were recessive, although some did not segregate in a 1:3 ratio. Because the copy number of *Tos17* is low enough to visualize each transposed copy, the *Tos17* copy causing the specific mutation can be identified by a simple DNA gel-blot analysis. Finally, the causative gene can be isolated by using IPCR (inverse PCR) or TAIL-PCR (thermal asymmetric interlaced PCR). The first direct evidence for the feasibility of tagging using *Tos17* was shown by cloning genes involved in abscisic acid biosynthesis (*OsABA1*, *OsTATC*: Agrawal et al 2001). By using this forward genetics strategy, many other genes have been cloned. Those include genes involved in gibberellin biosynthesis (genes for *ent*-copalyl diphosphate synthase and *ent*-kaurene synthase), meiosis (*PAIR1*: Nonomura et al 2004a; *PAIR2*: Nonomura et al 2004b), sporogenesis (*MSP1*: Nonomura et al 2003), and secondary cell-wall biosynthesis (*OsCesA4, OsCesA7, OsCesA9*: Tanaka et al 2003).

Although gene tagging with *Tos17* is a powerful strategy for cloning genes, one fundamental problem should be noted: the relatively low tagging efficiency (5–10%) (Hirochika 2001, Agrawal et al 2001). This indicates that tissue culture–induced mechanisms other than *Tos17* insertions cause untagged mutations with high frequency. Several lines of evidence indicated that untagged mutations involve deletions, possibly caused by double-strand break repair and point mutations (G.K. Agrawal et al, unpublished results). Recent studies showed that these deletion mutations, providing complete loss-of-function mutations, are also important resources for functional analysis of genes (Agrawal et al 2005). These mutations can also be used as a resource for reverse genetics by using TILLING (Till et al 2003) and PCR-based screening of deletions.

Reverse genetics

For reverse genetics, DNA pools for PCR screening or the FST database are needed to identify mutations in genes of interest. The first direct evidence for the feasibility of the use of *Tos17* for the PCR-screening strategy was shown by screening for a mutant of the homeobox gene (*OSH15*) (Sato et al 1999). DNA pools from 50,000 *Tos17* insertion lines have been constructed and results from screening these DNA pools suggest a success rate of 50% for a given target sequence (H. Hirochika, unpublished results). Because of the finite nature of the DNA pool, it is not practical to distribute DNA pools publicly. In addition, PCR screening is labor- and time-intensive. Thus, generating an FST database as a public resource is a more desirable and practical strategy. As of October 2005, 21,223 independent *Tos17*-flanking sequences from 6,713 lines have been determined and the FST database is open to the public (Miyao et al 2003; http://tos.nias.affrc.go.jp). It is expected that FSTs from 50,000 lines will be determined within a few years. Some other research groups constructing insertional mutant lines have also constructed FST databases, such as T-DNA FST databases at Pohang University (www.postech.ac.kr/life/pfg/risd) and at Génoplante (http://genoplante-info.infobiogen.fr/oryzatagline). Currently, FST data totaling more than 56,000 tags are generated. All the FST data are expected to be anchored to the public IRGSP genome sequence and positions of both predicted genes and FST data will be shown graphically on the physical map. *Tos17*-induced mutant lines linked with FST data are made available through the Rice Genome Resource Center (www.rgrc.dna.affrc.go.jp).

By using reverse genetics strategies with *Tos17*, functions of many genes have been demonstrated. Those include genes such as *phytochrome A* (Takano et al 2001), *phytochrome C* (Takano et al 2005), *GAMYB* (Kaneko et al 2004), *OsMADS1* (Agrawal et al 2005), *OsMADS3* (Yamaguchi et al 2006), *Mg-chelatase* (Jung et al 2003), *chlorophyll a oxygenase* (Lee et al 2005), and a putative voltage-gated Ca^{2+} channel gene, *OsTPC1* (Kurusu et al 2004, 2005).

To facilitate the functional analysis of genes, it is also important to make databases on phenotypes. All the *Tos17*-induced mutant lines were observed at four different growth stages and classified based on 54 phenotypes. About 50% of the lines exhibited at least one of the phenotypes. To maximize the utility of phenotype

databases, a common vocabulary for describing mutants should be adopted among the mutant producers in different projects. The Plant Ontology Consortium vocabulary provides such a common framework and integrates different fields of expertise specific to plants. If the phenotype database is linked to the FST database, functional assignment of genes will be greatly accelerated.

Limitations and future prospects

Several problems or limitations associated with features inherent in insertional mutagenesis were noted when functional analysis was carried out. One obvious problem is the lack of a mutant phenotype. In *Arabidopsis*, less than 2% of the insertional mutant lines showed a mutant phenotype. This problem may be largely due to gene redundancy. One possible solution is a combination of mutations by crossing, as has been successfully demonstrated in *Arabidopsis* (Liljegren et al 2000, Pelaz et al 2000). Considering the higher gene redundancy in rice (International Rice Genome Sequencing Project 2005), this problem may become more serious. Another problem is background mutations induced during the production of mutant lines, which make it unreliable to directly correlate phenotype with insertion mutations. Most of the background mutations are recessive and probably deletions and point mutations as described in the previous section.

To solve those problems, activation tagging inducing dominant mutations will be a complementary strategy. More than 50,000 activation-tagged lines have been generated already by using T-DNA with multimerized 35S enhancers (Jeong et al 2002). Considering the random insertion of T-DNA, a large number of tagged lines may be required to overexpress all the genes. Recently, a more efficient activation tagging strategy has been developed in *Arabidopsis* (T. Ichikawa et al, unpublished results). This system, called FOX Hunting (full-length cDNA over-expressor gene hunting), uses transformation with T-DNA carrying an *Arabidopsis* full-length cDNA library under the control of the modified 35S promoter. By using this system, random overexpression of the *Arabidopsis* full-length cDNA is possible. About 8% of the FOX lines showed visible mutant phenotypes, whereas only 0.1% to 1% of the activation-tagged lines showed the phenotypes. Another advantage of the FOX hunting system is that the genes responsible for the phenotype can be easily determined. Considering gene redundancy in rice and the availability of 32,000 full-length cDNAs (Kikuchi et al 2003), we have begun producing FOX lines of rice in which 15,000 independent rice full-length cDNAs are overexpressed under the control of the ubiquitin promoter (H. Ichikawa et al, unpublished results). So far, about 5,000 lines have been produced and about 10% of the lines showed visible mutant phenotypes. The FOX hunting system will greatly accelerate the functional analysis of genes.

References

Agrawal GK, Yamazaki M, Kobayashi M, Hirochika R, Miyao A, Hirochika H. 2001. Screening of the rice viviparous mutants generated by endogenous retrotransposon *Tos17* insertion: tagging of a zeaxanthin epoxidase gene and a novel *OsTATC* gene. Plant Physiol. 125:1248-1257.

Agrawal GK, Abe K, Yamazaki M, Miyao A, Hirochika H. 2005. Conservation of the E-function for floral organ identity in rice revealed by the analysis of tissue culture-induced loss-of-function mutants of the *OsMADS1* gene. Plant Mol Biol. 59:125-135.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana.* Science 301:653-657.

An G, Jeong DH, Jung KH, Lee S. 2005. Reverse genetic approaches for functional genomics of rice. Plant Mol. Biol. 59:111-123.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.

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.

Hirochika H. 2001. Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr. Opin. Plant Biol. 4:118-122.

Hirochika H, Guiderdoni E, An G, Hsing Y, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang Q, Pereira A, Sundaresan V, Leung H. 2004. Rice mutant resources for gene discovery. Plant Mol. Biol. 54:325-334.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Jeong DH, An S, Kang HG, Moon S, Han JJ, Park S, Lee HS, An K, An G. 2002. T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. 130:1636-1644.

Jung K-H, Hur J, Ryu C-H, Choi Y-J, Chung Y-Y, Miyao A, Hirochika H, An G. 2003. Characterization of a rice chlorophyll-deficient mutant using the T-DNA gene-trap system. Plant Cell Physiol. 44:463-472.

Kaneko M, Inukai Y, Kaneko M, Ueguchi-Tanaka M, Itoh H, Izawa T, Kobayashi Y, Hattori T, Miyao A, Hirochika H, Ashikari M, Matsuoka M. 2004. Loss-of-function mutations of the rice *GAMYB* gene impair the α-amylase expression in aleurone and flower development. Plant Cell 16:33-44.

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379.

Kurusu T, Sakurai Y, Miyao A, Hirochika H, Kuchitsu K. 2004. Identification of a putative voltage-gated Ca^{2+}-permeable channel (OsTPC1) involved in Ca^{2+} influx and regulation of growth and development in rice. Plant Cell Physiol. 45:693-702.

Kurusu T, Yagala T, Miyao A, Hirochika H, Kuchitsu K. 2005. Identification of a putative voltage-gated Ca^{2+} channel as a key regulator of elicitor-induced hypersensitive cell death and mitogen-activated protein kinase activation in rice. Plant J. 42:798-809.

Lee S, Kim JH, Yoo ES, Lee CH, Hirochika H, An G. 2005. Differential regulation of chlorophyll a oxygenase genes in rice. Plant Mol. Biol. 57:805-818.

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF. 2000. *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature 404:766-770.

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H. 2003. Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771-1780.
Nonomura K, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, Kurata N. 2003. The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. Plant Cell 15:1728-1739.
Nonomura K, Nakano M, Fukuda T, Eiguchi M, Miyao A, Hirochika H, Kurata N. 2004a. The novel gene *HOMOLOGOUS PAIRING ABERRATION IN RICE MEIOSIS1* of rice encodes a putative coiled-coil protein required for homologous chromosome pairing in meiosis. Plant Cell 16:1008-1020.
Nonomura KI, Nakano M, Murata K, Miyoshi K, Eiguchi M, Miyao A, Hirochika H, Kurata N. 2004b. An insertional mutation in the rice *PAIR2* gene, the ortholog of *Arabidopsis ASY1*, results in a defect in homologous chromosome pairing during meiosis. Mol. Genet. Genomics 271:121-129.
Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000. B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature 405:200-203.
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.
Takano M, Kanegae H, Shinomura T, Miyao A, Hirochika H, Furuya M. 2001. Isolation and characterization of rice phytochrome A mutants. Plant Cell 13:521-534.
Takano M, Inagaki N, Xie X, Yuzurihara N, Hihara F, Ishizuka T, Yano M, Nishimura M, Miyao A, Hirochika H, Shinomura T. 2005. Distinct and cooperative functions of phytochromes A, B, and C in the control of de-etiolation and flowering in rice. Plant Cell 17:3311-3325.
Tanaka K, Murata K, Onosato K, Miyao A, Hirochika H. 2003. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiol. 133:73-83.
Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S. 2003. Large-scale discovery of induced point-mutations with high-throughput TILLING. Genome Res. 13:524-530.
Yamaguchi T, Lee DY, Miyao A, Hirochika H, An G, Hirano H. 2006. Functional diversification of the two C-class genes, *OsMADS3* and *OsMADS58*, in *Oryza sativa*. Plant Cell 18:15-28.

Notes

Authors' address: National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan.

Analysis of genome sequences from the maternal and paternal parents of an elite rice hybrid

Jun Yu, Gane K.-S. Wong, Siqi Liu, Jian Wang, and Huanming Yang

We have initiated a genome project in China, the Superhybrid Rice Genome Project (SRGP), to understand the molecular basis of hybrid vigor. The early phase of the project is to sequence 93-11 and PA64S, the paternal and maternal parents of the hybrid rice strain LYP9. Preliminary analysis on genomic sequences from the parental cultivars indicates that hybrid vigor may be more complex at the molecular level than previously proposed, which is shaped, through complex and meticulous breeding practices, by intricate genetic and functional complementation processes attributable largely to variations in protein-coding sequences, regulatory elements, epigenetics, and posttranslational modifications of gene products. We are in a process and collaborating with other research groups in rice biology to acquire, in a broad spectrum, transcriptomic and proteomic data of the triad from different tissues, at multiple developmental stages and with different methods. The study should yield useful candidate genes and genetic markers for further investigations in molecular and functional details. Based on the information acquired, the SRGP initiated at the Beijing Genomics Institute will continue to map domestication-related genes and loci, based on the rich diversity of resources available in rice.

Keywords: superhybrid rice, genome sequence, hybrid vigor, gene expression

The Superhybrid Rice Genome Project (SRGP) in China has entered its last phase. The first phase of the project was to produce a complete genome sequence of 93-11, and the paternal cultivar of LYP9 (Lu and Zhou 2000), which is a popular indica rice strain in southern China. We first published a draft sequence of this indica rice variety to reveal its genes and genome organization (Yu et al 2002) and then finished the draft by adding more sequence data. We also compared the finished sequences with those of a japonica rice cultivar (Goff et al 2002). This analysis allowed us to reveal different types of sequence duplications and to gain insight into the evolutionary history of domesticated rice (Yu et al 2005).

The main theme of the SRGP's Phase II efforts is to develop the genome sequence of PA64S, the maternal cultivar of LYP9. PA64S has a complex breeding history, with

a major genetic background of indica (estimated to be about 55%) and minor gene flows from japonica and javanica. In this phase of the project, we also need to acquire a large amount of gene expression data that are generated with various technologies, including EST, SAGE, microarray (cDNA-based or genome-sequence-based), and proteomics. Data about protein modifications and epigenetic changes would also be generated. These different types of data are important in deciphering the molecular profile of hybrid vigor since we already have the full sequences of the two parents of a rice hybrid. Comparative analyses among the three basic rice genomes will allow us to look into many of the molecular details, including precise sequence, sequence variations, and expression changes among unique and shared genes among the maternal, paternal, and F_1 strains as well as other members of their close pedigrees. Rice scientists in China are preparing adequate genetic materials essential for mapping discrete phenotypic traits related to the triad.

The final phase (Phase III) of the SRGP has been aiming at producing molecular and informatic tools for rice biologists and breeders (Zhao et al 2004); it is most likely a composite of several open-ended efforts in collaboration with plant biologists, geneticists, and rice breeders who are interested in in-depth studies in rice biology at the molecular level. We have put together rice microarrays and publicized our databases to facilitate such collaborative efforts, which have been and will be updated over time when new data are generated. We believe that efforts to discover adequate numbers of genetic markers and to collect germplasm, including that of wild rice and breeding/crossing intermediates, are of essence for understanding the complex molecular mechanisms of hybrid vigor.

Rice genome sequences produced through a whole-genome shotgun approach are of high quality

The map-based sequence of the japonica rice genome published in *Nature* by the International Rice Genome Sequencing Project (IRGSP 2005) is a major accomplishment in rice biological research, adding the third complete rice genome sequence in addition to the two previously published genomes completed with a whole-genome shotgun (WGS) approach (Beijing indica and Syngenta japonica; Yu et al 2002, 2005, Goff et al 2002). Although finished according to different principles (such as map-based or gene-centric sequence assembly) and resources, the qualities of the sequences are comparable by most of the mapping and sequencing quality measures, including order and orientation of the assembled contigs and scaffolds (clone-based and/or sequence-based), nucleotide quality of the sequences, and completeness of gene coverage. When effort and cost are both carefully considered, the WGS approach is supreme since a comparative-grade sequence finished with a clone-by-clone (CBC) approach requires approximately 40-fold more reagents and approximately 10-fold more personnel effort than the WGS-derived near-perfect finished sequence (Blakesley et al 2004). This statement is proven valid for mammalian-sized animal genomes and the rice genomes. Furthermore, over the last decade, Bennetzen and colleagues (SanMiguel et al 1996, Bennetzen 2000) have demonstrated that intergenic regions of grass genomes

are full of nested retrotransposons. It was clear that any WGS method would have trouble assembling these regions, but it was just as clear that no such effort would be needed, especially for plant genomes that have repetitive sequence contents of nearly 50% (such as the rice genome) or over (such as the maize and wheat genomes). Intergenic regions of the plant genomes evolve very rapidly and are essentially devoid of functional sequences. For instance, the enormous variation in the intergenic regions between two rice subspecies sequenced thus far is 27.6 SNP per kb compared with 3.0 SNP per kb in coding regions (Yu et al 2005). In addition, research to study gene and genome variations in a context of population genetics has not begun on a reasonable scale in rice as what was reported for *Arabidopsis*, for which a genome-wide excess of rare alleles and variations between genomic regions were found (Nordborg et al 2005). Aside from the evaluation of various quality issues, discrepancies among sequence variations can be resolved only with empirical data by population genetics studies (e.g., Tang et al 2004, Tian et al 2006).

Although differences do exist between the sequences acquired through the two distinct techniques, the final genome assemblies are actually in remarkable agreement if 411-Mb assembled sequences from 93-11 are used for an orthodox comparison. Unfortunately, the published analysis by IRGSP used only 258 Mb of the 411 Mb in our submission and, not surprisingly, only 68.3% of the rice genes discovered in the japonica variety were found in our indica assembly. This result directly contradicts our analyses and series of reports on not only genome assemblies of indica rice strain 93-11 but also those of the same cultivar produced by Syngenta and re-assembled by BGI. We are in a process to produce more detailed annotations for the rice genomes, especially to use a combination of aligned genomic sequences and empirical gene expression data generated with a variety of methods for RNA and protein analyses.

Another misunderstood parameter is the mean size of the sequence contigs (8.2 kb on average). First, contigs are linked together, with the correct order and orientation, to create much larger scaffolds and super-scaffolds. Second, most of the genome is in a small number of large sequences, but there are also a large number of small sequences. Short and unanchored sequences often contain genes that are surrounded by large clusters of repetitive sequences that are mutation-prone in bacterial cloning systems as well as fast-changing intergenic sequences. The more useful metric popularized by the Human Genome Project is N50 size, that size above which half the genome length is found. By this metric, our contigs and super-scaffolds are 23 kb and 8.3 Mb, respectively. From the beginning of our sequencing efforts, we have used a gene-centric approach for the rice genome sequencing projects: to discover all genes, to achieve moderate sequence contiguity, and to close gaps in repeat-rich regions.

Gene content can be more reliably determined by direct comparisons of experimentally confirmed genes and their sequences to the assembled genome. We use 19,079 nonredundant full-length cDNAs (Kikuchi et al 2003), referred to as nr-KOME, to verify and evaluate sequence assemblies, gene predictions, and gene annotations. When we insist that these genes be aligned in one piece, without fragmentation, both WGS-derived genomes (the Syngenta japonica and the Beijing indica) are over 91.2% complete. The union is 98.1% complete, and requiring that the genes be anchored

to the physical map only brings this number down to 97.7%. Across all 19,079 nr-KOME defined genes that are aligned over WGS- and CBC-derived assemblies of chromosome 1 (arguably the best finished chromosome from all the projects), only 72 bp (0.0004%) of the exons and 0.02% of the introns are collectively discrepant. Other chromosomes are not easily compared because they are finished at different quality. We noticed that 287 of 575 unfinished gaps, denoted in the IRGSP sequence by a string of 50 to 100 N's, are found within chromosome 11.

Another important issue for genome analysis and gene annotation is that it is absolutely essential to align the genome sequence rather than just compare sequences with certain blast-derived computing tools because faulty gene predictions, true sequence variations, and massive gene duplications all interfere with gene identification and subsequent functional annotation. In handling piles of orthologous and paralogous genes, it is very easy to make simple mistakes, such as setting inappropriate thresholds or filters, resulting in biased data sets. Genome sequences, WGS- or CBC-derived, are certainly alignable based on genetic and physical markers that are positioned (ordered and oriented) on the chromosomes. The alignment can be checked with a set of experimentally verified gene sequences, such as full-length cDNAs. The error rates, such as false positives or false negatives in gene alignment that are pinpointed to each nucleotide sequence or exons (introns), can be readily calculated (Wang et al 2003). A subtle yet important point to be aware of is that the sequence variations between indica and japonica are different from those of population-based data within a single species.

Preliminary analysis of the PA64S genome assembly

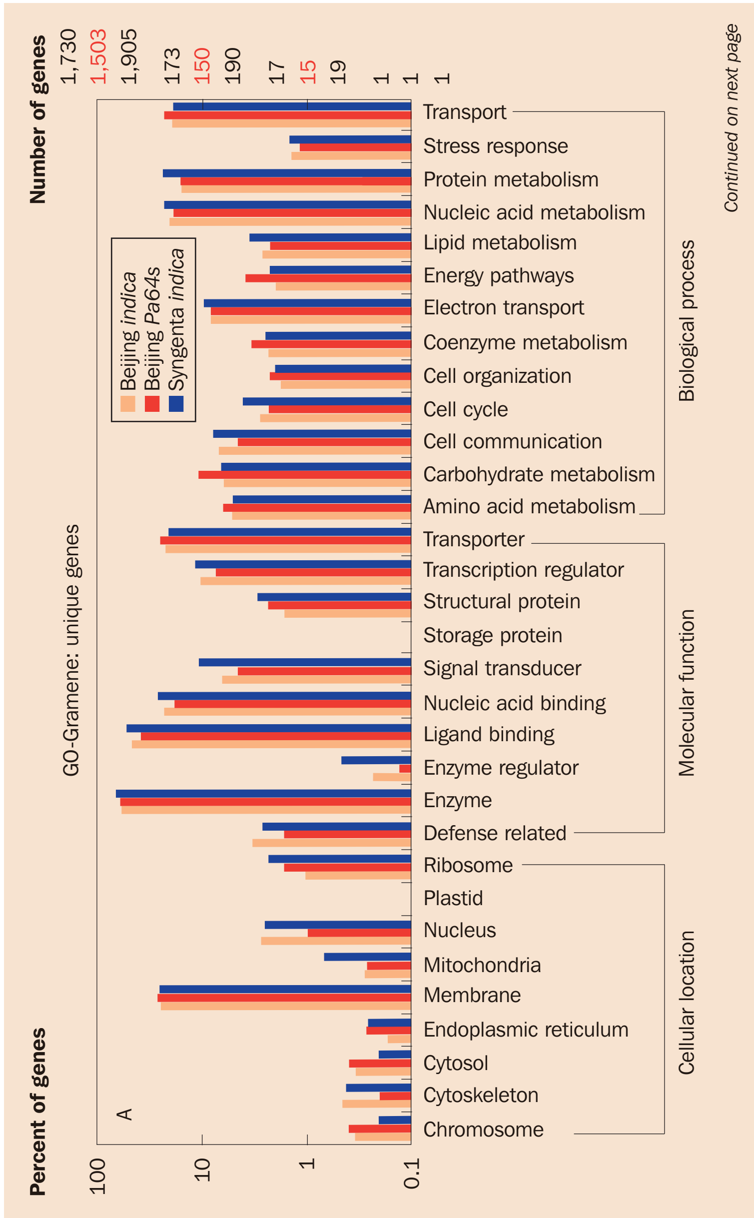
We acquired the PA64S genome sequences with the WGS approach to an effective coverage over sixfold of the genome. The sequences have been assembled with an improved procedure to eliminate interference from repetitive sequences, based on a new protocol to remove and assemble repeats (Zhong et al 2003, Li et al 2005). The PA64S genome is assembled into 105 super-scaffolds with an accumulated size of over 404 Mb, slightly smaller than those of Nipponbare (Syngenta japonica) and 93-11 (Table 1). The contiguity of the assembly can be compared to other assemblies by N50 (at the scaffold size, 50% of the genome is covered by larger scaffolds); for PA64S, the N50 value is about 9 Mb, slightly larger than for 93-11 but slightly smaller than for Nipponbare.

We have predicted genes from PA64S and compared them with those from other genome assemblies with identical procedures. We classify the genes into two basic categories, shared and unique. The genes shared by all three genomes have very small differences and so do the genes shared by two of them (data not shown). However, there are obvious but minor differences among the unique genes between the two Nipponbare assemblies produced by different groups and by using different sequencing and assembling methods, although their genes can be predicted with a single method (Fig. 1). For instance, there are many over- and underpredictions in several categories found in the IRGSP assembly, but they are not obvious in the Syngenta assembly.

Table 1. Summary of sequence assemblies of the Beijing indica (93-11), PA64S, and Syngenta japonica.

	Mapped			Unmapped (≥2K) total size	Unmapped (<2K) total size	Unassembled equivalent size	Total
	No. of pieces	N50 size[a]	Total size				
Contigs							
Maternal PA64S				334,143,633	35,500,346	34,120,601	403,764,580
Paternal 93-11				389,556,605	36,210,950	39,516,973	465,284,528
Syngenta japonica				368,152,741	16,954,192	47,158,342	432,265,275
Scaffolds							
Maternal PA64S				338,821,818	30,822,161	34,120,601	403,764,580
Paternal 93-11				399,914,958	26,914,997	39,516,973	466,346,928
Syngenta japonica				372,391,160	13,633,973	47,158,342	433,183,475
Super-scaffolds							
Maternal PA64S	105	9,012,520	329,551,027	18,662,668	22,105,424	34,120,601	404,439,720
Paternal 93-11	149	8,274,516	373,860,499	52,481,238	15,143,962	39,516,973	481,002,672
Syngenta japonica	119	11,616,395	352,843,970	38,292,863	10,911,800	47,158,342	449,206,975

[a]Contig sizes are in basepairs. The equivalent sizes of unassembled sequences are calculated based on the effective coverage of 6.01x.



Continued on next page

Figue 1 continued.

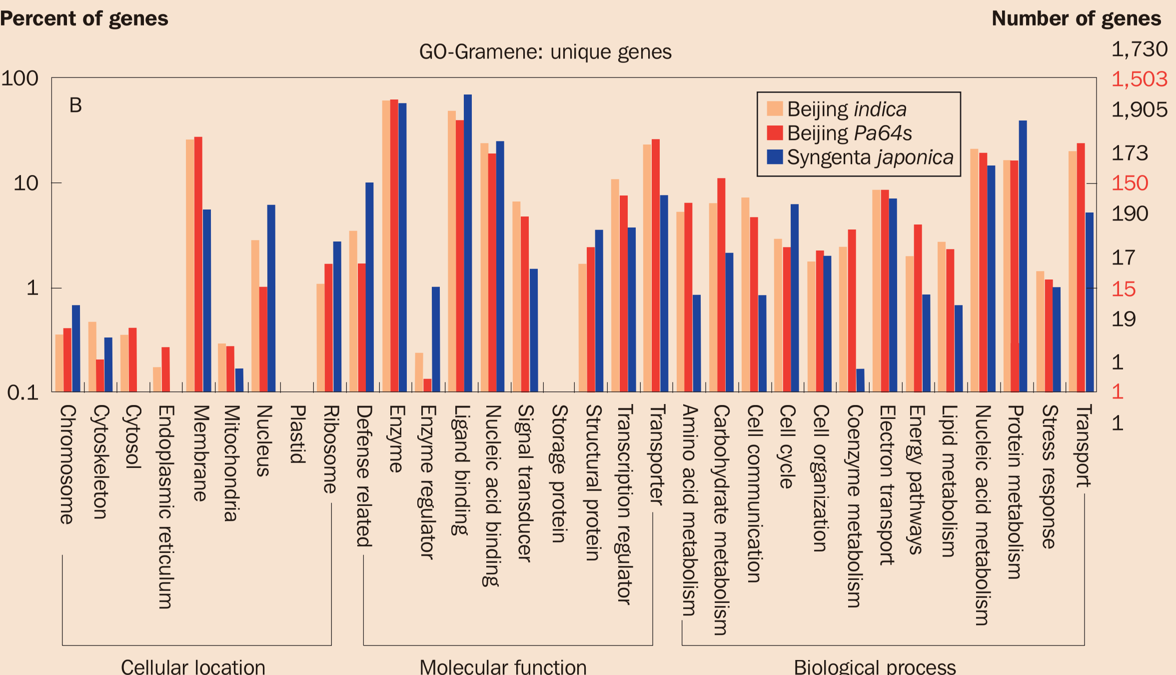


Fig. 1. A comparison of unique genes among three sequenced rice genomes with a whole-genome shotgun approach (A) and a clone-by-clone approach (B). The numbers are genes predicted with Fgenesh for four genome assemblies: Beijing indica (93-11), PA64S, IRGSP japonica (Nipponbare), and Syngenta japonica (Nipponbare). The red horizontal lines in the two panels serve as a visual aid to emphasize the number of genes at the levels. The genes are classified with GO database annotations at a shallow level.

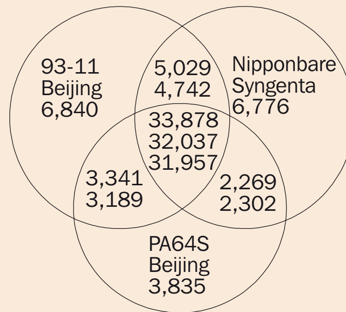


Fig. 2. Gene distribution among three rice genome assemblies. Only the predictions based on a whole-genome shotgun approach are shown due to subtle differences in gene identification and genome assembling methods. The number of genes shared by all three genomes is listed as 93-11 at the top, Nipponbare in the middle, and PA64S at the bottom. The number of genes shared by two genomes is ordered as follows: 5,029 from 93-11, 2,269 from Syngenta Nipponbare, and 3,189 from PA64S.

These categories concern genes in rather diverse functional classes, such as nucleus, defense-related, energy regulators, signal transducers, co-enzyme metabolism, and lipid metabolisms, making the interpretation unfeasible. We currently are not taking any of these particular results for serious interpretations but are starting to verify some of these genes experimentally in the three sequenced subspecies as well as in other related rice strains. Since it is difficult to explain the differences in genes identified between the two japonica assemblies, we decided to take the WGS product for further analysis at this point.

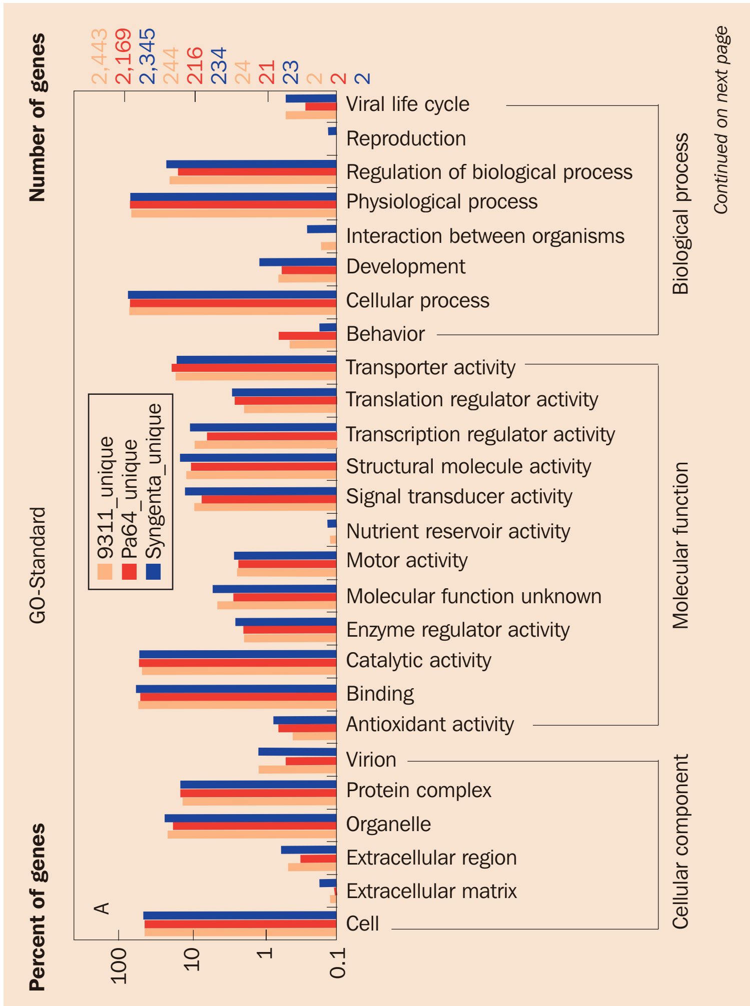
We predicted around 32,000 to 34,000 genes that are shared by all three genome assemblies (Fig. 2). These numbers were generated by alignment-based cross-comparisons after the removal of intergenic repeats. More genes were identified in 93-11 than in what Syngenta found in Nipponbare and PA64S. Pairwise comparisons yielded similar results regarding the number of shared genes. The PA64S assembly also gave rise to fewer unique genes. The smaller number of genes discovered in PA64S is in part due to a slightly lower effective coverage of the genome sequences compared with that of other genome assemblies; the PA64S genome also appears significantly smaller than the other two (Table 1). About 10% to 15% of the genes are shared between the two assemblies or are unique to each assembly, with an exception in which the shared genes between Nipponbare and PA64S or unique to PA64S are only 6% of the total.

It is very important to understand these unique or shared genes in each comparison before identifying their functions and relatedness to hybrid vigor. Some of them may turn out to be true-genome-specific or even cultivar-specific.

We have further investigated these genome-related groups of genes according to the Gene Ontology database by comparing the classification patterns (relative number of genes in each category) among total, shared, and unique genes from all three assemblies. First, the distribution patterns of genes in the specified groups are similar, regardless of whether they are the total from each genome assembly, shared among them, or unique to each (data not shown). This result suggests that these unique or partially shared genes from each assembly are randomly distributed in each genome; this may be a simple consequence of frequent gene loss or duplication over time. We did not reveal, by simple classification schemes, major specific classes of genes among the high-level functional groups although there are many subtle differences in almost every category of genes. Second, we classified the genes further into subcategories to show numerous subtle differences among the three sets of cultivar-specific genes in detailed categories (Fig. 3). This result emphasizes that it is essential to look into functional and expression details of these genes individually and we are in the process of verifying the expression of these genes, following some early attempts by using proteomic and gene-profiling techniques (Bao et al 2005, Zhao et al 2005).

In search of genes and their variations involved in hybrid vigor

Based on the limited data generated at BGI, we believe that hybrid vigor is not the effect of a few genes and its molecular mechanisms are not going to be worked out by characterizing a few functionally important genes or by simply comparing gene and protein expression profiles between the parental and hybrid cultivars since the genetically distant lines may have unparallel gene expressions and even growth timing. A series of experiments has to be carefully designed to follow a set of genes or relevant regulatory or functional networks among the triad and also its segregation lines. The experimental design can be very complex and even endless. An example of such a design and some of the preliminary results are shown in Table 2. We have learned a few important lessons in this type of study. First, only a few genes were found differentially expressed when gene expression patterns were compared in a pair-wise fashion and between two cultivars among the developing leaf tissues. Other comparisons, such as root and panicle tissues, should yield similar results. Second, the relationship among these genes requires further investigation since these genes may or may not overlap when they are compared vertically (over different developmental stages) and across three cultivars. Third, when we examined the genes expressed (with normalization of the experimentally acquired raw data) in all samples, we found that only less than one-third of them are detectable in all tissue samples (data not shown). The result suggested that genes expressed at low levels may not be detected by microarray alone and confirmation with different types of experiments (such as EST, SAGE, quantitative PCR, and allele-specific PCR) is of essence. Fourth, most of the popular methods are not ideal for detecting alternatively spliced gene variants



Number of genes
GO-Standard
Percent of genes
A
2,443
2,169
2,345
244
216
234
24
21
23
2
2
2
9311_unique
Pa64_unique
Syngenta_unique
100
10
1
0.1
Viral life cycle
Reproduction
Regulation of biological process
Physiological process
Interaction between organisms
Development
Cellular process
Behavior
Transporter activity
Translation regulator activity
Transcription regulator activity
Structural molecule activity
Signal transducer activity
Nutrient reservoir activity
Motor activity
Molecular function unknown
Enzyme regulator activity
Catalytic activity
Binding
Antioxidant activity
Virion
Protein complex
Organelle
Extracellular region
Extracellular matrix
Cell
Biological process
Molecular function
Cellular component
Continued on next page

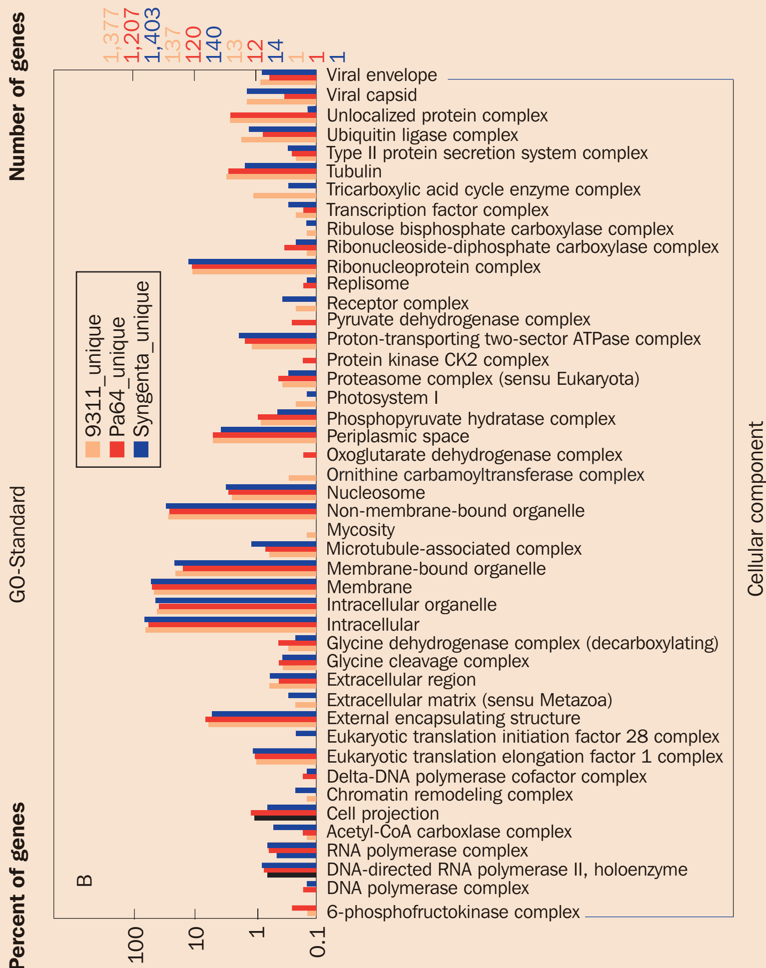
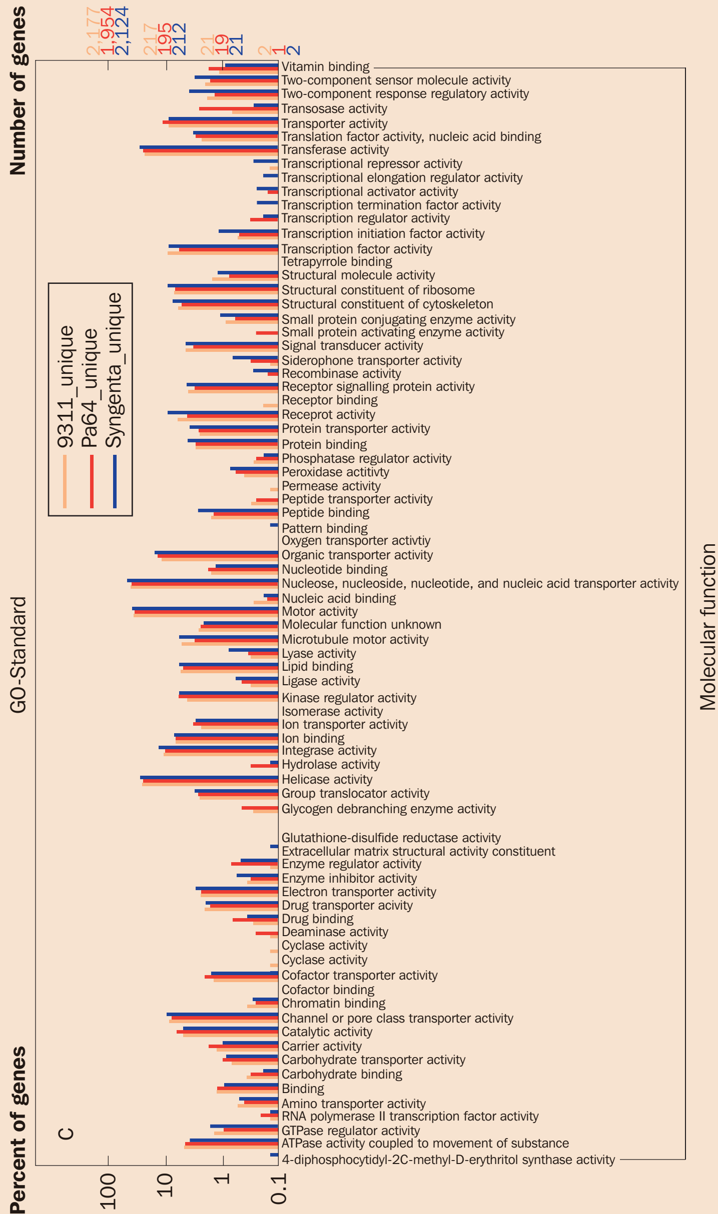


Figure 3 continued.

Continued on next page

Figure 3 continued.



Continued on next page

Figure 3 continued.

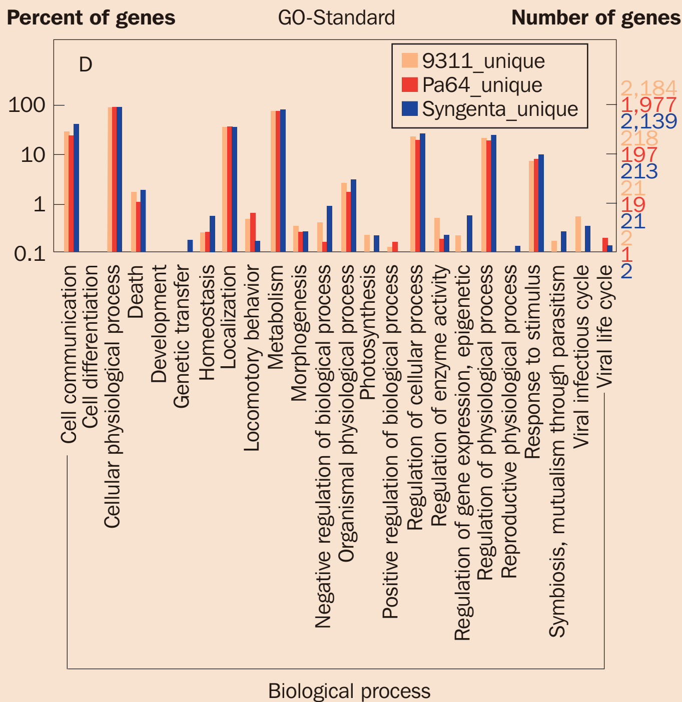


Fig. 3. GO classifications of the genes that are unique to 93-11, PA64S, and Nipponbare. The first panel shows a higher level classification (A) and the other three panels are breakdowns in three subcategories: cellular components (B), molecular function (C), and biological process (D).

and small noncoding RNAs. Special experimental designs are needed to look into these subtle yet important details. We are in the process of acquiring more data and attempting to answer these interesting questions.

Obviously, we are entering a critical phase in searching for a mechanistic explanation of plant hybrid vigor in molecular detail. For this reason, we should not jump too fast on the bandwagon of massive microarray studies, but step back for a moment to think harder and take a deep breath before we go too far into the experimental details. First, tissues from each related cultivar, the parental and the hybrid, should be collected at high density (or at short timing intervals) to assess gene expression and to determine what developmental stages in terms of both timing and biological traits are actually comparable for an analysis because field observations have shown that these related rice cultivars often have different growth potential, appearance, and maturation timing. Second, hybrid vigor may be a collective effect of many other cellular

Table 2. Microarray studies on differential gene expression in the developing leaves of LYP9 and its parental lines.

	93-11 paternal parent		LYP9 hybrid		PA64S maternal parent	
	Expressed[b]	D-regulated[c]	Expressed	D-regulated	Expressed	D-regulated
Flag leaf (filling)[a]	16,867	306	15,888	416	15,694	433
Flag leaf (flowering)	24,603	1,066	19,338	706	19,621	757
Flag leaf (heading)	32,823	661	17,433	580	22,799	971
Leaf (booting)	26,522	600	20,363	400	29,336	665
Leaf (tilling)	23,225	895	16,213	682	17,235	625
Leaf (seedling)	25,498	1,157	17,673	1,100	23,697	1,482

[a]Leaf samples collected at different developing stages are indicated in the parentheses. [b]Number of genes that were found expressed in the specified tissue samples. [c]Number of genes that were found differentially regulated in the specified tissues.

mechanisms in addition to gene expression and its regulation. For instance, we have observed, in our proteomics studies, that some of the differentially expressed gene products (proteins) in the same tissue samples from different cultivars detected by 2D gel electrophoresis actually resulted from different posttranslational modifications of the same gene product (Zhao et al 2005). Some of them were clearly alternatively spliced gene variants when highly expressed and they may not be relevant to hybrid vigor, unless proven functionally. Third, allelic differences between maternal and paternal genomes may also play a role in the manifestation of hybrid vigor, especially in the regulatory sequences that are not easily detectable unless large-scale sequencing is undertaken in a massive discovery process for all differentially expressed genes once proven functionally. Fourth, in order to approve that a set of genes or QTLs is related to hybrid vigor, we need to create intermediate breeding and transgenic lines for validating results from molecular studies and predictions. Our SRGP will acquire

more data on this front in building a long-term research platform for such studies. Many other cellular mechanisms regulate genes and their expression, such as chromosome methylation, histone acetylation, and RNA-based regulatory mechanisms (e.g., small and microRNAs; Sunkar et al 2005, Lu et al 2005). Nevertheless, complete sequences of the two parental cultivars from a rice hybrid as well as differential gene expression among them at different developmental stages in different tissues certainly provide powerful yet basic information for future genetic and molecular studies.

References

Bao J, Lee S, Chen C, Zhang X, Zhang Y, Liu S, Clark T, Wang J, Cao M, Yang H, Wang S, Yu J. 2005. Serial analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. Plant Physiol. 138:1216-1231.

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

Blakesley RW, Hansen F, Mullikin J, Thomas P, McDowell J, Maskeri B, Young A, Benjamin B, Brooks S, Coleman B, et al. 2004. An intermediate grade of finished genomic sequence suitable for comparative analyses. Genome Res. 14:2235-2244.

Goff SA, Ricke D, Lan T, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296:92-100.

IRGSP (International Rice Genome Sequencing Project). 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Kikuchi S, Satoh K, Nagata T, Kawagashia N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379.

Li R, Ye J, Li S, Wang J, Han Y, Ye C, Wang J, Yang H, Yu J, Wong G, Wang J. 2005. ReAS: recovery of ancestral sequences for transposable elements from the unassembled reads of a whole genome shotgun. PLoS Computational Biology 1(4):e43.

Lu C, Tej S, Lou S, Haudenschild C, Meyers B, Green P. 2005. Elucidation of the small RNA component of the transcriptome. Science 309:1567-1569.

Lu C, Zhou J. 2000. Breeding and utilization of two-line interspecific hybrid rice LYP9. Hybrid Rice 15:4-5. (In Chinese.)

Nordborg M, Hu T, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. 3(7):e196.

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

Sunkar R, Girke T, Jain P, Zhu J. 2005. Cloning and characterization of microRNAs from rice. Plant Cell 17:1397-1411.

Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, Tong W, Wang J, Wang J, Yu J, Yang H, Zhu L. 2004. A comparison of rice chloroplast genomes. Plant Physiol. 135:412-420.

Tian X, Hu S, Yu J. 2006. The rice (*Oryza sativa* L.) mitochondrial genomes and their variations. Plant Physiol. 140:401-410.

Wang J, Li S, Zhang Y, Zheng H, Xu Z, Ye J, Yu J, Wong G. 2003. Vertebrate gene predictions and the problem of large genes. Nature Rev. Genet. 4:741-749.

Yu J, Hu S, Wang J, Wong G, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79-92.
Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, et al. 2005. The genomes of *Oryza sativa*: a history of duplications. PLoS Biol. 3:e38.
Zhao C, Wang J, Cao M, Zhao K, Shao J, Lei T, Yin J, Hill G, Xu N, Liu S. 2005. Proteomic changes in rice leaves during development of field-grown rice plants. Proteomics 5:961-972.
Zhao W, Wang J, He X, Huang X, Jiao Y, Dai M, Wei S, Fu J, Chen Y, Ren X, et al. 2004. BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. Nucl. Acids Res. 32:D377-382.
Zhong L, Zhang K, Huang X, Ni P, Han Y, Wang K, Wang J, Li S. 2003. A statistical approach designed for finding mathematically defined repeats in shotgun data and determining the length distribution of clone-inserts. Genomics Proteomics Bioinformatics 1:43-51.
Zhou Y, Tang J, Walker MG, Zhang X, Wang J, Hu S, Xu H, Deng Y, Dong J, Ye L, et al. 2003. Gene identification and expression analysis of 86,136 expressed sequence tags (EST) from the rice genome. Genomics Proteomics Bioinformatics 1:26-42.

Notes

Authors' address: Beijing Genomics Institute, Chinese Academy of Sciences, Beijing, China, 100301, e-mail:junyu@genomics.org.cn.
Acknowledgments: This work has been supported by major Chinese funding agencies, including the Chinese Academy of Sciences, the State Commission of Development and Reform, the Ministry of Science and Technology, the National Science Foundation, the Municipal Government of Hangzhou, and the Provincial Government of Zhejiang. We are grateful to many past and present faculty and staff members of the Beijing Genomics Institute who have contributed to this significant project.

Developmental biology and gene regulation

T-DNA tagging for developmental biology

G. An, D.-H. Jeong, S. An, and S. Park

We have generated 47,932 T-DNA tag lines in japonica rice using activation tagging vectors that contain tetramerized 35S enhancer sequences. To facilitate use of those lines, we isolated the genomic sequences flanking the inserted T-DNA via inverse polymerase chain reaction. For most of the lines, we performed four sets of amplifications using two different restriction enzymes toward both directions. In analyzing 41,234 lines, we obtained 27,621 flanking sequence tags (FSTs), among which 12,505 were integrated into genic regions and 15,116 into intergenic regions. Mapping of the FSTs on chromosomes revealed that T-DNA integration frequency was generally proportional to chromosome size. However, T-DNA insertions were nonuniformly distributed on each chromosome, that is, higher at the distal ends and lower in regions close to the centromeres. In addition, several regions showed extreme peaks and valleys of insertion frequency, suggesting hot and cold spots for T-DNA integration. The density of insertion events was somewhat correlated with the expressed, rather than the predicted, gene density along each chromosome. Analyses of expression patterns near the inserted enhancer showed that at least half the test lines displayed greater expression of the tagged genes. Although in most of the increased lines expression patterns after activation were similar to those in the wild type, thereby maintaining the endogenous patterns, the remaining lines showed changes in expression in the activation tagged lines. In this case, ectopic expression was most frequently observed in mature leaves. Currently, the database can be searched with the gene locus number or location on the chromosome at www.postech.ac.kr/life/pfg/risd. Upon request, seeds of the T_1 or T_2 plants will be provided to the scientific community.

Keywords: activation tagging, database, flanking sequences, insertional mutation, T-DNA

Since the release of the rice genome sequence, the most significant challenge has been the large-scale identification of gene functions (Feng et al 2002, Goff et al 2002, Sasaki et al 2002, Yu et al 2002). Recently, about 370 Mb, or >97% of the genome, have been assembled as reference molecules with the release of the 3.0 pseudomolecules

by the International Rice Genome Sequencing Project (IRGSP) (Sasaki et al 2005). Using these nonoverlapping genome sequences as templates for annotation, 57,888 genes have now been predicted by the annotation team of The Institute for Genomic Research (TIGR). In addition, the rice cDNA project has generated sequence data for 175,642 full-length cDNAs clustered into 28,469 nonredundant clones (Kikuchi et al 2003). These data, available through the Knowledge-based Oryza Molecular Biological Encyclopedia (KOME), facilitate gene prediction to make this information more valuable. As expected, these recent successes are also accelerating the need for functional genomics in this genus.

Various methods have been applied to generate loss-of-function mutations, including the use of ethyl methanesulfonate, fast-neutron treatment, antisense and RNA interference technology, and insertion mutations by a transposable element or T-DNA (Hirochika et al 2004, Jeon et al 2000, Kolesnik et al 2004, Miki and Shimamoto 2004, Miyao et al 2003). One limitation to these approaches is that they rarely uncover function when the genes are either redundant or essential for early embryo or gametophyte development. Functional redundancy in most eukaryotic genes provides a significant obstacle to the assignment of gene function (Normandy and Bartel 1999). Among the numerous approaches that have emerged to overcome these problems, the enhanced expression of genes that provide gain-of-function phenotypes has proven a productive identification strategy.

The first direct method for performing gain-of-function genetics in plants used the enhancer element from the cauliflower mosaic virus (CaMV) 35S gene (Odell et al 1985). T-DNA vectors containing four copies of this element were used successfully for generating activation tagging lines and mediating transcriptional activation of nearby genes in *Arabidopsis* (Weigel et al 2000). For example, the mechanism for auxin biosynthesis, consisting of multiple pathways, was elucidated by this activation tagging approach (Zhao et al 2001). This method has also been widely used in mutant screening to uncover enhancers or suppressors of given mutations (Li et al 2001, 2002). Although activation tagging has predominantly been applied to gene mining in *Arabidopsis*, this technology is now being deployed in diverse plant species, such as petunia (Zubko et al 2002), tomato (Mathews et al 2003), poplar (Busov et al 2003), Madagascar periwinkle (van der Fits and Memelink 2000), and rice (Jeong et al 2002).

The CaMV 35S enhancers, used for most activation tagging, function both upstream and downstream of a gene, in either orientation, and at a considerable distance from the coding regions. Furthermore, in some activation tagging lines of *Arabidopsis* or rice, those enhancers cause greater endogenous expression rather than ectopic expression (Jeong et al 2002, Neff et al 1999, Weigel et al 2000). In these cases, identified phenotypes are more likely to reflect the endogenous function of a given tagged gene. Researchers have also developed an alternative gain-of-function approach, which relies on either a tissue-specific promoter to mis-express a gene in a particular tissue or else an inducible promoter to overexpress a gene at a specific time and under certain conditions (Matsuhara et al 2000, Zuo et al 2002).

Despite the usefulness of a phenotype-driven genetic approach, it is somewhat inconvenient for high-throughput screening of rice mutations. First, mutant screening of plants requires more effort because of their larger size and longer life cycle. Second, the phenotypic alterations observed in a T-DNA tagged line are not necessarily due to insertional mutation events but, instead, to the transposition of endogenous mobile elements such as *Tos17* (Miyao et al 2003). Finally, tissue culture often causes point mutations as well as small deletions or insertions.

These difficulties can be overcome through reverse genetics, in which a database for T-DNA insertion sites is first established and then used for functional analysis of the T-DNA-tagged genes. Although large-scale application of this strategy requires considerable effort (Parinov and Sundaresan 2000), this database can be easily shared with other scientists, facilitating the distribution of mutant materials and analysis of gene functioning (An et al 2005). In *Arabidopsis*, more than 88,000 independent T-DNA insertion site sequences have been isolated from 127,706 T_1 plants, resulting in the identification of mutations in more than 21,700 of the 29,454 predicted genes (Alonso et al 2003). Here, we report the generation of nearly 50,000 activation tagging lines and 27,621 insertion site sequences in rice.

Results

Generation of activation tagging lines and isolation of tag end sequences

We previously reported the generation of 13,450 activation tagging lines of japonica rice using binary vector pGA2715 (Jeong et al 2002). We have now developed another activation tagging vector, pGA2772, which is a modified version of pGA2715 containing the pUC18 vector backbone. This new vector is useful for retrieving T-DNA flanking regions if routine PCR methods fail to identify them. Using the *Agrobacterium*-mediated transformation method, we have established an additional 23,009 lines by pGA2715, plus 11,473 lines by pGA2772. Altogether, we have now generated 47,932 activation tag lines in rice.

To facilitate the use of these tagged lines, we isolated genomic sequences flanking the inserted T-DNA via inverse PCR (An et al 2003). Cutting the genomic DNA with restriction enzymes and using self-circularization yielded a molecule that could be PCR-amplified with primers located in the T-DNA. We designed the primer sets to amplify the genomic sequences flanking either the left border or right border of T-DNA. For most of the lines, we performed four sets of amplifications using two different restriction enzymes in both directions.

By analyzing 31,100 lines of pGA2715 and 10,134 lines of pGA2772, we obtained 22,114 and 5,507 flanking sequence tags (FSTs), respectively. The isolated FSTs were analyzed by the BlastN homology search program, using the rice genome sequence database version 3.0 in TIGR. Of the total of 27,621 insertions, 12,505 (45.3%) of the T-DNAs were integrated into genic regions and 15,116 (54.7%) were integrated into intergenic regions (Table 1). We considered the 300-bp flanking sequences outside the start ATG and stop codon to be untranslated regions of the genic

Table 1. Distribution of T-DNA insertions in genic and intergenic regions.

Location of T-DNA insertions	pGA2715	pGA2772	Total	%
Genic	10,017	2,488	12,505	45.3
5′UTR (300 bp upstream)	1,749	459	2,208	8.0
Coding exon	3,248	820	4,068	14.7
Intron	3,790	884	4,674	16.9
3′UTR (300 bp downstream)	1,230	325	1,555	5.6
Intergenic	12,097	3,019	15,116	54.7
Total	22,114	5,507	27,621	100.0

Table 2. Distribution of predicted genes, expressed genes, and T-DNA insertions in rice chromosomes.

Chromosome	Size		Predicted genes		Expressed genes		T-DNA insertions	
	Mb	%	n	%	n	%	n	%
1	43.2	11.7	6,905	11.9	1,853	12.2	4,126	14.9
2	35.9	9.7	5,422	9.4	1,774	11.7	3,177	11.5
3	36.3	9.8	5,986	10.3	2,067	13.6	3,642	13.2
4	35.0	9.4	5,534	9.6	1,370	9.0	2,505	9.1
5	29.7	8.0	4,636	8.0	1,319	8.7	2,065	7.5
6	31.2	8.4	4,837	8.4	1,313	8.7	2,137	7.7
7	29.7	8.0	4,635	8.0	1,147	7.6	2,015	7.3
8	28.3	7.6	4,326	7.5	984	6.5	1,721	6.2
9	22.7	6.1	3,409	5.9	790	5.2	1,544	5.6
10	22.7	6.1	3,743	6.5	816	5.4	1,565	5.7
11	28.4	7.7	4,286	7.4	876	5.8	1,562	5.7
12	27.5	7.4	4,169	7.2	857	5.7	1,562	5.7
Total	370.6	100.0	57,888	100.0	15,166	100.0	27,621	100.0

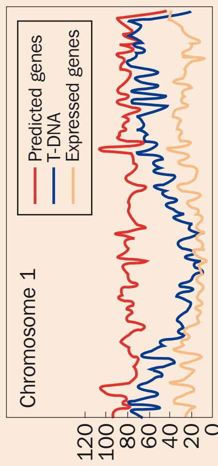
region (An et al 2003, Szabados et al 2002). Our results are similar to those reported previously for nonactivation T-DNA tagging lines (An et al 2003, Chen et al 2003, Sallaud et al 2004).

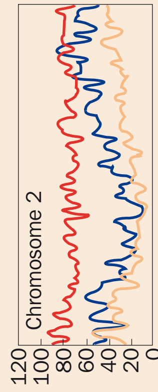
Distribution of T-DNA insertions

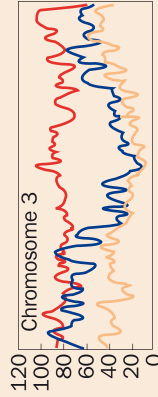
Mapping of the 27,621 FSTs revealed that T-DNA integration frequency was generally proportionate to chromosome size (Table 2). Insertion was most frequent on the largest one, chromosome 1, which also had the greatest number of predicted genes. Chromosomes 9 and 10 were the smallest and carried the fewest T-DNA insertions.

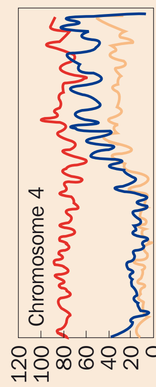
We found nonuniform distribution of T-DNA insertions when their numbers were plotted per 500-kb interval over the length of each of the 12 chromosomes (Fig. 1).

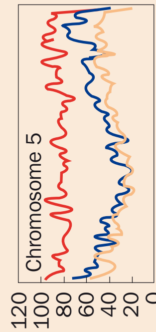
No. of events/500 kb

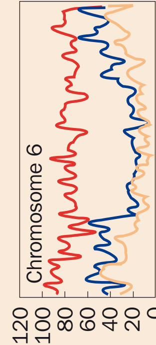












Continued on next page

Figure 1 continued

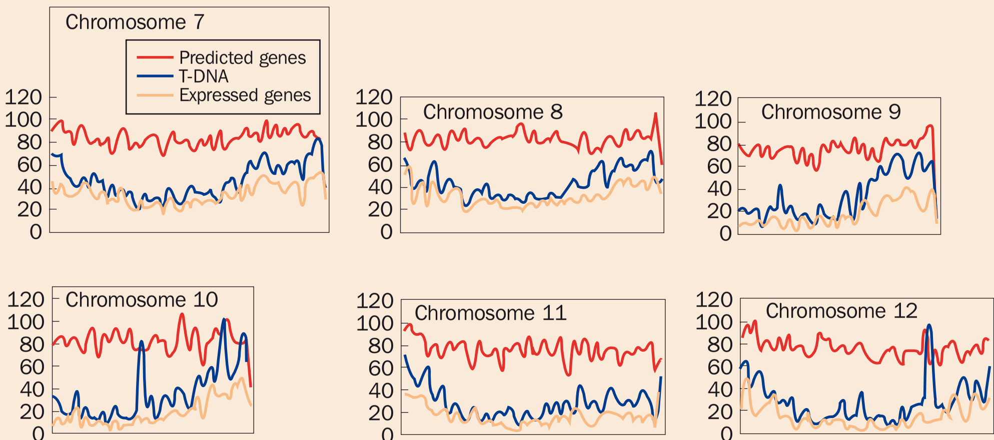


Fig. 1. Distribution of T-DNA insertions, expressed genes, and predicted genes along rice chromosomes, divided into 500-kb windows.

Table 3. Genes tagged by T-DNA or *Tos17*. A total of 27,621 T-DNA insertions and 14,681 *Tos17* insertions were analyzed. The number of predicted genes and expressed genes in the rice genome was 57,888 and 15,166, respectively. T-DNA insertions into intergenic regions were classified by the intergenic regions of predicted or expressed genes adjacent to the left border containing the 35S enhancer. In the case of *Tos17* insertions into intergenic regions, the genes adjacent to the 3′-LTR of the insertion element were presented.

	Genic regions				Intergenic regions			
	Predicted genes		Expressed genes		Predicted genes		Expressed genes	
	No.	%	No.	%	No.	%	No.	%
T-DNA	9,911	17.1	4,216	27.8	11,309	19.5	4,403	29
Tos17	3,380	5.8	1,408	9.3	1,783	3.1	705	4.5
Total	13,291	20.8	5,072	33.4	12,520	21.6	4,832	31.9

Insertion frequency was higher at the distal ends and lower in regions close to the centromeres. In addition, several regions showed extreme peaks and valleys of frequency, suggesting hot and cold spots for T-DNA integration along each chromosome. These results are largely similar to those previously reported with *Arabidopsis* and rice (Alonso et al 2003, An et al 2003, Chen et al 2003, Sallaud et al 2004).

To examine whether this bias was due to unequal distribution of genic regions, we downloaded 57,888 predicted genes and 15,166 expressed genes (with EST and/or full-length cDNA evidence from predicted genes) from the TIGR rice genome annotation database version 3.0. The density of T-DNA insertion events was somewhat correlated to the expressed gene density rather than the predicted gene density along each chromosome (Fig. 1).

Again, our analysis of 27,621 FSTs revealed that about 45% of the T-DNAs were inserted into the genic regions and about 55% into the intergenic regions (Table 1). Among the 12,505 T-DNA located within the former, some were inserted into the same genes. Therefore, removing the multiple alleles within the same gene resulted in the identification of T-DNA knockouts in 9,911 (17.1%) predicted genes. When we examined the expressed genes, 4,216 (27.8%) had T-DNA inserts (Table 3). Analysis of T-DNA insertions in the intergenic regions showed that 11,309 of the independent predicted genes had the insertion in either the 5′ or 3′ regions. Likewise, 4,403 of the expressed genes contained the T-DNA in the flanking regions.

We also analyzed the *Tos17*-tagged genes, using publicly available information (http://tos.nias.affrc.go.jp/). Examination of 14,681 *Tos17* insertion sequences showed that 3,380 predicted genes and 1,408 expressed genes were tagged by the element. Among these *Tos17*-tagged genes, 1,251 predicted genes and 552 expressed genes were also tagged by T-DNA. Consequently, 12,040 predicted genes and 5,072

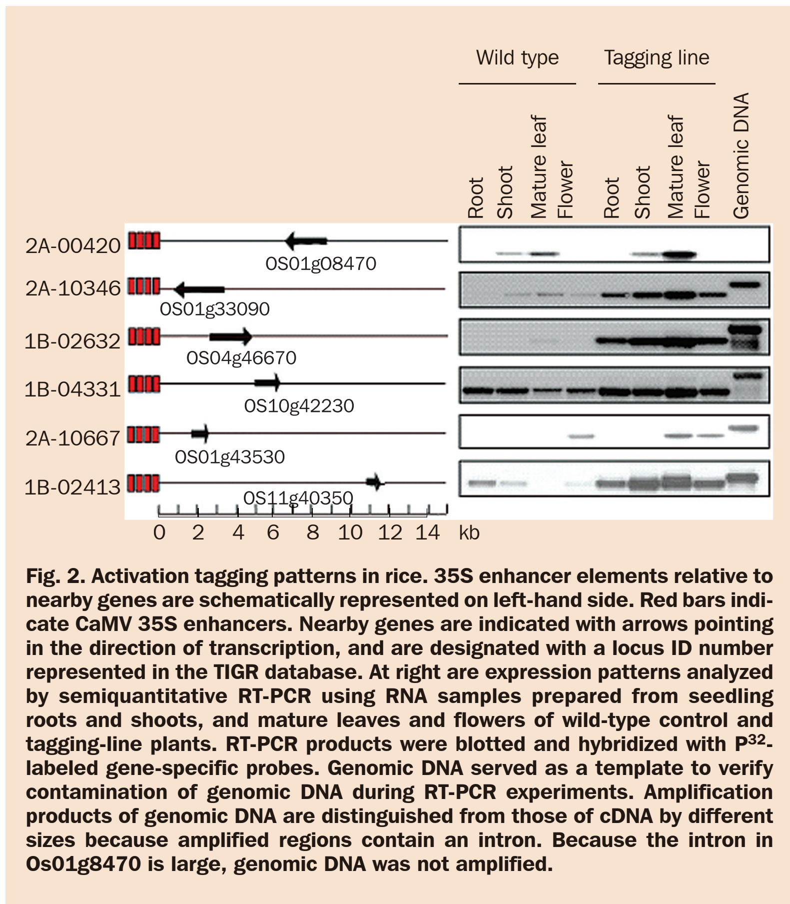


Fig. 2. Activation tagging patterns in rice. 35S enhancer elements relative to nearby genes are schematically represented on left-hand side. Red bars indicate CaMV 35S enhancers. Nearby genes are indicated with arrows pointing in the direction of transcription, and are designated with a locus ID number represented in the TIGR database. At right are expression patterns analyzed by semiquantitative RT-PCR using RNA samples prepared from seedling roots and shoots, and mature leaves and flowers of wild-type control and tagging-line plants. RT-PCR products were blotted and hybridized with P^{32}-labeled gene-specific probes. Genomic DNA served as a template to verify contamination of genomic DNA during RT-PCR experiments. Amplification products of genomic DNA are distinguished from those of cDNA by different sizes because amplified regions contain an intron. Because the intron in Os01g8470 is large, genomic DNA was not amplified.

expressed genes were tagged by T-DNA or *Tos17*. This result demonstrates that the chance of finding an insertional mutation in a given gene is higher from the T-DNA insertional database established in this laboratory.

To examine a genome-wide correlation between distribution of T-DNA insertion and genic region, we plotted the numbers of predicted genes in the 500-kb window against the number of tagged genes in that same window (Fig. 2). The correlation coefficiency (r) was 0.34, indicating little relationship between insert distribution and the predicted gene. A similarly low value was obtained between the distribution of intergenic T-DNA insertions and predicted genes. In contrast, we observed a high correlation between tagged genes and expressed genes, as well as between intergenic tags and expressed genes. Therefore, this genome-wide comparison strongly suggests that T-DNA insertion prefers the region where expressed genes are clustered.

We also analyzed *Tos17* distribution by the same method and found a low correlation efficiency between *Tos17* and predicted genes. This value was higher with expressed genes, although not as high as that obtained with T-DNA, suggesting that T-DNA insertion is more randomly distributed in the rice genome.

Distribution of T-DNA insertions into intergenic regions

Because the tagging vectors pGA2715 and pGA2772 contain multimerized 35S enhancers in the T-DNA, tagging lines transformed with these vectors can be used for both insertional tagging and activation tagging (Jeong et al 2002). To investigate how many such lines are candidates for activation of nearby genes, we examined the distributions of intergenic lengths for 57,888 predicted genes. Distribution here displayed an inclined pattern toward shorter lengths. Among the predicted genes, 32,381 (55.9%) intergenic regions were <3.0 kb long, which was the average length of those regions. We also examined the 15,116 T-DNA insertions located in the intergenic regions and found that their distribution displayed a bell-shape pattern with an average length of 5.5 kb in those tagged regions. The probability of finding an insertion in the intergenic regions was about 50% when the length was between 3 and 4 kb. This frequency reached 80% when the length was increased.

We also analyzed the distribution of T-DNA insertions from the start ATG and stop codons of the next genes. Here, 14,548 sites were located within 5 kb upstream from the start ATG, whereas 12,221 sites were found within 5 kb downstream from the stop codons. The results showed that regions near the start ATG and stop codon had a higher frequency of insertions than those far from the coding sequences.

Analysis of activation tagging patterns

To monitor how activation tagging might perturb gene expression, we randomly selected insertion lines with T-DNA in their intergenic regions. Expression patterns of nearby genes, closest to the tetramerized 35S enhancer sequences at the T-DNA left border, were studied via semiquantitative RT-PCR using gene-specific primers. Levels of expression were measured in the roots and shoots of seedlings, as well as in the mature leaves and panicles from tagged lines and wild-type plants. At least half of the test lines (52.7%, 59/112) displayed greater expression of the tagged genes (data not shown). In most of the increased lines (69.5%, 41/59), patterns after activation were similar to those in the wild type, maintaining their endogenous expression patterns (Fig. 2, lines 2A-00420, 2A-10364, and 1B-02632). In the remaining lines (30.5%, 18/59), the patterns changed in the activation-tagged lines, with ectopic expression being most frequently observed in the mature leaves (Fig. 2, lines 1B-04331, 2A-10667, and 1B-02413).

No good relationship was found between frequency of activation and distance from the 35S enhancers to the gene (data not shown). Similarly, we observed no correlation between degree of activation and distance (Fig. 2). For example, strong enhancement was noted in line 1B-02413, where the 35S enhancers were located 10.7 kb upstream from the start codon of the Os11g40350 gene. Enhancement was observed both upstream and downstream of the tagged genes (Fig. 2).

Discussion

Generation of activation tagging lines to provide a wide variety of mutants

Nearly 2,000 traits, including both single Mendelian loci/genes and quantitative trait loci (QTLs), have been identified in rice (Kurata et al 2005). However, the number of mutants is much smaller than the number of predicted genes found during recent genome sequencing of that species. This might be mainly due to redundancy because most of its genes are members of one family (Goff et al 2002, Sasaki et al 2002, Yu et al 2002). Therefore, classical loss-of-function mutants have limitations when one attempts to elucidate gene function. To provide a wide variety of mutants, we have generated binary vectors that contain multimerized 35S enhancer elements immediately next to the left border. Similar vectors have been successfully used to produce activation tagging populations in *Arabidopsis* and other dicot species (Busov et al 2003, Li et al 2001, 2002, Mathews et al 2003, van der Fits and Memelink 2000, Zubko et al 2002). In those tagging lines, expression of the gene near the enhancer elements is enhanced, causing dominant gain-of-function phenotypes. Thus, activation tagging mutagenesis presents a phenotypic spectrum that is different from phenotypes generated by loss-of-function mutations. In this study, we created nearly 50,000 activation tagging lines in japonica rice. Because each line contains an average of 1.4 insertion loci (Jeon et al 2000), approximately 70,000 T-DNA inserts were made. This population should be a valuable resource to researchers in the plant community for functional analysis of rice genes.

Establishing a database of T-DNA insertion sites for reverse genetics approaches

To efficiently use the mutant population and facilitate the sharing of resources within the scientific community, we determined the genomic sequences flanking the T-DNA insertions, using inverse PCR because that method provides an average of one band after amplification. This is an important factor because this approach does not require gel-separation of PCR bands followed by elution and purification. We directly sequenced the PCR product, thereby analyzing a large number of samples with only limited resources.

From the analysis of 41,234 lines, we obtained 27,621 FSTs. Considering that each tagging line carries an average of 1.4 T-DNA insertion loci, up to 57,700 FSTs might have been retrieved from the analysis. This indicates that our efficiency was about 48%. One of the difficulties in isolating FSTs is a high GC content at the tag sites, which inhibits PCR amplification. Another problem is repetitive sequences that lack the enzyme sites used for iPCR analyses. Currently, we are improving the efficiency rate by employing a high GC buffer. We also plan to determine the flanking sequences of the pGA2772-tagged lines by the plasmid rescue method.

We have now generated a database with FSTs obtained from the activation tagging lines. It can be searched with the gene locus number or location on the chromosome at www.postech.ac.kr/life/pfg/risd. We are in the process of improving the

search engine so that the database can be searched with DNA sequences or key words. Upon request, 15 seeds of the T_1 plants can be made available when more than 100 seeds are present in the seed stock. If that number is less than 100, we provide the seeds after their multiplication.

More than a quarter of the expressed genes are tagged

Approximately 45% of the FSTs are present in the genic region. Our analysis showed that 17.1% of the predicted genes had at least one T-DNA insertion there. In contrast, 27.8% of the expressed genes were tagged by T-DNA. Similar, higher efficiency in the expressed genes was obtained with the intergenic insertions. Therefore, it seems that T-DNA prefers highly expressed genes compared with those that are poorly expressed. Alternatively, the predicted gene number may be overestimated. Among the 57,888 genes predicted by the TIGR rice genome database, 14,196 genes are transposable elements, which are considered transcriptionally silent. These elements are usually clustered near the centromeres where T-DNA insertion frequency is low. However, even if those transposable elements are not considered, the frequency of FSTs in the predicted genes is still lower than that in the expressed genes. This suggests that the total number of functional genes in rice may be much smaller than the number of annotated genes.

Gene expression in mature leaves is more preferentially enhanced by activation tagging

Transcript levels for genes near the 35S enhancer were increased in about half of the activation tagging lines. In the remaining half, levels and patterns of expression were not changed significantly. In most of the increased lines, the expression patterns were conserved but overall enhancement was observed. In these cases, the genes were more preferentially expressed in mature leaves. In the remaining lines, expression patterns of the tagged genes were not mature-leaf-preferential, but became leaf-preferential after activation tagging. Therefore, it appears that the 35S enhancer elements increase the expression of nearby genes more preferentially in mature leaves, especially when the tagged gene is originally expressed preferentially in other organs. Because only half of our tagged genes were enhanced by the tagging vector, a silencing mechanism might inactivate the action of 35S enhancer elements. One possible mechanism is methylation, which is induced by multiple-copy T-DNA integration (Chalfun-Junior et al 2003).

References

Alonso JM, Stepanova A, Leisse T, Kim C, Chen H, Shinn P, Stevenson D, Zimmerman J, Barajas P, Cheuk R, et al. 2003 Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653-657.

An G, Lee S, Kim SH, Kim SR. 2005. Molecular genetics using T-DNA in rice. Plant Cell Physiol. 46:14-22.

An S, Park S, Jeong D, Lee D, Kang H, Yu J, Hur J, Kim S, Kim Y, Lee M, et al. 2003. Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiol. 133:2040-2047.

Busov VB, Meilan R, Pearce DW, Ma C, Rood SB, Strauss SH. 2003. Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature. Plant Physiol. 132:1283-1291.

Chalfun-Junior A, Mes JJ, Mlynarova L, Aarts MG, Angenent GC. 2003. Low frequency of T-DNA based activation tagging in *Arabidopsis* is correlated with methylation of CaMV 35S enhancer sequences. FEBS Lett. 555:459-463.

Chen S, Jin W, Wang M, Zhang F, Zhou J, Jia Q, Wu Y, Liu F, Wu P. 2003. Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J. 36:105-113.

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X, et al. 2002. Sequence and analysis of rice chromosome 4. Nature 420:316-320.

Goff S, Ricke D, Lan T, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100.

Hirochika H, Guiderdoni E, An G, Hsing YI, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang Q, Pereira A, Sundaresan V, Leung H. 2004. Rice mutant resources for gene discovery. Plant Mol. Biol. 54:325-334.

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.

Jeong DH, An S, Kang HG, Moon S, Han JJ, Park S, Lee HS, An K, An G. 2002. T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. 130:1636-1644.

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379.

Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG, Sundaresan V, Ramachandran S. 2004. Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. Plant J. 37:301-314.

Kurata N, Miyoshi K, Nonomura K, Yamazaki Y, Ito Y. 2005. Rice mutants and genes related to organ development, morphogenesis and physiological traits. Plant Cell Physiol. 46:48-62.

Li J, Lease KA, Tax FE, Walker JC. 2001. BRS1, a serine carboxypeptidase, regulates BRI1 signaling in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 98:5916-5921.

Li J, Wen J, Lease KA, Doke JT, Tax FE, Walker JC. 2002. BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell 110:213-222.

Mathews H, Clendennen SK, Caldwell CG, Liu XL, Connors K, Matheis N, Schuster DK, Menasco DJ, Wagoner W, Lightner J, Wagner DR. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell 15:1689-1703.

Matsuhara S, Jingu F, Takahashi T, Komeda Y. 2000. Heat-shock tagging: a simple method for expression and isolation of plant genome DNA flanked by T-DNA insertions. Plant J. 22:79-86.

Miki D, Shimamoto K. 2004. Simple RNAi vectors for stable and transient suppression of gene function in rice. Plant Cell Physiol. 45:490-495.

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H. 2003. Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771-1780.
Neff MM, Nguyen SM, Malancharuvil EJ, Fujioka S, Noguchi T, Seto H, Tsubuki M, Honda T, Takatsuto S, Yoshida S, Chory J. 1999. *BAS1*: a gene regulating brassinosteroid levels and light responsiveness in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 96:15316-15323.
Normandy J, Bartel B. 1999. Redundancy as a way of life: IAA metabolism. Curr. Opin. Plant Biol. 2:207-213.
Odell JT, Nagy F, Chua NH. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 313:810-812.
Parinov S, Sundaresan V. 2000. Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. Curr. Opin. Biotechnol. 11:157-161.
Sallaud C, Gay C, Larmande P, Bes M, Piffanelli P, Piegu B, Droc G, Regad F, Bourgeois E, Meynard D, Perin C, et al. 2004. High-throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. Plant J. 39:450-464.
Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, Nagamura Y, et al. 2002. The genome sequence and structure of rice chromosome 1. Nature 420:312-316.
Sasaki T, Matsumoto T, Antonio BA, Nagamura Y. 2005. From mapping to sequencing, post-sequencing and beyond. Plant Cell Physiol. 46:3-13.
Szabados L, Kovacs I, Oberschall A, Abraham E, Kerekes I, Zsigmond L, Nagy R, Alvarado M, Krasovskaja I, Gal M, et al. 2002. Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. Plant J. 32:233-242.
van der Fits L, Memelink J. 2000. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science 289:295-297.
Weigel D, Ahn J, Blázquez M, Borevitz J, Christensen S, Fankhauser C, Ferrándiz C, Kardailsky I, Malancharuvil E, Neff M, et al. 2000. Activation tagging in *Arabidopsis*. Plant Physiol. 122:1003-1013.
Yu J, Hu S, Wang J, Wong G, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79-92.
Zhao Y, Christensen SK, Fankhauser C, Cashman JR, Cohen JD, Weigel D, Chory J. 2001. A role for flavin monooxygenase-like enzymes in auxin biosynthesis. Science 291:306-309.
Zubko E, Adams CJ, Machaekova I, Malbeck J, Scollan C, Meyer P. 2002. Activation tagging identifies a gene from *Petunia hybrida* responsible for the production of active cytokinins in plants. Plant J. 29:797-808.
Zuo J, Niu QW, Frugis G, Chua NH. 2002. The WUSCHEL gene promotes vegetative-to-embryonic transition in *Arabidopsis*. Plant J. 30:349-359.

Notes

Authors' address: Department of Life Science and National Research Laboratory of Plant Functional Genomics, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Republic of Korea.
Acknowledgments: This work was funded in part by grants from the Crop Functional Genomic Center, the 21st Century Frontier Program (CG-1111); from the Biogreen 21 Program, Rural Development Administration; and from POSCO.

Novel insights into the genomics of rice root adaptive development

C. Périn, J. Rebouillat, A.M.C. Brasileiro, A. Diévart, P. Gantet, J.C. Breitler, A.A.T. Johnson, B. Courtois, N. Ahmadi, M. de Raissac, D. Luquet, M. Conte, D. This, P.K. Pati, Q.H. Le, D. Meynard, J.L. Verdeil, and E. Guiderdoni

Deciphering the genetic and molecular mechanisms controlling the development of the root system and its adaptive plasticity under adverse environments is of primary importance for the sustainable establishment of the rice crop. Rice displays a complex root structure comprising several root types mostly of postembryonic origin. The large natural variation in root architecture among cultivars reflects their adaptation to contrasting agro-environmental conditions. This article reviews the current knowledge on the organization and anatomy of the various types of roots of the fibrous root system of rice, the diversity and genetic basis of natural variation of root system architecture and performance, and the molecular mechanisms underlying constitutive and adaptive root development. This paper also throws light on how the integrated approach of new tools in high-resolution microscopy imaging, expression profiling, mutant screening, and reverse genetics could facilitate the rapid discovery and analysis of the key genes and regulatory networks involved in root architectural traits affecting plant performance under field conditions.

Keywords: Functional genomics, molecular genetics, rice, root development

The root system displays several essential functions for whole-plant development and growth, ranging from anchoring and water and nutrient uptake to hormone synthesis. There is an established correlation between root system size and resistance to water stresses (e.g., Price et al 2002a, Tuberosa et al 2002). Plants with a larger root system have been shown to have an increased ability to compete for nutrients and survive nutritional deficits (Liao et al 2001). Moreover, roots show adaptive morphogenesis, hydrotropism, and accelerated elongation in response to moisture gradient and water deficit (Takahashi 1997, 2003), and this is more evident in the case of upland cultivars (Yang et al 2003), and the increased lateral root elongation to target nutrients (nitrate or phosphate) (e.g., Zhang and Forde 1998, Hodge 2004). In rice, earlier work suggests that, under a broad range of conditions, a deep and thick root system is critical for maintaining yield under stress and this indicates that plant root architecture should be

a major determinant of drought tolerance under different field conditions (Mambani and Lal 1983). However, the molecular control of constitutive and adaptive pathways of root morphogenesis remains poorly understood in cereals even though some studies have begun to elucidate the genetic control of root system architecture in maize and rice (reviewed in Hochholdinger et al 2004). Most studies have focused on *Arabidopsis thaliana,* a model dicotyledonous species exhibiting a single primary root with a simple and detailed anatomy (reviewed in Scheres et al 2002). Molecular control of the adaptive response of root development, so-called plasticity, has also been investigated more recently, for example, in response to phosphate availability or water moisture gradients (see López Bucio et al 2005 and Malamy 2005 for recent reviews).

Interest has increased in understanding the molecular control of root development in rice because of the importance of root system establishment and plasticity in plant performance in the field, notably under nonbiotic and weed competition constraints. We focus hereafter on the fundamentals of rice shoot:root developmental synchronism, the morphology and anatomy of the various root types building the complex fibrous root system of rice, and some striking differences with the *Arabidopsis* system. We also present current knowledge on genetic and molecular controls of constitutive and adaptive root development in rice. Recent progress in cell-imaging techniques, isolation and functional analysis of developmental genes, analysis and manipulation of gene expression at a cell-specific resolution, all of which will facilitate the investigation of root developmental processes in rice, is also highlighted.

Rice root system organization, morphology, and anatomy

Root development follows a typical monocotyledonous pattern that can be resolved into three successive stages that overlap in both duration and ontogeny:

- *Seminal root development*. During grain germination, a unique seminal root (sr) originating from the root pole of the embryonic axis emerges and elongates.
- *Nodal (crown) root development*. As the shoot develops, nodal roots—or crown roots (cr)—are formed by the differentiation of stem parenchyma cells. This shoot ontogeny is a specific feature of monocotyledonous plants and is often known as adventitious roots. The roots emerge and elongate at successive node ranges throughout the course of phytomer (a unit of stem, leaf, and axial bud) development.
- *Branching pattern*. When root length exceeds a certain size, the branching process starts by initiation, emergence, and growth of lateral roots (lr, also called secondary roots) from the root pericycle and endodermis. Lateral roots of a higher order are also produced. It is reported that rice generally exhibits roots up to the tertiary order, but superficial roots exhibiting up to fifth-order branches have been described (Morita and Yamazaki 1993).

Because of a high synchronism between shoot and root development in rice, one must assess performance of the root system with respect to a reference developmental stage of the shoot, which can be easily measured by the plastochron (time interval

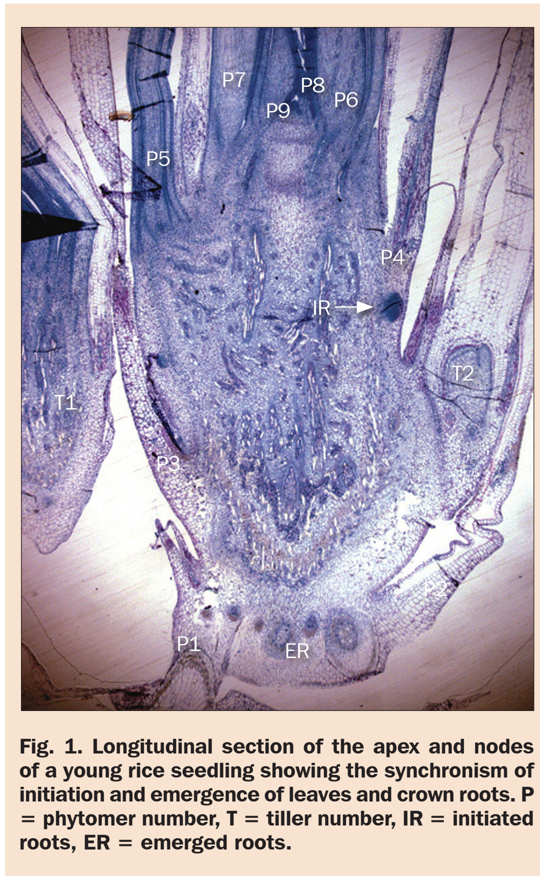


Fig. 1. Longitudinal section of the apex and nodes of a young rice seedling showing the synchronism of initiation and emergence of leaves and crown roots. P = phytomer number, T = tiller number, IR = initiated roots, ER = emerged roots.

between initiation of two successive leaves on a culm) index. This process has been quite extensively described in the literature and strong correlations between the respective rates of root and leaf production have been highlighted (Yoshida 1981, Nemoto et al 1995). This can be illustrated by a longitudinal section of an IR64 seedling that shows the synchronous events of initiation at the ninth phytomer, root initiation at the fourth phytomer, and root emergence at the first phytomer (Fig. 1). Since plastochron and phyllochron [(thermal) time interval between the emergence of two successive leaves on a culm] are equal in rice (Nemoto et al 1995), the Haun index, defined as the decimal number of leaves emerged from the main stem, provides a relevant indicator of the number of phytomers that bear crown roots and allows us to correlate the observed number of crown roots with a number of contributing nodes.

Rice root morphology

The fibrous mature root system is made up of at least four distinct root types that originate from embryonic and postembryonic development. The single sr that is vis-

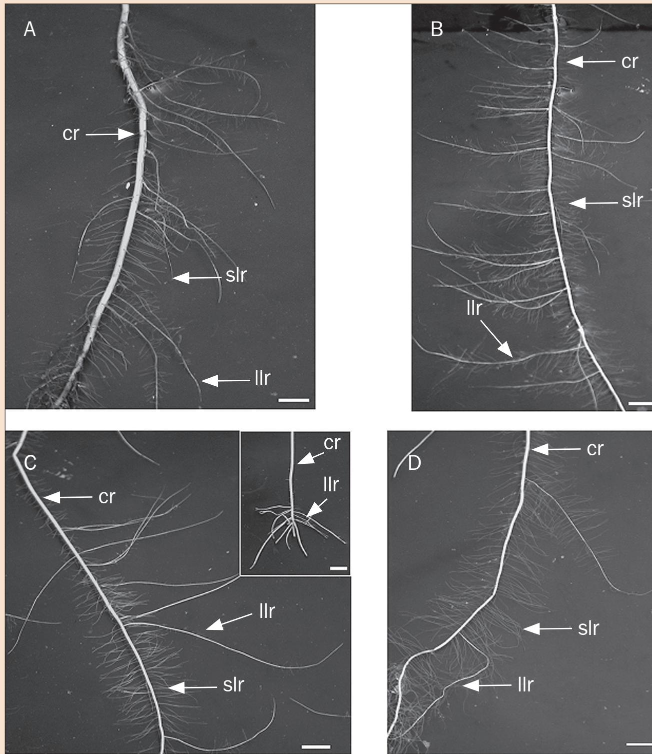


Fig. 2. Root morphology in different rice varieties. (A) Azucena, (B) IR64, (C) Kasalath, detail of a so-called "lion tail" structure, (D) Nipponbare. Cr = crown root, llr = long lateral root, slr = small lateral root. Scale of bar is 1 cm.

ible 1 day after seed imbibition and emerges by breaking the coleorhiza has mainly a role restricted to the seedling development phase. The sr primordium is formed during embryo development. After 2–3 days, five coleoptilar cr emerge from the coleoptile by breaking the sheath, resulting from *de novo* postembryonic root meristem differentiation events. The mature root system is made up of numerous cr, more than 100 in the model japonica cultivar Nipponbare, grown under hydroponics. Both sr and cr have many branching roots. Two different types of branching roots can be distinguished. Thick, ageotropic, and growth-determined small lateral roots (slr) are the most numerous, whereas a few large lateral roots (llr), thicker in diameter than the cr and with a moderate positive geotropism, sometimes appear on the crown roots. Large lateral roots also bear small lateral roots.

These categories of root axes observed in Nipponbare are also found in different japonica and indica rice varieties such as Azucena, IR64, and Kasalath (Fig. 2).

Qualitatively, the hierarchy of root development and differentiation is similar. Nevertheless, some quantitative differences in the frequency of development of llr can be pointed out, which explains in part the differences observed *in fine* for the whole aspect of the root system in these varieties. As illustrated in Figure 2, it is clear that the frequency of development of llr from cr is higher in IR64 than in Nipponbare. Under hydroponic conditions, we have frequently observed a profile of ramification of the so-called "lion tail" for Kasalath (Fig. 2). These lion-tail ramification profiles can be explained by the death of the apical meristem of the cr and by a differentiation of the near lateral roots in llr. This suggests that the differentiation of several lateral roots of llr or slr types likely involves physiological correlations between lateral root developing primordia and the apical meristem. Moreover, it is important to stress that the different roots observed in rice originate from dedifferentiation of a few cells from different tissues. Crown root meristems differentiate from stem tissues (parenchyma adjacent to xylem poles), whereas lateral root meristems differentiate from a few endodermis and pericycle cells adjacent to xylem bundles.

Comparative radial anatomy of embryonic and postembryonic roots

The radial anatomy of a crown root displays highly specialized tissues typical of a semiaquatic plant, which allows growth in flooded conditions. The transverse section through a cr shows a separation between three functional anatomical groups (Fig. 3A): first, the epidermal tissue group, then the ground tissue made up of a variable number of cortex cell layers, and then the central cylinder with conducting vessels. The epidermis is composed of two cell types: trichoblasts bearing root hairs and atrichoblasts. Water and nutrient absorption is mainly ensured by epidermal cells with or without root hairs. Following the hypodermis, the sclerenchyma differentiates as a highly lignified layer with a main function of limiting oxygen diffusion outside the root, and also plays a role as supporting tissue. The ground tissue is composed of several cortex cell layers and endodermis (Fig. 3A and B). Most of the cortical cells, representing up to 10 cell layers, will differentiate into aerenchyma with intercellular spaces developing into large lacunae. Aerenchyma have both a schizogenous and lysigenous origin. Aerenchyma allow gas transport and oxygen reserve in respiration for roots growing in flooding conditions, with limited or no access to oxygen. Endodermal cells contain a Casparian strip made of suberin extending around the cell. Endodermal cells represent a resistant apoplastic barrier to water and nutrition flow and are necessary to ensure selectivity of transfer through the symplasm. Central cylinder organization is typical of a monocotyledonous plant with a polyarch structure of xylem vessel that alternates with phloem (Fig. 3B). In the cr, four large metaxylem vessels are found in the middle of the stele surrounded by 10 smaller metaxylem vessels. The xylem forms ridges with the earliest element called protoxylem extending through the pericyle (Fig. 3B and C). Each protophloem vessel is associated with two companion cells forming a symmetrical pattern (Fig. 3D). All of the central cylinder cells are surrounded by the pericycle, defined as the limiting layer (Fig. 3B), and, in rice, with the endodermal cell being the starting point of lateral root branching. Rice, as most of the monocotyledons,

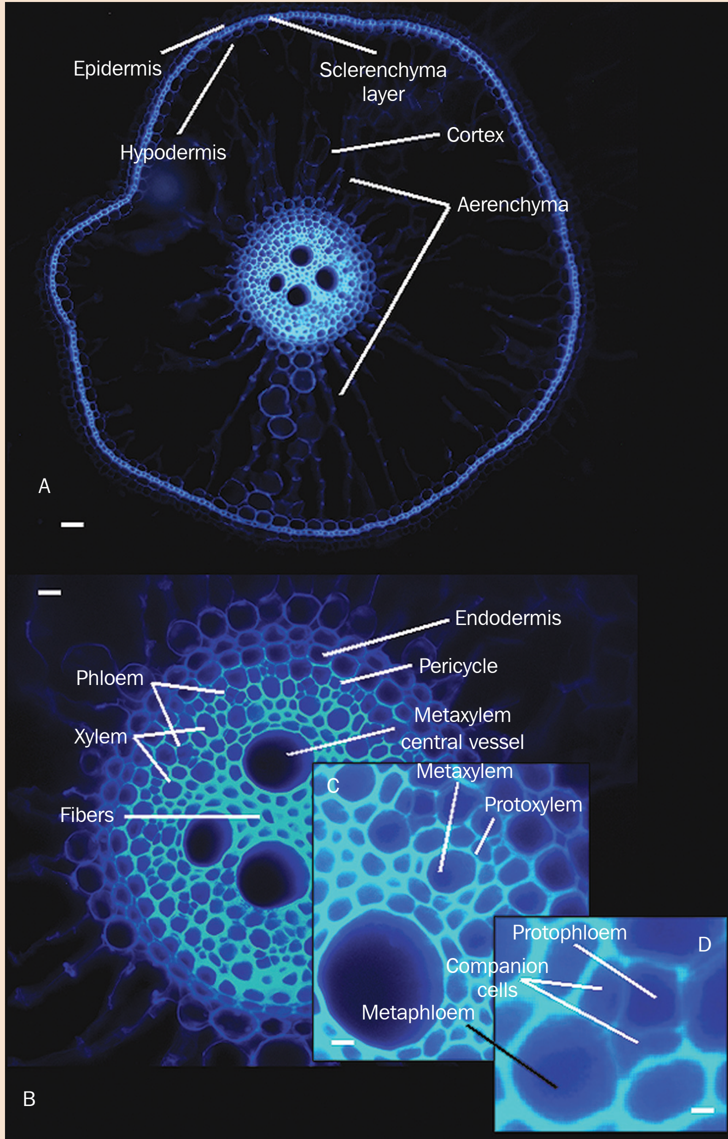


Fig. 3. Radial anatomy of a crown root. (A) Transverse section of a crown root about 6 cm above the root tip. (B, C, D) Details of the stele tissues. Scale of bars: A = 60 μm, B = 20 μm, C = 10 μm, D = 5 μm. All observations were made under UV illumination.

does not have secondary growth and all cells found between the xylem and phloem differentiate in sclerenchyma fibers (Fig. 3B).

The radial anatomy of sr and llr is similar to that of cr, whereas small lateral roots have a much simpler internal structure. The epidermis, hypodermis, and sclerenchyma are found in all root types. The internal structure is comparable among llr, sr, and cr except for a difference in cortex cell number in the ground tissue. The number of cortex cell layers (including hypodermis and sclerenchyma) is 3, 6, 8, and 13 for slr, llr, sr, and cr, respectively. These differences could explain the differences observed in diameter of the different types of roots. Slr do not have any cell layer between the endodermis and sclerenchyma and consequently do not form aerenchyma. A common central cylinder organization is shared among cr, sr, and llr even though the number of cell layers in the stele increases gradually for llr, sr, and cr. A polyarch structure of xylem vessels is visible in all rice types, with one or two large central metaxylem vessels in sr and llr and four to five metaxylem vessels surrounding sclerenchyma fibers in cr.

Root meristem structure

The cell files converge at the putative quiescent center (QC) and a small group of initials. The initials are arranged in three tiers: (1) lateral root cap and columella initial cells, (2) epidermis/endodermis initial cells, and (3) stele initial cells. The columella cells are produced through periclinal divisions of columella initials. The lateral root cap cells are formed through anticlinal division of root cap cell initials followed by additional rounds of periclinal/anticlinal divisions, whereas the production of root cap cells is independent of that of other tissues. All the ground tissues derive from a unique concentric cell layer of initials: the first anticlinal cell division generates two daughter cells, one regenerating the initial cell, the other one being the endodermis initial, which generates all the ground tissues after seven to eight asymmetric periclinal divisions. The first asymmetric periclinal division produces the epidermis and endodermis. The external epidermis cells repeat only anticlinal divisions forming a unique cell layer. The internal endodermis cell undergoes several rounds of asymmetric periclinal divisions to generate successively the exodermis, sclerenchyma, and all cortex layers. Differences in cortical layers of sr, cr, llr, and slr can be explained by a difference in number of periclinal cell divisions, ranging from 3 (slr) to 15 (cr).

Lateral root initiation starts from the protoxylem pole and proceeds through the first anticlinal divisions from single endodermal and pericycle cells, giving rise, respectively, to slr and stele. The cell number increases through periclinal divisions: three periclinal divisions give rise to the epidermis, hypodermis, sclerenchyma layer, and endodermis as in sr, but no additional cortex layer is formed.

In contrast to *Arabidopsis thaliana*, rice forms a massive postembryonic fibrous root system of adventitious origin. *Arabidopsis* has no exodermis and sclerenchyma and exhibits a single cortical layer with no aerenchyma. In *Arabidopsis*, lateral roots originating from pericycle cells of parent roots have the same anatomy as the seminal root, suggesting that they result from a reiterative process. Moreover, the division pattern of initial cells and consequently the origin of root tissues is different in

Arabidopsis and rice (e.g., in *Arabidopsis*, a common initial cell generates epidermis and root cap cells), except for the columella cells. The morphological differences and distinct anatomical organizations of the different types of roots of the rice root system may reflect different functions. Such a distinct level of differentiation of the axes constituting the root system does not exist in *A. thaliana*. The rice root anatomy has more tissues with specific cell differentiation than in *Arabidopsis*. For this reason, rice is treated as a unique model system to study new and original questions in root morphogenesis.

Genetic and molecular control of constitutive root development

Natural genetic variation and QTL mapping of constitutive root developmental traits

Rice is characterized by a shallow and limited root system compared with other crops. However, significant genetic variation occurs among different rice varieties for morphological traits such as root number, diameter, depth, branching, and vertical density distribution, and root to shoot ratio, or traits linked to root activity such as water extraction pattern, activity at depth, and root pulling force (O'Toole and Bland 1987). Genetic variability of the rice root system mostly results from contrasting cr emergence and elongation rates, final lengths, diameters, and branching densities. Differences are also observed for root plasticity in response to water stress, with some varieties being able to adjust root growth to maintain water supply.

The organization of the variability of root morphology in *Oryza sativa* reflects the overall group structure of the species. This was shown by both Courtois et al (1996) studying a set of traditional and improved upland rice varieties belonging to the indica, temperate and tropical japonica, and aus groups, and Lafitte et al (2001) studying a set of traditional varieties from all ecosystems belonging to the six isozyme groups (Glaszmann 1987). Variability between groups was observed. Indica types (enzymatic group 1) had thin, highly branched superficial roots with narrow vessels and a low root to shoot ratio. Japonica types (group 6) had thick, less-branched long roots and a larger root to shoot ratio. The aus group (group 2) was intermediate, with a root distribution profile similar to that of japonica, but thin roots. The other groups, including deepwater and floating rice (groups 3 and 4) and sadri and basmati rice originating from the Himalayan border (group 5), had root thickness and root distribution profiles closer to those of indicas (Lafitte et al 2001).

Variability within groups was limited except when the group showed a substructure linked to ecosystem adaptation. This was particularly noticeable for groups 2 and 6, the less homogeneous in terms of adaptation. The boro accessions that are grown under irrigation during the cool, dry winter season had unusually thin roots. The temperate japonica or some of the bulus, which regroup lowland varieties from Indonesia, can have a very shallow root system. For most traditional upland genotypes, already deep-rooted, it may be difficult to find donors in the *O. sativa* species to further improve the trait.

The root development of wild species, mostly belonging to the primary gene pool, has also been studied (Liu et al 2004). Although some accessions had a large constitutive root mass at depth, wild species did not appear to be a better source of alleles for improved adaptive root distribution than japonica varieties from *O. sativa*. With the development of molecular markers, the detection of QTLs controlling constitutive root traits was undertaken in recombinant inbred or doubled-haploid (DH) line mapping populations, mostly derived from indica × japonica crosses.

Root morphological traits such as maximum root length, root thickness, and vertical distribution of root mass were studied in five different populations, under either well-drained upland conditions (Champoux et al 1995, Yadav et al 1997, Courtois et al 2003) or anaerobic conditions representing rainfed lowlands (Kamoshita et al 2002a,b). Root penetration through a 5-mm-thick wax-petrolatum layer simulating the impedance of compacted rice fields was studied in the same populations, focusing on root number, number of roots penetrated and their length and thickness after penetration, and root penetration index (Ray et al 1996, Ali et al 2000, Price et al 2000, Zheng et al 2000, Zhang W et al 2001). For traits that have been studied by more than one team, a tentative synthesis is presented in Table 1 with the chromosome arm location of the QTLs. For space reasons, the synthesis does not take into account studies conducted on very young plants. A set of QTLs seems to be regularly identified on the long arm of chromosomes 1, 2, 4, 7, and 9, and, to a lesser extent, on the short arm of chromosome 3 and the long arm of chromosomes 3, 6, and 11. Those are generally the QTLs with the highest LOD scores. However, most QTLs are specific to a given study. This limited reproducibility can be due to G × E interactions, although the experimental conditions were relatively homogeneous for a group of traits, or due to threshold effects. These studies also highlighted the links between root and shoot development and the influence on root trait expression of the semidwarfism gene *sd1* located on the long arm of chromosome 1.

The genetic dissection of constitutive root traits allowed the identification of QTLs with a reasonably general effect and hence attempts were made to transfer these QTLs into elite material. Based on the QTL analysis of Yadav et al (1997), Shen et al (2001) conducted a study for introgressing alleles for root depth from Azucena, a deep-rooted variety, into IR64, which was used as the recurrent parent. DH lines of the IR64 × Azucena population carrying the favorable alleles at markers surrounding the QTLs were chosen as donor parents. Four QTLs on chromosomes 1, 2, 7, and 9 were manipulated to produce near-isogenic lines (NILs) introgressed with one or two QTLs. Three cycles of marker-assisted backcrossing were used with selection based on molecular information alone. The phenotypic evaluation of BC_3F_3 families showed that, except in the case of the QTL on chromosome 2, the introgressed QTLs were expressed in the recipient background with the expected effects. Comparisons of the NIL allelic profiles, however, did not allow locating the QTLs more precisely. The analysis of the results showed that the most efficient approach was one that localized the QTL with the smallest confidence interval possible, allowed detection of QTLs linked in a repulsion phase, and took epistasis into account. Drawing benefits from the

Table 1. Constitutive and inducible QTLs detected for rice root traits under three different growing conditions.[a]

Traits	Growing conditions			Relative value	
	Control (T)	Phosphorus deficiency (P)	Osmotic stress (O)	(P/T)	(O/T)
R/S					**C1-150(2.8)**
		C2-86(4.6)		**C2-87(4.7)**	
		C5-52(3.2)		**C5-15(3.0)**	
	C9-68(6.7)		C9-60(2.8)		C9-58(3)
	C11-51(7.5)	C11-51(3.6)	C11-54(2.7)	C11-51(5.2)	
MRL	C1-199(3.8)				C1-101(4.2)
			C2-90(3.0)		
	C3-128(3.4)	C3-138(3.5)		C3-91(3.4)	
	C4-123(3.2)	C4-113(3.8)			C4-34(4.2)
		C5-38(3.0)			
	C6-78 (3.0)	C6-81(3.6)			
				C9-73(3.2)	
			C11-52(2.8)		
					C12-114(3.6)
RB	C1-157(3.9)	C1-157(3.0)			
		C3-161(2.8)		C3-75(2.8)	**C3-80(2.9)**
	C5-86(3.0)	C5-89(3.6)			
			C11-49(3.1)		
NNR			**C1-85(2.5)**		C1-79(2.6)
		C3-39(3.0)			**C12-32(2.8)**
				C5-56(2.9)	
	C6-114(4.5)				
	C8-124(3.5)				

[a]R/S = root/shoot biomass ratio, MRL = maximum root length, RB = root biomass, NNR = number of nodal roots. C = chromosome number: 1 to 12; value after chromosome number represents QTL position on the map (cM); values in parentheses indicate LOD of the QTLs identified. Boldface data indicate inducible QTLs expressed only under stress.

availability of the rice sequence, new microsatellite markers are now used to fine-map the significant areas.

Steele et al (2006) have undertaken the introgression of QTLs from Azucena into an elite Indian line, Kalinga III. Similarly, japonica QTLs were introgressed into an indica background, but with the difference that no study had been conducted to locate QTLs in Kalinga III. This situation was closer to a normal situation for breeders than the previous one. Steele et al (2006) manipulated four QTLs on chromosomes 2, 7, 9, and 11 and pyramided them to obtain a line carrying the four QTLs in a 6-year marker-aided backcross program. These authors concluded that the QTL from chromosome 9 significantly increased root depth under both irrigated and drought-stress

treatments, whereas the other QTLs did not show significant effects in their trials. These studies, which involved a long effort, showed that QTL introgression is feasible but not without a loss of efficiency because of the limited precision in QTL location. Identifying the genes underlying the QTLs and developing better markers based on the genes themselves would be of considerable help in the use of QTLs.

Molecular control

Numerous studies have begun to identify genes involved in crown root primordia formation and emergence in monocotyledons. Several root-specific mutants have been identified in maize and rice (Hochholdinger et al 2004), without cr or sr. Isolation of a mutant without cr but with sr suggests that independent mechanisms are involved for embryonic and postembryonic root formation. *Crl1* and *Arl1* rice mutants are defective in cr formation (Inukai et al 2005, Liu et al 2005). Indeed, both mutants are allelic to the same gene. *CRL1/ARL1* encodes a member of the LOB family (of plant-specific proteins). Interestingly, other members of this protein family have been identified, which controls the development of other lateral organs in plants. Expression of *CRL1/ARL1* is regulated by auxins through the *AUX/IAA/ARF* transduction pathway by direct binding of an ERF transcription factor with an auxin-responsive element located in the *CRL1/ARL1* promoter. *Crl1* mutants have normal embryonic roots but did not form any cr primordium, indicating that *CRL1* was a master gene acting in the early steps of crown root differentiation. Crown root primordia formation is under genetic and physiological control and belongs to a normal developmental process, whereas emergence is controlled by environmental conditions. Actually, cr emergence depends on flooding of the rice plant and is mediated by ethylene action (Mergemann and Sauter 2000). Crown root growth is preceded by the induced death of epidermal cells of the node external to the tip of the root primordium (Mergemann and Sauter 2000, Steffens and Sauter 2005). Besides *CRL1/ARL1,* other genes involved in rice root development have been recently isolated and characterized. *OsRAA1* (root architectural associated) is also involved in the development of the root system mediated by auxin (Ge et al 2004), *OsGAP* (encoding GTPase-activating protein) may be involved in the mediation of root development by regulating auxin level (Zhuang et al 2005), *GNA* (glucosamine-6-phosphate acetyltransferase) is required to maintain normal root cell shape (Jiang et al 2005), and *OsPIN1* may play an important role in auxin-dependent crown root emergence and tillering (Xu et al 2005).

Root meristem organizations in *Arabidopsis thaliana* and rice are similar but not identical. Both meristems belong to the closed meristem type with initial layers surrounding a quiescent center and arranged in three tiers. A quiescent center made of four cells seems to be present in rice. Genes involved in the maintenance and function of the root meristem have been characterized. *QHB* was specifically expressed in quiescent center cells of the root apical meristem (RAM) (Kamiya et al 2003b) and OsSCR expression overlaps QHB expression in these cells (Kamiya et al 2003a), confirming that these cells formed in the quiescent center. A difference in the number of cells expressing QHB during seminal root and crown root primordia formation suggests that the mechanisms governing the ontogeny of these two types

of RAM may be different (Kamiya et al 2003b). *QHB* is the ortholog of the *WOX5* *Arabidopsis* gene, which is also specifically expressed in QC of the *Arabidopsis* RAM (Haecker et al 2004). *OsSCR* is involved in asymmetrical division of the initial cell, which will generate separate endodermal and cortical cell lineage. The continuous endodermal OsSCR expression in rice could be correlated well with the proposed function in *Arabidopsis SCR* (Di Laurenzio et al 1996). Expression of OsSHR in rice is different from that in *Arabidopsis*. *AtSHR* (Helariutta et al 2000) is expressed only in the stele region, whereas *OsSHR* is expressed in the stele as well as endodermis and in one cortex cell layer (Kamiya et al 2003a), suggesting the existence of a different molecular mechanism for cortex formation in rice and *A. thaliana*. Moreover, the additional periclinal divisions that increase the number of cortical cell layers in rice point toward another difference with *A. thaliana* during ground tissue formation.

Plasticity of root architecture

Natural genetic variation and QTL mapping of adaptive root developmental traits

Phenotypic plasticity, the ability of a plant to dynamically adjust its growth and development to environmental constraints, is of great adaptive importance (Wright and McConnaughay 2002). Adjustments in the rice root system include various changes in morphology, architecture, and biomass partitioning between roots and shoots. The environmental stimuli that induce a plastic response of the rice root system include not only soil characteristics (mineral nutrients and water status, compactness, and temperature) but also aboveground resources such as radiation. These stimuli affect root system behavior through their direct role in plant biomass production (passive plasticity in response to carbon assimilate availability) as well as acting as a signal for the regulatory systems involved in root morphogenesis (ontogenetic plasticity). Sugar concentration can also act as a signal modifying morphogenetic rules (Black et al 1995). Nitrogen application increases the total rice root length (passive plasticity) but decreases the root length/leaf area ratio (ontogenetic plasticity) and leads to a higher capacity of specific water extraction rate per unit length of root (Kondo et al 2000). Nitrate acts not only as a nutrient but also as a signal that modulates root branching. Phosphorus application enhances the depth of nodal roots and the development of fine secondary roots (Okada et al 2004) but phosphorus deficiency reduces shoot growth to the benefit of the root system and a relative increase in root growth is associated with an increased sucrose concentration in roots (Luquet et al 2005). Water deficit reduces overall plant growth but, in the case of phosphorus deficiency, the root and shoot systems are not affected to the same extent: nodal root depth is enhanced and root/shoot biomass ratio is modified to the benefit of the root system (Price et al 2002a). Soil compactness induces reduced branching and radial thickening of nodal roots and also induces changes in plant growth regulator levels, including ethylene and auxin, as well as changes in root carbohydrate levels and metabolism (Atwell 1993, Ray et al 1996).

Analysis of genetic diversity (Ingram et al 1994, O'Toole 1982, Price et al 2002a) and research on QTLs involved in root system plasticity (Hemamalini et al 2000, Price et al 2002b, Zhang J et al 2001, Zheng et al 2003) have been undertaken mainly under contrasting water supply conditions. Almost all these authors report the absence or the relatively small number of common QTLs for root traits between different growing conditions. This suggests that there is a considerable amount of QTL × environment interactions and that constitutive and water deficit–specific QTLs may both contribute to root development under water deficit. The latter may be considered as QTLs involved in phenotypic plasticity, or "response" QTLs. The most complete analysis of the genetic basis of root plasticity has been done by Zheng et al (2003), who compared root development (seminal root length, adventitious root number, and lateral root length) of a recombinant inbred line (RIL) population derived from the IR1552 × Azucena cross under flooding and under aerobic conditions. They identified nine QTLs in both conditions, but only one was common. They also identified five other QTLs for relative responses under aerobic conditions and flooding. Among them, only one co-locates with the QTLs identified in both growth conditions. The authors concluded that relative root responses should be used to evaluate root adaptation to water-limited stress.

More recently, using a new RIL population of the IR64/Azucena cross, our team analyzed rice phenotypic plasticity during early vegetative growth by comparing root/shoot biomass partitioning and plant architecture and development under three growth conditions: phosphorus deficiency, osmotic stress, and suboptimal (control) hydroponic culture. The two environmental stimuli lead to a similar expression of phenotypic plasticity: a higher allocation of biomass to roots versus shoots and enhanced root growth and development when the ontogenetic differences between environments are taken into account. Root/shoot biomass partitioning was found to be under polygenic determinism and five QTLs were detected on five different chromosomes (Table 2). Among them, two had a constitutive expression, whereas two others had inducible stress-related expression, and the last one was detected only for relative response variables. Several other inducible QTLs were detected controlling maximum root length, root biomass, and the number of nodal roots. Inducible QTLs related to either phosphorus deficiency or osmotic stress do not share common chromosomal segments. These results suggest that the response of rice to these stimuli involves different regulation and signaling networks. Looking for common points between these networks, we are currently investigating the relationships between biomass and sugar allocation to roots, gene expression, and activities of some enzymes (e.g., invertases, sucrose synthase) involved in sugar metabolism. Indeed, these relationships could be of essential importance in providing a generic explanation for root phenotypic plasticity (Black et al 1995).

Molecular response of root developmental genes to stresses

Roots are one of the primary sites for stress signal perception in which a signaling mechanism triggers a cascade of gene expression responses. These transcriptional changes can result in successful adaptations leading to stress tolerance by regulating

Table 2. Position on rice chromosome arms of QTLs for root morphology and penetration detected in different studies.[a]

Trait	Studies in which the trait was analyzed	1S	1L	2S	2L	3S	3L	4S	4L	5S	5L	6S	6L	7S	7L	8S	8L	9S	9L	10S	10L	11S	11L	12S	12L
Root thickness	1 2 3 4 5 10	1	1 2 3 4 5	1	2 4 10	1	1 3 10		1 3 4 5 10	2	2	1 2	4		1 2 5	3	1 2 3	1 10	1 2 3		1		1 4	1	10
Total root mass	2 5 10		2 5 10			2 5		10			2	2	2 10		2				2		10		5		
Deep root mass	2 3 4 5		2 4		3 4	2 3 4	2 4	2	3 4	4		2	2	2 5	2				2 3				3 4		
Deep root/tiller ratio	1 2 3 4 5		2 4 5	1 3	1 2 3 4 5			5	1 3 4	4		2 3	2	3	1 2	2	1 2		1 2				4		1
Deep root/shoot ratio	2 3 5		2		2 3 5	3		2 5	3	2			2		2		2		2 3 5				3		
Maximum root length	2 3 4 5		2 3 5		2 4		2	3	3	2 4		2			2 4		2		2 5						2 5
Root number	6 8 7 9	6 8 9	6 8		6	6	7		6		6	6	6	6 79	6 7		6	6	6			6	6	6	
Number penetrated roots	6 8 7 9	6 7			8 7 9	8	6 8 7						6		9										6
Root penetration index	6 8 7 9 10				6 8 7 9	8 9 10	8 7		6 10	6			6		9	9					7	6		10	
Penetrated root thickness	7 9 10		7 9 10		7 10	9		7	9 10				7 10	10				10	9 10		7			10	
Penetrated root length	7 10		7	7			7								7								7 10		

[a]S = short arm of the chromosome, L = long arm, I = indica, J = japonica. 1 = Champoux et al (1995), population Co39 × Moroberekan (I × J), well-drained conditions; 2 = Yadav et al (1997), population IR64 × Azucena (I × J), well-drained conditions; 3 = Kamoshita et al (2002a), population IR58221 × IR52561 (I × I), anaerobic conditions; 4 = Kamoshita et al (2002b), population CT9993 × IR62266 (J × I), anaerobic conditions; 5 = Courtois et al (2003), population IAC165 × Co39, well-drained conditions; 6 = Ray et al (1996), population Co39 × Moroberekan; 7 = Ali et al (2000), population IR58821 × IR52561; 8 = Price et al (2000), population Bala × Azucena (I × J); 9 = Zheng et al (2000), population IR64 × Azucena; 10 = Zhang et al (2001), population CT9993 × IR62266.

gene expression and signal transduction in the stress response (regulatory proteins) or directly protect the plant against environmental stress (functional proteins). Because of its high plasticity, root system architecture can be dramatically altered in response to changes in soil osmotic potential (Deak and Malamy 2005, Malamy 2005). This osmotic-responsive regulatory mechanism affects the initiation, development, and emergence of the primordium and the subsequent activation of the lateral root apical meristem.

Although so far numerous osmotic-responsive genes have been described in rice (Reddy et al 2002, Cooper et al 2003, Rabbani et al 2003, Yang et al 2004), only some of them are actually identified as potential regulators of root system architecture. Consequently, the mechanisms controlling root system architecture and developmental plasticity under osmotic stress are still poorly understood.

High-throughput detection of differentially expressed genes could help to identify genes and pathways that could play a role in early changes in rice root architecture under different osmotic conditions. Yang et al (2003) used a cDNA-AFLP analysis to examine gene expression in seminal root types at four time points during water deficit in Azucena, an upland rice variety. They identified 106 unique genes induced by the water deficit. Sixty genes of known function were involved in transport facilitation, metabolism and energy, stress- and defense-related proteins, cellular organization and cell-wall biogenesis, signal transduction, expression regulators, and transposable elements. Among transcripts reaching a maximum accumulation within 16 h of water deficit, they identified an aquaporin (PIP2a), a 9-*cis*-epoxycarotenoid dioxygenase (NCED1) involved in ABA biosynthesis, and a negative regulator of gibberellin signal transduction (SPY).

Nguyen et al (2004) refined map positions of key QTL regions for drought tolerance–related root traits by candidate genes, including drought-related ESTs, and used differential display (DD) analysis on the parental lines to define and map new full-length cDNA differentially expressed during drought stress. Twelve new cDNAs were mapped onto the genetic linkage map and six of them mapped with QTLs: the cDNA clone I12A1 that mapped within the QTL region for root thickness on chromosome 7 showed significant similarity to a UDP-glucose-4-epimerase gene. The cDNA clone T5-I23C1 mapping in a QTL region for basal root thickness matched an oryzain gamma sequence, whereas the cDNA I28C1 mapped in close proximity to a QTL determining root penetration and thickness on chromosome 6 was similar to a homeodomain transcription factor implicated in disease susceptibility; other mapped cDNAs did not match any entry in the available databases.

In our laboratory, we have used two complementary approaches, SSH (suppression subtractive hybridization) and microarray-based expression profiling, to isolate and identify candidate genes. Mannitol treatment (100 mM) was applied on rice plantlets under hydroponic conditions and 12 SSH libraries were constructed using mRNA isolated from osmotic-stressed and unstressed control tissues (leaf and root) harvested at 15 min, 1 h, 5 h, and 7 d after treatment. A set of 10,000 SSH cDNA clones selected among these libraries, together with another 12,000 rice cDNA clones from a progressive salt-stress SSH collection, were spotted onto glass slides.

The 22k microarray was hybridized with RNA from short (pool of 15 min and 1 h) and medium-term (5 h) treatments from both roots and leaves of osmotic-stressed and unstressed plants and four hybridizations were performed for each experiment. Some 43 clones were significantly under- or overexpressed by osmotic treatment. Sequence analysis and BLAST search revealed that functions could be assigned to some of these clones on the basis of homology: they encode proteins involved in signal transduction (kinases, GTP-binding protein, 14-3-3 protein, zinc-finger protein), in ROS scavenging (glutathione S-transferase), and adaptation processes (invertase, expansin, glycosyl hydrolase, sodium/calcium exchange, helicase). Moreover, two of them were identified as an *LRP1* (*LATERAL ROOT PRIMORDIUM 1*) (Smith and Fedoroff 1995) and an auxin transport BIG protein (López-Bucio et al 2005), proteins that might be involved in root system architecture adaptation in response to osmotic stress and that have so far not been described in rice. RT-PCR and RT-qPCR as well as functional validation are under way to validate these candidate genes (Brasileiro et al, unpublished).

Novel tools for investigating root developmental processes

From this discussion, it is apparent that a high degree of complexity exists in the morphology and anatomy of the cereal root system. Although rice will derive benefit from past and future discoveries in *Arabidopsis* root development, there is still certainly a need for high-throughput and high-resolution analytical tools for accelerating the identification of rice root developmental genes. During the last few years, much advancement has been made in developing a range of novel tools for such analysis.

Gene discovery and HTP gene validation through knockout (KO) analyses

In the last few years, besides classical mutant (insertion/deletion point analyses) analysis combined with TILLING-mediated searches of lesions in specific sequences, large libraries of insertion lines (reviewed in Hirochika et al 2004) and related flanking sequence tag databases (see OrygenesDB as an example, Droc et al 2006) have been generated. Most of these lines are obtained by use of the *Ac* element (Kolesnik et al 2004), the tissue culture–activated endogenous retrotransposon *Tos 17* (Miyao et al 2003), T-DNAs carrying a gene trap (An et al 2003, Sallaud et al 2004), or an activation tagging system (Jeong et al 2002). The most advanced initiative is that of NIAS in Tsukuba (Japan), where 50,000 lines totaling 250,000 *Tos17* insertions have been produced (Miyao et al 2003). One can estimate the number of insertion lines generated to be more than 250,000 and the number of insert-characterized FSTs at 70,000 (Hirochika et al 2004). Such resources will both hasten the discovery of novel root developmental and adaptive response genes through systematic screening of insertion libraries for alteration in root traits. However, this requires the setting up of efficient root-phenotyping platforms. These mutant collections will further facilitate the implementation of reverse genetics strategies for validating the function of putative rice orthologs of *Arabidopsis* root-development genes. An alternative or complement of insertion-mediated KOs lies in the use of high-throughput RNA interference vectors

based on recombinational cloning technology (Gateway) for triggering constitutive or inducible gene-specific silencing in rice plants. Inducible vectors will be very useful for root traits since regenerated T_0 plants often exhibit a non-true-to-type root system. Comparison of root architectures in seed-derived progeny, with or without induced silencing, will allow a more confident examination of phenotypic alterations related to gene silencing. An example of an inducible promoter is the dexamethasone-inducible pINDEX system, which has been found to be functional in rice (Ouwerkerk et al 2001). We recently produced co-suppression lines using gene-specific 3′ and 5′ regions of the two orthologs of *SCR* (*SCR1* and *SCR2*) and *SHR* (*SHR1* and *SHR2*) existing in the rice genome. Preliminary analysis in co-suppression lines revealed, as expected, a range of more or less severe root alterations, which are being confirmed in subsequent generations (Breitler et al, unpublished).

Two-photon laser scanning microscope

Progress has been made recently in cell imaging, allowing an *in situ* analysis through the whole-root structure in rice (Rebouillat et al, unpublished). This technical breakthrough is fundamental to further describe localization, dynamics, trafficking, and interactions of gene products using green fluorescent protein (GFP) fusions and fluorescent probes. The classical imaging system dedicated to monitoring GFP expression in plants is the confocal laser scanning microscope (CLSM). Double labeling using propidium iodide that specifically stains cell walls is by far the most popular system used by plant molecular biologists to track GFP in plant models such as *A. thaliana*. However, *A. thaliana* roots are thin enough (30–50 µm) to generate an optical median view of the whole root using CLSM and this is in constrast to roots of other species, which are generally several-fold thicker. For instance, the primary root of rice in a region close to the apex has an estimated diameter of 300 µm; therefore, for a median view, one needs to penetrate as deep as 150 µm inside the root apex. CLSM is therefore not of much use if we want to view the whole root structure at the cellular level without fixing and sectioning. In the last few years, the two-photon laser scanning microscope (TLSM) has increasingly become a standard technology for cell biology laboratories, notably in neurobiology. The TLSM allows analyses at >500 µm depth in nervous tissues (Levene et al 2004, Oheim et al 2001). Applications for plants are scarce, but Feijo and Moreno (2004) reviewed most of the TLSM applications for plants and gave insight into the high potential of the TLSM for plant cellular biology. We recently demonstrated the feasibility of deep characterization of cell-specific patterns in living primary rice roots using the TLSM. The TLSM penetrates three times deeper than the CLSM on the same sample in optimal conditions, reaching 200 µm at the focal limit of the objective we were using. As the primary rice root tip is around 300–350 µm in diameter, this is enough for a median optical view of a root.

Catalogue of lines with distinct GAL4:GFP-marked cell types

Among the T-DNA constructs used to generate insertion libraries in rice, an elegant enhancer trap (ET) system made use of the GAL4-VP16 composite transcriptional activator, which can specifically control transcription of genes driven by UAS (up-

stream activating sequence) elements. In these T-DNAs, the UAS elements control either the *gusA* (Wu et al 2003) or the *gfp* (Johnson et al 2005) reporter gene acting as a gene detector through the GAL4 enhancer trap. Screening of T_1 seedlings of more than 2,680 GAL4:UAS:GFP enhancer trap lines led researchers to identify GFP expression patterns in the root system and root organs or tissue-specific expression with frequencies of 25.7% and 5.4%, respectively. Seventy-five lines displaying either cell-specific or root-specific GFP expression were characterized more deeply in the T_2 generation (Rebouillat et al, unpublished). Sixty-five lines were confirmed in the T_2 generation for GFP expression and six were characterized at the cellular level using a two-photon microscope to gain clear insight into the tissue(s) expressing the GFP. Flanked sequence tags (FSTs) of T-DNA borders of all positive T_2 lines were then isolated, sequenced, and located on the rice genome. We are extending such analysis to all the root organ-specific and tissue-specific ET lines. This library of the rice root-specific ET library will be an efficient tool to identify root-associated genes and promoters or developmental molecular markers.

Cell-specific ectopic expression and inactivation of genes

As recently demonstrated for *Arabidopsis* lateral root founder cells and primordia (Laplaze et al 2005), the GAL4 transactivation system can be used to direct the expression of a deleterious protein such as the diphtheria toxin in specific cell lines, leading to their genetic ablation. This is a helpful alternative to laser-mediated destruction for determining the influence of cell ablation on the identity of neighboring cells and overall anatomical and architectural changes. As the transactivation has been shown to be functional in rice (Johnson et al 2005), it could be possible to test genetic cell ablation using this system in rice. Transactivation of cell-type-specific ectopic expression of genes or RNAi-mediated inactivation of gene expression is also possible. This will allow the genetic manipulation of developmental processes at a high resolution in specific tissues or cells.

Global expression profiles at a cell-specific resolution

Beyond the classical laser microdissection of tissues, the possibility of generating a global map of expression of genes within the tissues of the root organs following sorting of cells/protoplasts from GFP-marked tissues (Birnbaum et al 2003) has recently allowed a breakthrough in establishing the transcriptional profile of the *Arabidopsis* quiescent center (Nawy et al 2005). This approach should be applicable to rice roots using GFP-marked root-specific lines, with adaptations due to the presence of sclerenchyma and a higher number of cortex layers, which may complicate the protoplast study from the innermost cells of the root stele.

Conclusions and prospects

As the root system of rice appears to be more complex than that of *Arabidopsis*, some aspects related to root development will be tractable only in the model cereal species. The influence of flooding and ethylene response on the molecular control of root de-

velopment could be an issue that should be addressed in the specialized root system of rice. However, despite the recent increase in isolation and characterization of rice root development genes, knowledge accumulated on rice root development still appears rather scarce. Future research should therefore take advantage of *Arabidopsis* findings to assess the conservation of developmental controls while focusing on specificities of monocotyledonous root development, using high-throughput technologies.

Though a large amount of information on the genetic control of root constitutive and adaptive development has been accumulated, allowing the identification of many chromosomal regions influencing root developmental traits, no QTL cloning has been achieved so far. Such work is in progress for the root-depth QTL located on chromosome 9 (Ahmadi et al, unpublished). Co-localization of root morphology QTLs with root-specific, constitutive, and adaptive development genes isolated from mutant screenings, *in silico* identification of orthologs, and expression profiling studies will provide additional clues on candidate genes. Association genetics also has the power to identify new candidate genes but again points to the need for efficient root-phenotyping protocols.

Among these genes, transcription factors often represent a key intersection in the regulatory framework controlling developmental programs and the response of genes to growth regulators and environmental signals (Montiel et al 2004). High-throughput analysis of TF and identification of targets of TF expressed in the cells of interest, which has allowed tremendous progress in yeast, should now be employed in model plants such as rice. Many auxin homeostasis and plant development TF have been shown to be putative or validated targets of miRNAs in plants (Dugas and Bartel 2004), including rice (Sunkar et al 2005). Recently, *Arabidopsis NAC1*, a NAM/ATAF/CUC NAC-domain TF that transduces auxin signals for lateral root formation, has been found to be the target miR164 (Guo et al 2005). As numerous miRNA, including miR164, are conserved between *Arabidopsis* and rice (Jones-Rhoades and Bartel 2004), rice should benefit from *Arabidopsis* findings. Other efforts should be dedicated to investigating the conservation of the CLAVATA pathway in the rice root meristem. Expression of two *CLAVATA*-like and *WUS-RELATED HOMEOBOX*1(*WOX5*) genes has been detected at a low level in *Arabidopsis* roots. To date, all the phenotypes that have implicated the *WUS-CLV*1 pathway in root meristem maintenance have relied on constitutive ectopic expression, which may not reflect the effects of genes at their endogenous expression levels (Birnbaum and Benfey 2004). This holds true for the recently isolated QC-specific homeobox (*QHB*) gene that is predominantly expressed in the rice quiescent center and encodes a WUSCHEL homolog (Kamiya et al 2003b). The use of mutant libraries, which have been instrumental in isolating genes controlling root stem-cell maintenance and cell-fate decisions in *Arabidopsi*s, in combination with cell-type-specific expression data should allow the unambiguous identification of the pathway regulating root meristem maintenance and function in rice.

References

Ali ML, Pathan MS, Zhang J, Bai G, Sarkarung S, Nguyen HT. 2000. Mapping QTLs for root traits in a recombinant inbred population from two indica ecotypes in rice. Theor. Appl. Genet. 101:756-766.

An S, Park S, Jeong DH, Lee D-Y, Kang H-G, Yu J-H, Hur J, Kim S-R, Kim Y-H, Lee M, et al. 2003. Generation and analysis of end-sequence database for T-DNA tagging lines in rice. Plant Physiol. 133:2040-2047.

Atwell B. 1993. Response of roots to mechanical impedance. Env. Exp. Bot. 33:27-40.

Black C, Loboda T, Chen J, Sung S. 1995. Can sucrose cleavage enzymes serve as markers for sink strength and is sucrose a signal molecule during plant sink development. In: Pontis H, Salerno G, Echeveria E, editors. International Symposium on Sucrose Metabolism. Rockville, MD (USA): Am. Soc. Plant Physiol. p 49-64.

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN. 2003. A gene expression map of the *Arabidopsis* root. Science 302:1956-1960.

Birnbaum K, Benfey PN. 2004. Network building: transcriptional circuits in the root. Curr. Opin. Plant Biol. 7:582-588

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.

Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S, Guimil S, Dunn M, Luginbuhl P, Ellero C, Goff SA, Glazebrook J. 2003. A network of rice genes associated with stress response and seed development. Proc. Natl. Acad. Sci. USA 100:4945-4950.

Courtois B, Chaitep W, Moolsri S, Sinha PK, Trébuil G, Yadav R. 1996. Drought resistance and germplasm improvement: on-going research in the Upland Rice Research Consortium. In: Piggin C, Courtois B, Schmit V, editors. Upland Rice Research in Partnership. IRRI Discussion Paper Series 16. Los Baños (Philippines): International Rice Research Institute. p 154-175.

Courtois B, Shen L, Petalcorin W, Carandang S, Mauleon R, Li Z. 2003. Locating QTLs controlling constitutive root traits in the rice population IAC 165 × Co39. Euphytica 134(3):335-345.

Deak KI, Malamy J. 2005. Osmotic regulation of root system architecture. Plant J. 43(1):17-28.

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. 1996. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. Cell 86:423-433.

Droc G, Ruiz M, Larmande P, Pereira A, Piffanelli P, Morel J, Dievart A, Courtois B, Guiderdoni E, Perin C. 2006. OryGenes DB: a database for rice reverse genetics. Nucl. Acids Res. 34:D736-D740.

Dugas DV, Bartel B. 2004. MicroRNA regulation of gene expression in plants. Curr. Opin. Plant Biol. 7:512-520.

Feijo JA, Moreno N. 2004. Imaging plant cells by two-photon excitation. Protoplasma 223:1-32.

Ge L, Chen H, Jiang JF, Zhao Y, Xu ML, Xu YY, Tan KH, Xu ZH, Chong K. 2004. Overexpression of OsRAA1 causes pleiotropic phenotypes in transgenic rice plants, including altered leaf, flower, and root development and root response to gravity. Plant Physiol. 135(3):1502-1513.

Glaszmann JC. 1987. Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. 74:21-30.

Guo HS, Xie Q, Fei JF, Chua NH. 2005. MicroRNA directs mRNA cleavage of the transcription factor NAC1, to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell 17:1376-1386.

Haecker A, Gross-Hardt R, Geiges B, Sarkar A, Breuninger H, Herrmann M, Laux T. 2004. Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. Development 131:657-668.

Helariutta Y, Fukaki H, Wysocka-Diller JW, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN. 2000. The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signalling. Cell 101:555-567.

Hemamalini G, Shashidhar H, Hittalmani S. 2000. Molecular marker assisted tagging of morphological and physiological traits under two contrasting moisture regimes at peak vegetative stage in rice (*Oryza sativa* L.). Euphytica 112:69-78.

Hirochika H, Guiderdoni E, An G, Hsing YI, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang Q, Pereira A, Sundaresan V, Leung H, et al. 2004. Rice mutant resources for gene discovery. Plant Mol. Biol. 54:325-334.

Hochholdinger F, Park WJ, Sauer M, Woll K. 2004. From weeds to crops: genetic analysis of root development in cereals. Trends Plant Sci. 9:42-48.

Hodge A. 2004. The plastic plant: root responses to heterogeneous supply of nutrients. New Phytol. 162:9-24.

Ingram K, Bueno F, Namuco O, Yambao E, Beyrouty C. 1994. Rice root traits for drought resistance and their variation. In: Kirk G, editor. Rice roots: nutrient and water use. Manila (Philippines): International Rice Research Institute. p 67-77.

Inukai Y, Sakamoto T, Ueguchi-Tanaka M, Shibata Y, Gomi K, Umemura I, Hasegawa Y, Ashikari M, Kitano H, Matsuoka M. 2005. Crown rootless1, which is essential for crown root formation in rice, is a target of an AUXIN RESPONSE FACTOR in auxin signaling. Plant Cell 17:1387-1396.

Jeong DH, An S, Kang HG, Moon S, Han J, Park S, Lee HS, An K, An G. 2002. T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. 130:1636-1644.

Jiang H, Wang S, Dang L, Wang S, Chen H, Wu Y, Jiang X, Wu P. 2005. A novel short-root gene encodes a glucosamine-6-phosphate acetyltransferase required for maintaining normal root cell shape in rice. Plant Physiol. 138(1):232-242.

Johnson AA, Hibberd JM, Gay C, Essah PA, Haseloff J, Tester M, Guiderdoni E. 2005. Spatial control of transgene expression in rice (*Oryza sativa* L.) using the Gal4 enhancer trapping system. Plant J. 41:779-789.

Jones-Rhoades MW, Bartel DP. 2004 Computational identification of plant micro RNAs and their targets, including stress-induced miRNA. Mol. Cell 14:787-799.

Kamiya N, Itoh J, Morikami A, Nagato Y, Matsuoka M. 2003a. The SCARECROW gene's role in asymmetric cell divisions in rice plants. Plant J. 36:45-54.

Kamiya N, Nagasaki H, Morikami A, Sato Y, Matsuoka M. 2003b. Isolation and characterization of a rice WUSCHEL-type homeobox gene that is specifically expressed in the central cells of a quiescent center in the root apical meristem. Plant J. 35:429-441.

Kamoshita A, Wade LJ, Ali ML, Pathan MS, Zhang J, Sarkarung S, Nguyen HT. 2002b. Mapping QTLs for root morphology of a rice population adapted to rainfed lowland conditions. Theor. Appl. Genet. 104:880-893.

Kamoshita A, Zhang J, Sipongco J, Sarkarung S, Nguyen HT, Wade LJ. 2002a. Effect of phenotyping environment on identification of QTLs for rice root morphology under anaerobic conditions. Crop Sci. 42:255-265.

Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG, Sundaresan V, Ramachandran S. 2004. Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. Plant J. 37:301-314.

Kondo M, Murty MVR, Aragones DV. 2000. Characteristics of root growth and water uptake from soil in upland rice and maize under water stress. Soil Sci. Plant Nutr. 46:721-732.

Lafitte RH, Champoux MC, McLaren G, O'Toole JC. 2001. Rice root morphological traits are related to isozyme groups and adaptation. Field Crops Res. 71:57-70.

Laplaze L, Parizot B, Baker A, Ricaud L, Martinière A, Auguy F, Franche C, Hussaume L, Bagusz D, Haseloff J. 2005. Gal4-GFP enhancer trap lines for genetic manipulation of lateral root development in *Arabidopsis thaliana*. J. Exp. Bot. 56(419):2433-2442.

Levene MJ, Dombeck DA, Kasischke K, Molloy R, Webb W. 2004. In vivo multiphoton microscopy of deep brain tissue. J. Neurophysiol. 91:1908-1912.

Liao H, Rubio G, Yan X, Cao A, Brown KM, Lynch JP. 2001. Effect of phosphorus availability on basal root shallowness in common bean. Plant Soil 232:69-79.

Liu L, Lafitte R, Guan D. 2004. Wild *Oryza* species as potential sources of drought-adaptive traits. Euphytica 138:149-161.

Liu H, Wang S, Yu X, Yu J, He X, Zhang S, Shou H, Wu P. 2005. ARL1, a LOB-domain protein required for adventitious root formation in rice. Plant J. 43(1):47-56.

López-Bucio J, Hernández-Abreu E, Sánchez-Calderón L, Pérez-Torres A, Rampey RA, Bartel B, Herrera-Estrella L. 2005. An auxin transport independent pathway is involved in phosphate stress-induced root architectural alterations in *Arabidopsis*: identification of BIG as a mediator of auxin in pericycle cell activation. Plant Physiol. 137(2):681-691.

Luquet D, Zhang BG, Dingkuhn M, Dexe A, Vidal C. 2005. Phenotypic plasticity of rice seedlings: case of phophorus deficiency. Plant Prod. Sci. 8:145-151.

Malamy JE. 2005. Intrinsic and environmental response pathways that regulate root system architecture. Plant Cell Environ. 28(1):67-77.

Mambani B, Lal R. 1983. Response of upland rice cultivars to drought stress. III. Screening varieties by means of variable moisture along a toposequence. Plant Soil 73:73-94.

Mergemann H, Sauter M. 2000. Ethylene induces epidermal cell death at the site of adventitious root emergence in rice. Plant Physiol. 124:609-614.

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H. 2003. Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771-1780.

Montiel G, Gantet P, Jay-Allemand C, Breton C. 2004. Transcription factor networks: pathways to the knowledge of root development. Plant Physiol. 136:3478-3485.

Morita S, Yamazaki K.1993. Root system. In: Matsuo T, Hoshikawa K, editors. Science of the rice plant. Vol. 1. Morphology. Food and Agriculture Policy Research Center, Tokyo. p 161-186.

Nawy T, Lee JY, Colinas J, Wang JY, Thongrod SC, Malamy JE, Birnbaum K, Benfey PN. 2005. Transcriptional profile of the *Arabidopsis* root quiescent centre. Plant Cell 17:1908-1925.

Nemoto K, Morita S, Tadashi B. 1995. Shoot and root development in rice related to the phyllochron. Crop Sci. 35:24-29.

Nguyen TTT, Klueva N, Chamareck V, Aarti A, Magpantay G, Millena ACM, Pathan MS, Nguyen HT. 2004. Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. Mol. Gen. Genomics 272:35-46.

Oheim M, Beaurepaire E, Chaigneau E, Mertz J, Charpak S. 2001. Two-photon microscopy in brain tissue: parameters influencing the imaging depth. J. Neurosci. Methods 111:29-37.

Okada K, Kondo M, Ando H, Kakuda K. 2004. Phosphorus application affects root length distribution and water uptake of upland rice in a column experiment. Soil Sci. Plant Nutr. 50:257-261.

O'Toole J. 1982. Adaptation of rice to drought-prone environment. In: Drought resistance in crops with emphasis on rice. Manila (Philippines): International Rice Research Institute. p 195-213.

O'Toole JC, Bland WL. 1987. Genotypic variation in crop plant root systems. Adv. Agron. 41:91-145.

Ouwerkerk PBF, de Kam RJ, Hoge JH, Meijer AM. 2001. Glucocorticoid-inducible gene expression in rice. Planta 213:370-378.

Price AH, Steele KA, Gorham J, Bridges JM, Moore BJ, Evans JL, Richardson P, Jones RGW. 2002a. Upland rice grown in soil-filled chambers and exposed to contrasting water-deficit regimes. I. Root distribution, water use and plant water status. Field Crops Res. 76:11-24.

Price AH, Steele KA, Moore BJ, Barraclough PB, Clark LJ. 2000. A combined RFLP and AFLP linkage map of upland rice used to identify QTLs for root penetration ability. Theor. Appl. Genet. 100:49-56.

Price AH, Steele KA, Moore BJ, Jones RGW. 2002b. Upland rice grown in soil-filled chambers and exposed to contrasting water-deficit regimes. II. Mapping quantitative trait loci for root morphology and distribution. Field Crops Res. 76:25-43.

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol. 133(4):1755-1767.

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

Reddy AR, Ramakrishna W, Sekhar AC, Ithal N, Babu PR, Bonaldo MF, Soares MB, Bennetzen JL. 2002. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. indica cv. Nagina 22). Genome 45(1):204-211.

Sallaud C, Gay C, Larmande P, Bes M, Piffanelli P, Piegu B, Droc G, Regad F, Bourgeois E, Meynard D. 2004. High-throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. Plant J. 39:450-464.

Scheres B, Benfey PN, Dolan L. 2002. Root development. In: Somerville C, Meyerowitz E, editors. The *Arabidopsis* book. Rockville, MD (USA): American Soc. Plant Physiologists. p 1-18.

Shen L, Courtois B, McNally KL, Robin S, Li Z. 2001. Evaluation of near-isogenic lines introgressed with QTLs for root depth through marker-aided selection. Theor. Appl. Genet. 103:75-83.

Smith DL, Fedoroff NV. 1995. *LRP1*, a gene expressed in lateral and adventitious root primordia of *Arabidopsis*. Plant Cell 7(6):735-745.

Steele KA, Price AH, Sashidhar HE, Witcombe JR. 2006. Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theor. Appl. Genet. 112:208-221.

Steffens B, Sauter M. 2005. Epidermal cell death in rice is regulated by ethylene, gibberellin, and abscisic acid. Plant Physiol. 139:713-721.

Sunkar R, Girke T, Jain PK, Zhu JK. 2005.Cloning and characterization of microRNAs from rice. Plant Cell 17:1397-1411.

Takahashi A, Kobayashi A, Kakimoto Y, Fujii N, Takahashi H. 2003. Novel hydrotropism mutants of *Arabidopsis thaliana* and their altered waving response and phototropism. Biol. Sci. Space 17(3):209-210.

Takahashi H. 1997. Hydrotropism: the current state of our knowledge. J. Plant Res. 110(1098):163-169.

Tuberosa R, Sanguinetti MC, Landi P, Giuliani MM, Salvi S, Conti S. 2002. Identification of QTLs for root characteristics in maize grown in hydroponics and analysis of their overlap with QTLs for grain yield in a field at two water regimes. Plant Mol. Biol. 48:697-712.

Wright S, McConnaughay K. 2002. Interpreting phenotypic plasticity: the importance of ontogeny. Plant Spec. Biol. 17:119-131.

Wu C, Li X, Yuan W Chen G, Kilian A, Li J, Xu C, Li X, Zhou DX, Wang S, Zhang Q. 2003. Development of enhancer trap lines for functional analysis of the rice genome. Plant J. 35:418-427.

Xu M, Zhu L, Shou H, Wu P. 2005. A PIN1 family gene, *OsPIN1*, involved in auxin-dependent adventitious root emergence and tillering in rice. Plant Cell Physiol. 46:1674-1681.

Yang L, Zheng B, Mao C, Yi K, Liu F, Wu Y, Wu P. 2003. cDNA-AFLP analysis of inducible gene expression in rice seminal root tips under a water deficit. Gene 314:141-148.

Yang L, Zheng B, Mao C, Qi X, Liu F, Wu P. 2004. Analysis of transcripts that are differentially expressed in three sectors of the rice root system under water deficit. Mol. Genet. Genomics 272(4):433-442.

Yadav R, Courtois B, Huang N, McLaren G. 1997. Mapping genes controlling root morphology and root distribution in a double-haploid population of rice. Theor. Appl. Genet. 94:619-632.

Yoshida S. 1981. Fundamentals of rice crop science. Los Baños (Philippines): International Rice Research Institute. 269 p.

Zhang H, Forde BG. 1998. An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. Science 279:407-409.

Zhang J, Zheng HG, Aarti A, Pantuwan G, Nguyen TT, Tripathy JN, Sarial AK, Robin S, Babu RC, Nguyen BD, Sarkarung S, Blum A, Nguyen HT. 2001. Locating genomic regions associated with components of drought resistance in rice: comparative mapping within and across species. Theor. Appl. Genet. 103:19-29.

Zhang W, Shen X, Wu P, Hu B, Liao C. 2001. QTLs and epistasis for seminal root length under a different water supply in rice (*Oryza sativa* L.). Theor. Appl. Genet. 103:118-123.

Zheng B, Yang L, Zhang W, Mao C, Wu Y, Yi K, Liu F, Wu P. 2003. Mapping QTLs and candidate genes for rice root traits under different water-supply conditions and comparative analysis across three populations. Theor. Appl. Genet. 107: 1505-1515.

Zheng HG, Babu CR, Pathan MS, Ali L, Huang N, Courtois B, Nguyen HT. 2000. Quantitative trait loci for root penetration ability and root thickness: comparison of genetic backgrounds. Genome 43:53-61.

Zhuang X, Xu Y, Chong K, Lan L, Xue Y, Xu Z. 2005. OsAGAP, an ARF-GAP from rice, regulates root development mediated by auxin in *Arabidopsis*. Plant Cell Environ. 28(2):147-156.

Notes

Authors' addresses: C. Périn, J. Rebouillat, A. Diévart, J.C. Breitler, B. Courtois, N. Ahmadi, M. de Raissac, D. Luquet, M. Conte, D. Meynard, J.L. Verdeil, and E. Guiderdoni, CIRAD, Amis and Annual Crops Departments TA 40/03, Avenue Agropolis, 34398 Montpellier cedex 5, France, e-mail: guiderdoni@cirad.fr; A.M.C. Brasileiro, EMBRAPA-CENARGEN, Brasilia, Brazil; P. Gantet, Montpellier II University, Montpellier, France; A.A.T. Johnson, ACFPG, Adelaide, Australia; D. This, ENSAM, Montpellier, France; P.K. Pati, Guru Nanak Dev University, India; and Q.H. Le, Polytechnical Institute of Hanoi, Hanoi, Vietnam.

Acknowledgments: This work is conducted under the federative institutional project Oryzon and in the framework of the EMBRAPA Agropolis LABEX. BOYSCAT support from the Ministry of Science and Technology, India, to PKP is gratefully acknowledged. We also appreciate the support of the Generation Challenge Program.

Molecular signaling in disease resistance of rice

K. Shimamoto, A. Nakashima, M. Fujiwara, N.T. Phuong, L. Chen, H.L. Wong, D. Miki, K. Imai, S. Maisonneuve, H. Takahashi, Y. Kawaguchi, S. Hirai, and T. Kawasaki

Although impressive progress in the area of our understanding of molecular signaling in disease resistance of rice has been made recently, we still know relatively little about the molecular mechanisms of pathogen recognition and signal transduction leading to disease resistance. Increasing evidence indicates that Rac GTPase is an important molecular switch in disease resistance of rice. It activates the production of reactive oxygen species, defense gene expression, phytoalexin production, and lignin synthesis. Recent evidence suggests that it forms a protein complex with other factors involved in defense signaling. Two new technologies that are useful for the study of molecular signaling in defense responses in rice are discussed.

Keywords: Rac GTPase, proteomics. effector, protein network, reactive oxygen species

Recently, great progress has been made in our understanding of the molecular mechanisms of disease resistance in rice. A number of resistance genes against blast and bacterial blight have been isolated (Gu et al 2005, Jia et al 2000, Song et al 1995, Sun et al 2004, Wang et al 1999, Yoshimura et al 1998) and various genes that play roles in the defense against various pathogens were identified (Ahn et al 2005, Chern et al 2001, Xiong and Yang 2003, Yang et al 2004). One picture emerging from such studies is that the major factors involved in disease resistance in rice are similar to those in *Arabidopsis* and other dicots. For instance, rice homologs of the genes that play an important function in disease resistance in *Arabidopsis* often confer resistance when overexpressed in transgenic rice plants (Chern et al 2001, Fitzgerald et al 2005).

However, one critical difference demonstrated recently is that salicylic acid (SA) may not play a major role as an inducer of disease resistance as has been demonstrated in many dicots (Yang et al 2004). Instead, SA functions as an antioxidant to protect rice plants affected by oxidative damage induced by pathogen infection (Yang et al 2004). These conclusions were obtained by the analysis of transgenic rice plants in which the bacterial *nahG* gene encoding an enzyme degrading SA is expressed. Since SA plays

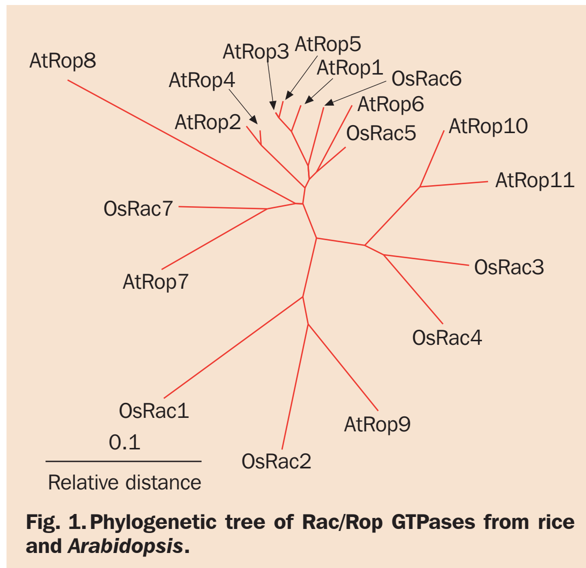


Fig. 1. Phylogenetic tree of Rac/Rop GTPases from rice and *Arabidopsis*.

a key role in the signal transduction pathway for disease resistance in *Arabidopsis* and other dicots, this finding is very important in the study of molecular signaling in rice disease resistance. It has been shown that rice has an extremely high SA level without any biotic or abiotic stress and that exogenous SA application to rice plants does not induce defense gene expression or resistance to pathogen infection (Silverman et al 1995, Yang et al 2004). If SA is not used as an inducer of disease resistance in rice, do any other chemicals play similar roles in rice? Alternatively, second messengers such as SA may not play significant roles in disease resistance in rice. New information obtained recently (Suharsono et al 2002) suggests that signaling pathways for disease resistance in rice may be partially different from those in *Arabidopsis*.

How much is known about the proteins involved in the early stages of defense responses after pathogen infection? Unfortunately, relatively little is understood. Although many R genes have been isolated in rice, no proteins that interact with those R proteins have been isolated. Furthermore, virtually nothing is known about factors that may play important roles during the early stages of defense responses in rice.

We have been studying signaling pathways involving Rac/Rop-type small GTPase. Rice Rac GTPase called OsRac1 has been shown to play an important role in disease resistance against infections by rice blast as well as bacterial blight (Kawasaki et al 1999). The involvement of Rac/Rop-type small GTPase in defense responses has been demonstrated in *Arabidopsis*, tobacco, soybean, and barley, suggesting that Rac GTPase signaling in defense responses is conserved in monocots and dicots (Moeder et al 2005, Park et al 2000, Potikha et al 1999). In this paper, we summarize our current understanding of the signaling pathways in rice in which Rac GTPase plays important roles in various steps of resistance reactions (Fig. 1).

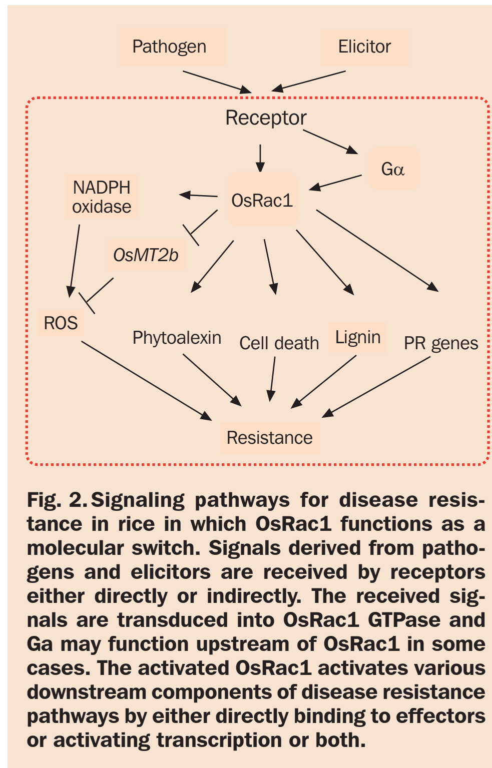


Fig. 2. Signaling pathways for disease resistance in rice in which OsRac1 functions as a molecular switch. Signals derived from pathogens and elicitors are received by receptors either directly or indirectly. The received signals are transduced into OsRac1 GTPase and Ga may function upstream of OsRac1 in some cases. The activated OsRac1 activates various downstream components of disease resistance pathways by either directly binding to effectors or activating transcription or both.

Seven *OsRac* genes in rice

Rice contains seven genes in the OsRac gene family (Fig. 2, Miki et al 2005) and OsRac1 protein was shown to have GTPase activity (Kawasaki et al 1999). In contrast to rice, *Arabidopsis* contains 11 genes (Gu et al 2004). These are highly conserved in amino acid sequences; however, their cellular functions are likely highly diverse (Gu et al 2004). It has been shown that barley may contain only six Rac genes based on an extensive search of an EST database (Schultheiss et al 2003). Interestingly, expression of the constitutively active form of one member of Rac, RACB, enhances susceptibility to powdery mildew (Schultheiss et al 2003). These results suggest that the Rac gene family may contain two distinct groups of genes with respect to the defense function. The identification of specific members of the Rac gene family for particular cellular functions is difficult since more than one member is likely to be involved in one particular function. Therefore, the constitutively active and dominant-negative forms of Rac GTPase are generated and introduced into rice plants and cell cultures and they are used for various analyses (Kawasaki et al 1999). For the cellular functions of Rac/Rop GTPase in plants, increasing numbers of cellular functions in addition to their role in defense such as pollen tube growth, cell shape control, ABA response, and auxin response have been identified recently (Gu et al 2004).

We have also generated knock-down mutants of each member of the OsRac gene family by using the RNAi method with highly diverged 3′ UTR regions as a target of silencing (Miki et al 2005). It was also shown that multiple members of a gene family can be simultaneously silenced by using either a highly conserved region to generate a trigger dsRNA or by fusing multiple gene–specific regions to make a dsRNA trigger. These RNAi-induced mutants will be used for the functional analysis of members of the gene family in the future.

Downstream factors of OsRac1 signaling in defense response

NADPH oxidase (RBOH)

We have previously shown that CA-OsRac1 is able to induce ROS production in transgenic rice plants (Kawasaki et al 1999). Rac-induced ROS production was inhibited by DPI, an inhibitor of NADPH oxidase in mammals; therefore, it was speculated that a similar oxidase is activated by CA-OsRac1 and induces ROS production in rice. We recently found that RBOHs (respiratory burst oxidase homolog), homologs of mammalian NADPH oxidase catalytic subunit gp91^{phox} (Groom et al 1996, Keller et al 1998, Torres et al 1998), are targets of Rac GTPase in rice (Nara Institute of Science and Technology, unpublished results). Plant RBOH protein has an extension at the N-terminus that is absent from the mammalian counterpart and this is the region that directly interacts with Rac GTPase. Therefore, Rac GTPase directly regulates the activity of NADPH oxidase in plants.

Metallothionein

It was shown that expression of one of the rice metallothionein (*MT*) genes, *MT2b*, is suppressed by CA-OsRac1 in rice cell cultures (Wong et al 2004). The suppression is transient. The rice MT2b protein was shown to possess antioxidant activity against superoxide and the hydroxyl radical. Therefore, it was speculated that CA-OsRac1 enhances ROS signaling during defense by transient suppression of MT gene expression. These results show that CA-OsRac1 has dual regulation of ROS production: first it promotes ROS production by activating NADPH oxidase and second it suppresses the antioxidant activity of the metallothionein protein.

Phytoalexin

We showed that synthesis of a rice phytoalexin, momilactone A, increased more than 100-fold in transgenic rice plants expressing CA-OsRac1, indicating that OsRac1 activates phytoalexin production (Ono et al 2001). Whether the activation of phytoalexin production occurred at the transcriptional level remains to be studied. Since some genes involved in phytoalexin biosynthesis have been cloned (Cho et al 2004), their induction by CA-OsRac1 can be tested in the near future.

Defense genes

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*,

was shown to be strongly down-regulated in transgenic rice expressing CA-OsRac1 (Ono et al 2001). In contrast, in transgenic rice plants expressing DN-OsRac1, CA-OsRac1 was highly expressed. *PBZ1* expression was strongly enhanced by CA-OsRac1, whereas its expression was completely suppressed by DN-OsRac1 (Suharsono et al 2002). The kinetics of *PBZ1* induction by CA-OsRac1 was similar to that in the wild type, except that it was constitutively activated without elicitor treatment in transgenic cells. In contrast, a very low level of expression was observed in cell cultures expressing DN-OsRac1. These results indicate that Rac GTPase regulates expression of defense genes.

MAP kinase

It was recently demonstrated that OsMAPK6, a homolog of tobacco SIPK and *Arabidopsis* AtMAPK, is regulated by OsRac1 (Lieberherr et al 2005). These MAPKs have been shown to be involved in defense responses in various plant species (Nakagami et al 2005). In OsRac1-RNAi plants, the level of OsMAPK protein and its kinase activity decreased, although its RNA level was not changed (Lieberherr et al 2005). These results suggest that OsRac1 is required for the stability of OsMAPK protein. The heterotrimeric G ptotein α subunit is similarly required for the stability as well as kinase activity of OsMAPK6 protein. Furthermore, results of co-immunoprecipitation experiments revealed that OsMAPK protein is in the same complex as OsRac1 (Lieberherr et al 2005). Therefore, these results strongly suggest that OsMAPK6 is one of the downstream components of Rac GTPase signaling in defense response.

Cinnamoyl-CoA reductase for lignin biosynthesis

We have recently found that cinnamoyl-CoA reductase (CCR), a key enzyme for lignin biosynthesis, is one of the downstream effectors of OsRac1 GTPase. Lignin has been shown to be one of the important components of the cell wall and it plays a role in defense against pathogen infection (Reimers and Leach 1991). We found that one of many *CCR* genes in rice, termed *CCR1*, is induced by elicitors in rice cell cultures. More importantly, it was shown that OsRac1 was able to interact with CCR1 protein and modify the enzyme activity of CCR (Kawasaki et al 2006). Therefore, direct binding of OsRac1 modifies CCR's enzyme activity. This observation is interesting since CCR functions to produce lignin monomer and ROS is required for lignin polymer formation. Since ROS is induced by Rac GTPase, ROS seems to play a dual role in lignin synthesis during defense response in rice.

Upstream factors of OsRac1 signaling in defense response

Heterotrimeric G protein α subunit

We analyzed the function of the heterotrimeric G protein α subunit during disease resistance responses by using *d1* dwarf mutants that carry mutations in Gα (Suharsono et al 2002). First, we found that its gene expression is induced by infection of leaves by rice blast fungus and treatments with elicitors (Suharsono et al 2002). Next, we found that *d1* mutants are more susceptible to infection by an avirulent race of rice

blast. Furthermore, ROS production and *PBZ1* expression induced by elicitors in wild-type cells in culture are abolished in cell cultures of *d1* mutants. However, these defects in the *d1* mutants are corrected by expression of CA-OsRac1. These results indicate that the heterotrimeric G protein α subunit is essential for R-gene-mediated disease resistance and that it functions upstream of OsRac1 in the signal transduction pathway of disease resistance in rice (Suharsono et al 2002). More recently, we found that the heterotrimeric G ptotein α subunit is also required for the stability as well as kinase activity of OsMAPK6 protein (Lieberherr et al 2005).

Other upstream factors that are possibly involved in OsRac1-mediated disease resistance in rice

Several genes involved in R-gene-mediated disease resistance have been recently identified in *Arabidopsis*, barley, and tobacco (Belkhadir et al 2004, Schulze-Lefert 2004). They include RAR1, SGT1, and HSP90 and they are suggested to function as molecular chaperones for stabilization of R proteins and other functions during defense against pathogen infection (Bieri et al 2004, Holt et al 2005). Since these proteins are also likely involved in disease resistance in rice, to understand how they are integrated into the Rac GTPase-mediated defense signaling pathway in rice will be of great interest.

Emerging technologies for the study of molecular signaling in disease resistance of rice

Proteomics

Proteomics is becoming an important technology in various areas of plant biology (Rose et al 2004, van Wijk 2001). However, this new method has not been extensively used yet in the study of molecular plant-pathogen interactions (Peck et al 2001). We have been developing methods for proteomics in the study of disease resistance in rice. We have developed a method for two-dimensional gene electrophoresis (2-DE) of rice proteins isolated from leaves and cell suspension cultures (Fig. 3). The 2-DE method has been applied for the study of *cdr* lesion-mimic mutants in rice (Takahashi et al 1999) and we found that prohibitin, a protein known to be a chaperone in the assembly of a mitochondrial respiratory chain complex in yeast and mammalian cells, is involved in programmed cell death in rice (Takahashi et al 2003). More recently, the *cdr2* mutant was used to study proteins that are possibly involved in cell death and it was found that a number of defense-related proteins and metabolic enzymes are up-regulated by the mutation (Tsunezuka et al 2005). Most recently, we applied the same technique to study the regulation of proteins by CA-OsRac1 and sphingolipid elicitors. We found that CA-OsRac1 alters the protein profile of rice cell cultures and that most of the elicitor-inducible proteins are activated by CA-OsRac1. These results suggest that CA-OsRac1 is able to reorganize protein expression for defense against pathogens and they are consistent with our previous studies demonstrating that CA-OsRac1 confers resistance to rice plants.

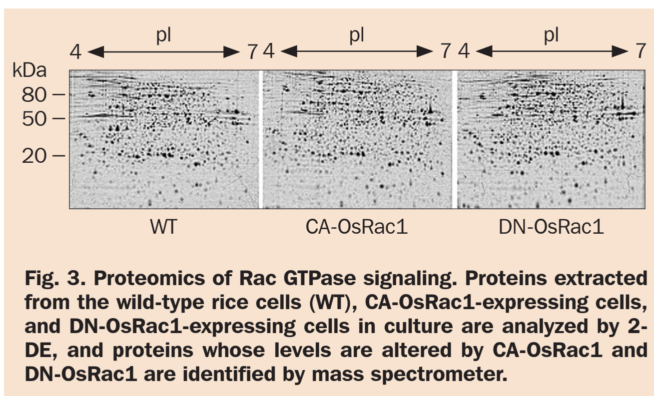


Fig. 3. Proteomics of Rac GTPase signaling. Proteins extracted from the wild-type rice cells (WT), CA-OsRac1-expressing cells, and DN-OsRac1-expressing cells in culture are analyzed by 2-DE, and proteins whose levels are altered by CA-OsRac1 and DN-OsRac1 are identified by mass spectrometer.

These studies clearly indicate the importance of proteomics in the study of disease resistance signaling in rice. This technique will be more widely used in rice in the future.

Gene-specific RNA silencing for the functional identification of rice genes

RNA silencing is becoming a powerful tool for identifying gene function in many organisms. To apply RNAi technology for rice, we developed simple Gateway-based vectors (Miki and Shimamoto 2004). The vectors can be used in either stable or transient suppression of gene function. Use of the vectors showed that the use of 3′ UTR for a double-stranded RNA trigger enables gene-specific inactivation of a member of a gene family (Miki et al 2005). This was made possible by the finding that transitive RNA silencing does not occur with endogenous rice genes. Furthermore, constructs in which the 3′ UTRs for multiple genes are fused can be used to simultaneously knock down expression of multiple genes in a gene family. These methods are very efficient and their applications for rice should generate much important information on the biology of rice in general.

References

Ahn IP, Kim S, Lee YH. 2005. Vitamin B1 functions as an activator of plant disease resistance. Plant Physiol. 138:1505-1515.

Belkhadir Y, Subramaniam R, Dangl JL. 2004. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Curr. Opin. Plant Biol. 7:391-399.

Bieri S, Mauch S, Shen QH, Peart J, Devoto A, Casais C, Ceron F, Schulze S, Steinbiss HH, Shirasu K, Schulze-Lefert P. 2004. RAR1 positively controls steady-state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. Plant Cell 16:3480-3495.

Chern MS, Fitzgerald HA, Yadav RC, Canlas PE, Dong X, Ronald PC. 2001. Evidence for a disease-resistance pathway in rice similar to the NPR1-mediated signaling pathway in *Arabidopsis*. Plant J. 27:101-113.

Cho EM, Okada A, Kenmoku H, Otomo K, Toyomasu T, Mitsuhashi W, Sassa T, Yajima A, Yabuta G, Mori K, et al. 2004. Molecular cloning and characterization of a cDNA encoding ent-cassa-12,15-diene synthase, a putative diterpenoid phytoalexin biosynthetic enzyme, from suspension-cultured rice cells treated with a chitin elicitor. Plant J. 37:1-8.

Fitzgerald HA, Canlas PE, Chern MS, Ronald PC. 2005. Alteration of TGA factor activity in rice results in enhanced tolerance to *Xanthomonas oryzae* pv. *oryzae*. Plant J. 43:335-347.

Fujiwara M, Umemura K, Kawasaki T, Shimamoto K. 2006. Proteomics of Rac GTPase signaling reveals its predominant role in elicitor-induced defense response of cultured rice cells. Plant Physiol. 140:734-745.

Groom QJ, Torres MA, Fordham-Skelton AP, Hammond-Kosack KE, Robinson NJ, Jones JD. 1996. rbohA, a rice homologue of the mammalian gp91^{phox} respiratory burst oxidase gene. Plant J. 10:515-522.

Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C, Yang F, Chu Z, Wang GL, White FF, Yin Z. 2005. R gene expression induced by a type-III effector triggers disease resistance in rice. Nature 435:1122-1125.

Gu Y, Wang Z, Yang Z. 2004. ROP/RAC GTPase: an old new master regulator for plant signaling. Curr. Opin. Plant Biol. 7:527-536.

Holt BF, Belkhadir Y, Dangl JL. 2005. Antagonistic control of disease resistance protein stability in the plant immune system. Science 309:929-932.

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.

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.

Kawasaki T, Koita H, Nakatsubo T, Hasegawa K, Wakabayashi K, Takahashi H, Umemura K, Umezawa T, Shimamoto K. 2006. Cinnamoyl-CoA reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice. Proc. Natl. Acad. Sci. USA 103:230-235.

Keller T, Damude HG, Werner D, Doerner P, Dixon RA, Lamb C. 1998. A plant homolog of the neutrophil NADPH oxidase gp91^{phox} subunit gene encodes a plasma membrane protein with Ca^{2+} binding motifs. Plant Cell 10:255-266.

Lieberherr D, Thao NP, Nakashima A, Umemura K, Kawasaki T, Shimamoto K. 2005. A sphingolipid elicitor-inducible mitogen-activated protein kinase is regulated by the small GTPase OsRac1 and heterotrimeric G-protein in rice. Plant Physiol. 138:1644-1652.

Miki D, Itoh R, Shimamoto K. 2005. RNA silencing of single and multiple members in a gene family of rice. Plant Physiol. 138:1903-1913.

Miki D, Shimamoto K. 2004. Simple RNAi vectors for stable and transient suppression of gene function in rice. Plant Cell Physiol. 45:490-495.

Moeder W, Yoshioka K, Klessig DF. 2005. Involvement of the small GTPase Rac in the defense responses of tobacco to pathogens. Mol. Plant-Microbe Interact. 18:116-124.

Nakagami H, Pitzschke A, Hirt H. 2005. Emerging MAP kinase pathways in plant stress signalling. Trends Plant Sci. 10:339-346.

Ono E, Wong HL, Kawasaki T, Hasegawa M, Kodama O, Shimamoto K. 2001. Essential role of the small GTPase Rac in disease resistance of rice. Proc. Natl. Acad. Sci. USA 98:759-764.

Park J, Choi HJ, Lee S, Lee T, Yang Z, Lee Y. 2000. Rac-related GTP-binding protein in elicitor-induced reactive oxygen generation by suspension-cultured soybean cells. Plant Physiol. 124:725-732.

Peck SC, Nuhse TS, Hess D, Iglesias A, Meins F, Boller T. 2001. Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. Plant Cell 13:1467-1475.

Potikha TS, Collins CC, Johnson DI, Delmer DP, Levine A. 1999. The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. Plant Physiol. 119:849-858.

Reimers PJ, Leach JE. 1991. Race-specific resistance to *Xanthomonas oryzae* pv. *oryzae* conferred by bacterial blight resistance gene *Xa-10* in rice (*Oryza sativa*) involves accumulation of a lignin-like substance in host tissues. Physiol. Mol. Plant Pathol. 38:39-55.

Rose JK, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS. 2004. Tackling the plant proteome: practical approaches, hurdles and experimental tools. Plant J. 39:715-733.

Schultheiss H, Dechert C, Kogel KH, Huckelhoven R. 2003. Functional analysis of barley RAC/ROP G-protein family members in susceptibility to the powdery mildew fungus. Plant J. 36:589-601.

Schulze-Lefert P. 2004. Plant immunity: the origami of receptor activation. Curr. Biol. 14: R22-24.

Silverman P, Seskar M, Kanter D, Schweizer P, Metraux JP, Raskin I. 1995. Salicylic acid in rice (biosynthesis, conjugation, and possible role). Plant Physiol. 108:633-639.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804-1806.

Suharsono U, Fujisawa Y, Kawasaki T, Iwasaki Y, Satoh H, Shimamoto K. 2002. The heterotrimeric G protein alpha subunit acts upstream of the small GTPase Rac in disease resistance of rice. Proc. Natl. Acad. Sci. USA 99:13307-13312.

Sun X, Cao Y, Yang Z, Xu C, Li X, Wang S, Zhang Q. 2004. *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. Plant J. 37:517-527.

Takahashi A, Kawasaki T, Henmi K, Shi I 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.

Takahashi A, Kawasaki T, Wong H L, Suharsono U, Hirano H, Shimamoto K. 2003. Hyperphosphorylation of a mitochondrial protein, prohibitin, is induced by calyculin A in a rice lesion-mimic mutant *cdr1*. Plant Physiol. 132:1861-1869.

Torres MA, Onouchi H, Hamada S, Machida C, Hammond-Kosack KE, Jones JD. 1998. Six *Arabidopsis thaliana* homologues of the human respiratory burst oxidase ($gp91^{phox}$). Plant J. 14:365-370.

Tsunezuka H, Fujiwara M, Kawasaki T, Shimamoto K. 2005. Proteome analysis of programmed cell death and defense signaling using the rice lesion mimic mutant *cdr2*. Mol. Plant-Microbe Interact. 18:52-59.

van Wijk KJ. 2001. Challenges and prospects of plant proteomics. Plant Physiol. 126:501-508.

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.

Wong HL, Sakamoto T, Kawasaki T, Umemura K, Shimamoto K. 2004. Down-regulation of metallothionein, a reactive oxygen scavenger, by the small GTPase OsRac1 in rice. Plant Physiol. 135:1447-1456.

Xiong L, Yang Y. 2003. Disease resistance and abiotic stress tolerance in rice are inversely modulated by an abscisic acid-inducible mitogen-activated protein kinase. Plant Cell 15:745-759.

Yang Y, Qi M, Mei C. 2004. Endogenous salicylic acid protects rice plants from oxidative damage caused by aging as well as biotic and abiotic stress. Plant J. 40:909-919.

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.

Notes

Authors' address: Laboratory of Plant Molecular Genetics, Nara Institute of Science and Technology (NAIST), 8916-5 Takayama, Ikoma 6300101, Japan.

Applied genetics

QTLs in rice breeding: examples for abiotic stresses

D.J. Mackill, B.C.Y. Collard, C.N. Neeraja, R.M. Rodriguez, S. Heuer, and A.M. Ismail

Despite the status of rice as a model agricultural crop and hundreds of studies identifying quantitative trait loci (QTLs), the applications of these results in breeding have been limited. However, the success of plant breeders in developing varieties with high yield, excellent grain quality, and wide adaptation that are widely grown by farmers (i.e., mega varieties) has provided an opportunity to deploy the most useful QTLs for rice improvement. Marker-assisted backcrossing (MAB) facilitates the precise introgression of a desired trait into the original genetic background of such mega varieties. QTLs with a large effect are rare for complex agronomic traits like yield, but are more common for other traits such as resistance to abiotic stresses. Here we discuss the example of submergence tolerance. Much of the tolerance in varieties such as FR13A has been shown to be under the control of the *Sub1* locus, which includes 2–3 tightly-linked putative transcription factors. *Sub1* was transferred into the Indian cultivar Swarna, resulting in a new version of this mega variety with tolerance of submergence. Large QTLs also exist for tolerance of salinity, P deficiency, Al toxicity, and low temperature. With some modifications, this approach may be applicable for traits controlled by multiple smaller QTLs. However, strategies for transferring multiple QTLs into mega varieties need to be developed such that negative effects of the transferred segments (linkage drag) do not adversely affect the resulting varieties. Furthermore, strategies for reducing the costs associated with marker genotyping and efficient phenotyping also need to be developed and adopted in order to apply MAB on a larger scale.

Keywords: quantitative trait loci, abiotic stress, rainfed lowland, marker-assisted selection, backcrossing

Plant breeders and scientists in related disciplines have been very successful in developing new rice varieties that have higher yield potential, resist major insect pests and diseases, and have good eating quality. The area planted with traditional rice varieties has decreased to around one-fourth in Asian countries (Evenson and Gollin 2003). However, many of these successes are temporary in nature, and additional

improvements are needed to meet the challenges of new races of diseases or insect biotypes, improved grain quality, and adaptation to diverse growing conditions. Although improvements in yield potential have been difficult to attain, many other plant traits have been manipulated to develop improved rice cultivars beneficial to farmers and consumers.

The plethora of new tools associated with the field of biotechnology has generated intense interest and optimism among scientists. However, significant advances are still to be made in employing these techniques in developing improved rice cultivars. DNA markers have been widely adopted by rice geneticists and breeders to assist in their breeding programs and the use of these markers in mapping populations has enabled discovery of the individual quantitative trait loci (QTLs) that underlie the agronomic traits that breeders want to improve. Despite such widespread use, examples of the practical applications of QTLs in plant breeding have been very limited, and few if any cultivars released to farmers are derived from targeted manipulation of known QTLs. In our paper, we provide an example of how the identification and transfer of QTLs into widely grown varieties through marker-assisted selection is an extremely powerful tool for developing improved rice for rainfed conditions.

Development and spread of mega varieties

After the rapid and inspiring success of the Green Revolution's high-yielding varieties (HYV) in the 1960s and 1970s, rice scientists began to turn their attention to the problem of improving rainfed lowland rice grown in areas where water control was poor or nonexistent. The ecosystem breeding approach as advocated by Buddenhagen (1978) suggested that rainfed environments were more complicated than irrigated rice environments, and improved varieties would have to be tailored to diverse situations: "Every field, every farm is its own reality" (Buddenhagen 1986). This was reinforced by the existence of many traditional rice varieties with varying characteristics grown in upland and rainfed lowland environments (Mackill et al 1996). The assumption was that each region and indeed individual fields within regions would have their own unique requirements for improved varieties.

The picture began to change with the development and spread of the variety Mahsuri, which can be considered the first rainfed "mega variety" (e.g., a variety grown on a very large area, usually several million hectares, and that is exceptionally popular with farmers) (Mackill 2006). This variety was developed from an indica/japonica crossing program initiated by FAO in the 1950s. Variety Mahsuri was selected in Malaysia (Samoto 1965) and subsequently evaluated in a number of Asian countries. It spread rapidly, mostly because of farmer-to-farmer contact, and was eventually being grown by millions of farmers in India, Nepal, Bangladesh, and Myanmar. This variety did not look like a modern HYV; it was taller and had a pale green leaf color. However, it had two major advantages: its ability to give a significant yield advantage in low-fertility soils in the region with no changes in crop management and its highly desirable grain quality. Despite its lack of tolerance of drought or submergence, it became widely grown in rainfed areas.

Plant breeders attempted to develop varieties for the rainfed lowlands that were tolerant of water excess or deficit and had higher yield potential than the traditional varieties. Some improved varieties were developed (see examples in Mackill et al 1996); however, they had limited success in terms of area covered. Plant breeders working in areas where Mahsuri was grown began working on developing higher-yielding varieties that would capture some of its advantages. The first success was variety Swarna (MTU7029), developed in the early 1980s in Andhra Pradesh, India. This variety began spreading and is now grown on more than 5 million ha in India and also in the northern areas of Bangladesh. Additional mega varieties are Samba Mahsuri (BPT5204) from India, BR11 from Bangladesh, and IR64 from IRRI, Philippines. These varieties are grown in both irrigated and rainfed lowland areas. They also have high yield potential and preferred quality characteristics. However, they are relatively poor in tolerance of abiotic stresses. In addition, they are usually susceptible to diseases and insects (IR64 being an exception, with resistance to many stresses).

The spread of mega varieties in rainfed lowland areas indicates that it is possible to replace the diverse traditional varieties in these areas with widely adapted varieties. However, these mega varieties lack many desirable features. The spread of a few varieties over large areas would indicate that they meet the most pressing needs of farmers, and that few other varieties measure up to these standards. In addition, once these varieties become widely established, they set the benchmark for any future cultivars that are being evaluated as replacements. It is possible that a completely new cultivar will be identified that replaces the existing mega variety and becomes the new benchmark. However, this type of success will be very rare. Although rice breeders will and should continue their efforts to develop replacements for the mega varieties, an alternative approach of an "upgrade" to the mega varieties can also be pursued as a parallel strategy to improve the well-being of farmers in rainfed areas.

Big QTLs versus small QTLs

The use of QTL analysis for understanding the genetic control of quantitative traits has been widely applied in rice since the early 1990s (see reviews in Yano and Sasaki 1997, Li 2001, Xu 2002). More than 7,000 rice QTLs are listed in the Gramene database (www.gramene.org/qtl/index.html). However, many problems limit the applications of QTLs in plant breeding (Holland 2004, Mackill 2006). These include the small effects of many QTLs, low resolution in locating the QTL on the chromosome, lack of polymorphic markers that can detect the QTL, and the effect of genetic background on expression of the QTL. Because of these limitations, the manipulation of QTLs in plant breeding is problematic.

Many of the limitations of QTLs could be overcome if QTLs of larger effect could be identified. For some quantitative traits like grain yield, this may not be possible, but, for other traits, larger QTLs have been identified. This is particularly true for tolerances of abiotic stresses, for which QTLs of significant effect can be discerned (Mackill 2006). Several of these QTLs have large LOD scores and explain an appreciable percentage of the phenotypic variance (Table 1). This indicates that

Table 1. Examples of QTLs with relatively large effects controlling tolerance of abiotic stresses.

Trait	Chr	Markers	LOD	R^2	Population[a]	Reference
Al toxicity (RRL)	3	CD01395-RG391	8.4	24.9	IR64/***O. rufipogon***	Nguyen et al (2003)
Submergence	9	C1232	36.0	69.0	**IR40931-26**/PI543851 (japonica)	Xu and Mackill (1996)
P deficiency	12	G2140-C443	10.7	27.9	Nipponbare/**Kasalath**	Wissuwa et al (1998)
Salt tolerance	1	C52903S-C1733S	6.7	43.9	**Pokkali**/IR29	Bonilla et al (2002)
Al toxicity	1	CD0345	8.1	24.1	**CT9993**/IR62266	Nguyen et al 2002)
Al toxicity	8	C1121	8.2	28.7	**CT9993**/IR62266	Nguyen et al (2002)
Al toxicity	12	RG9	6.8	20	IR1552/**Azucena**	Wu et al (2000)
P deficiency	12	RG9-RG241	16.5	54.0	**IR20**/IR55178-3B-9-3	Ni et al (1998)
Cold tolerance	12	RM101-RM292	18.5	41.7	**M-202**/IR50	Andaya and Mackill (2003)
Drought–cell membrane stability	3	RZ403	12.1	42.1	CT9993/**IR62266**	Tripathy et al (2000)
Drought–cell membrane stability	9	RZ698-RM219	10.4	37.4	**CT9993**/IR62266	Tripathy et al (2000)
Grain yield–drought	12	AFLP	7.5	22.3	**CT9993**/IR62266	Babu et al (2003)

[a]Boldface indicates the donor parent for abiotic stress tolerance.

these QTLs would provide a measurable improvement to a trait when introduced into a particular genetic background as a single locus. Most of the best QTLs for abiotic stress tolerance were identified in nonelite germplasm, suggesting that these alleles are different from those present in elite cultivars. These QTLs can therefore be useful for improving the elite mega varieties currently grown.

DNA (or molecular) markers have been advocated as a means to selectively transfer a small chromosomal fragment into a recurrent parent in an accelerated backcrossing scheme (Young and Tanksley 1989). This produces a plant with the desired graphical genotype, which is the recurrent parent plus the inserted fragment. Selection of the backcross (BC) plants occurs in three stages. First, "foreground selection" is performed by selecting plants possessing the target locus, using target gene- or QTL-specific markers (or a phenotypic screening). Individual selected plants are then screened for markers flanking the target locus. Recombinants for a marker on one side of the target locus are then selected (i.e., those plants having the recurrent parent marker for that flanking marker). In this way, linkage drag is minimized because the donor chromosomal segment is relatively small. In most papers, this is referred to as a type of background selection, but here we refer to this as "recombinant selection." Since double recombination events occurring on both sides of a target QTL are extremely rare events, plants that contain recombination events between flanking markers and a target QTL are selected using at least two BC generations (Frisch et al 1999b). The remaining plants are subsequently screened with markers that are unlinked to the target gene, and the plant with the fewest donor fragments is selected for the next cycle of backcrossing ("background selection"). In the second backcross, recombinant selection is performed for the other flanking marker.

This scheme works best when a large number of backcross seeds are available in each cycle. For example, Chen at al (2000) used this approach to transfer the bacterial blight resistance gene *Xa21* into a widely used restorer line in China. Frisch et al (1999a,b) have provided some guidelines for the sample sizes needed for this MAB approach. The transfer of a small donor segment is achieved with reasonable population sizes by using at least two backcrosses, and recombinant selection is performed for one flanking marker in the first BC and for the other side in the second BC. In fact, efficiency is actually improved by using three backcrosses, which minimizes the total number of marker data points needed for the introgression while adding one additional generation for the third backcross (Frisch et al 1999a).

A case study of QTL breeding: submergence tolerance

Submergence tolerance provides a good example of how the MAB approach can be applied to introduce a useful QTL into a mega variety. Intermittent flooding is common on 10–15 million ha of rainfed lowland rice in South and Southeast Asia. Although most varieties are intolerant of over 4 d of submergence, a handful of varieties are known to possess tolerance for up to 2 wk. One of the best studied varieties is FR13A, from Orissa, India. In contrast to floating rice grown in very deep water, submergence-tolerant varieties survive by inhibiting elongation growth during submergence, thereby

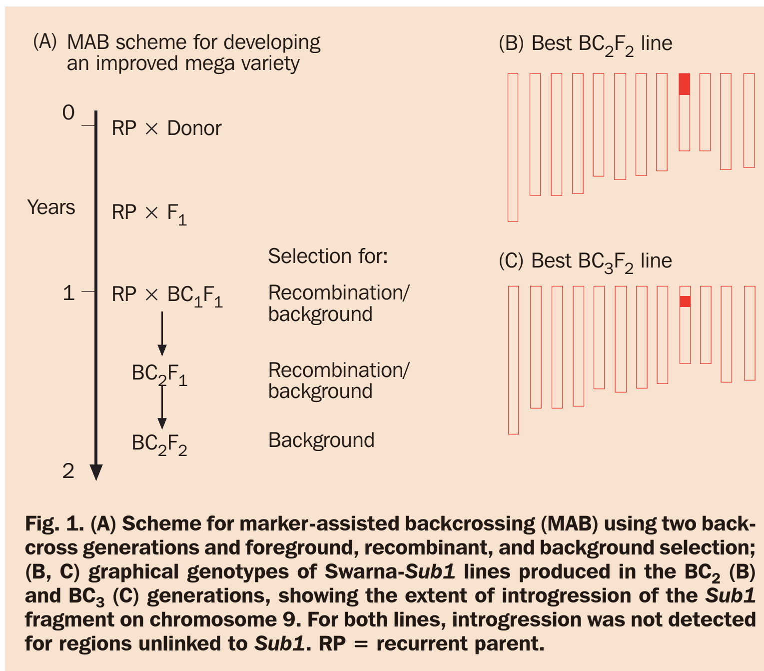


Fig. 1. (A) Scheme for marker-assisted backcrossing (MAB) using two backcross generations and foreground, recombinant, and background selection; (B, C) graphical genotypes of Swarna-*Sub1* lines produced in the BC_2 (B) and BC_3 (C) generations, showing the extent of introgression of the *Sub1* fragment on chromosome 9. For both lines, introgression was not detected for regions unlinked to *Sub1*. RP = recurrent parent.

conserving carbohydrate reserves. This allows them to survive for a longer period of time under water, and resume growth when desubmerged. Most of the tolerance in this trait is controlled by the *Sub1* locus, which has been mapped on rice chromosome 9 to very high resolution (Xu et al 2000). A transcription factor in this region that is induced under submergence (described below) was used to develop a gene marker in the region that is associated with tolerance.

Mega variety Swarna was chosen as the first target for MAB of submergence tolerance. In the BC_1 generation, 697 plants were screened for submergence tolerance, and the 376 tolerant plants identified were evaluated for two *Sub1* flanking markers, RM464 and RM219. Twenty-one plants recombinant for RM219 and *Sub1* were screened for unlinked SSR background markers, and one plant with nine donor chromosome fragments was selected for producing the BC_2. In the BC_2F_1, only a few of the 320 plants were recombinant with RM464 and *Sub1*. Unfortunately, these plants had too many donor fragments to produce the optimum genotype in the BC_2F_2 generation. One of these recombinant plants was crossed to produce a BC_3F_1 population and from 18 BC_3F_1 plants we were able to identify BC_3F_2 plants that were homozygous for all Swarna fragments and had the optimum genotype. The selected plant had an introgression around *Sub1* amounting to about 2.6 Mb of donor DNA, or less than 1% of the genome (Fig. 1A).

In the BC_2F_1 generation, one plant that was not recombinant for RM464 had only three donor chromosome fragments in addition to *Sub1*. This plant was self-pollinated

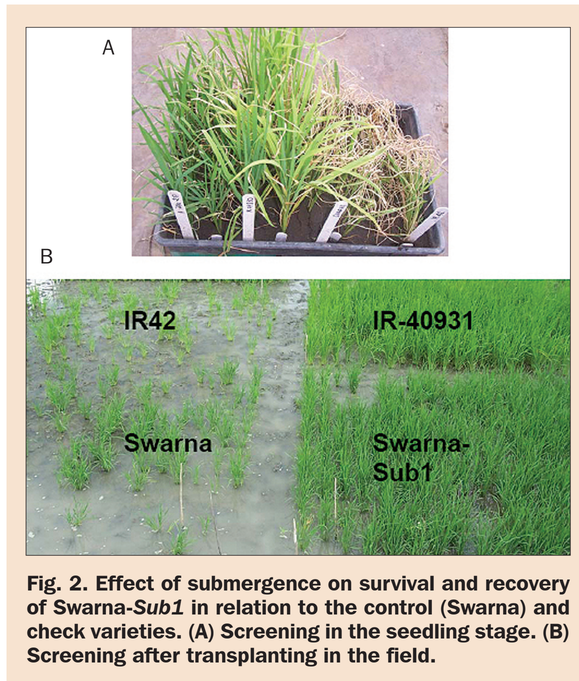


Fig. 2. Effect of submergence on survival and recovery of Swarna-*Sub1* in relation to the control (Swarna) and check varieties. (A) Screening in the seedling stage. (B) Screening after transplanting in the field.

and, from 973 BC_2F_2 plants, one plant was identified that was homozygous for all Swarna fragments except *Sub1*. This plant possessed the entire end of chromosome 9, which amounted to approximately 7 Mb of donor introgression around *Sub1* (Fig. 1B). Because this plant was available earlier than the BC_3 line, and had minimal introgression from the *Sub1* donor (*Sub1* is near the tip of chromosome 9), it was used for experiments to measure the effect of *Sub1* in the Swarna background.

The Swarna-*Sub1* line has been evaluated for agronomic performance at IRRI and is currently also being tested at rainfed lowland sites in eastern India and Bangladesh. Analysis of yield, morphological traits, and grain quality did not detect a measurable difference between Swarna and Swarna-*Sub1*, except for a change in hull color. Swarna has a gold-colored hull, whereas Swarna-*Sub1* has a straw-colored hull. This change appears to result from a gene linked to *Sub1*. Although the exact gene is not known, there is a locus designated "inhibiter of brown furrows" (*IBf*) at this position as described in the classical map (Takahashi 1964). This morphological change is not significant in terms of plant performance, and has the beneficial effect of making the *Sub1* version of Swarna distinguishable from the original. Swarna-*Sub1* has substantially improved submergence tolerance over Swarna (Fig. 2).

The beneficial effect of *Sub1* is most obvious when the screening includes a nearly complete desubmergence after the stress period, allowing plant recovery. Most varieties respond to submergence by rapid elongation of the shoots, a mechanism that

allows the plants to grow out of the water if the water stagnates for a longer duration as in deepwater areas. However, this elongation is disadvantageous if complete submergence occurs for only a brief period of up to 2 wk (flash flooding), particularly if the water level is too high. Tolerant plants show inhibited elongation, and this allows them to conserve reserves and resume growth upon desubmergence (Das et al 2005). For stagnant flooding at deeper water levels, shoot and internode elongation will sustain survival, but, for intermittent flooding (up to 14 d), submergence tolerance is a better mechanism and is more compatible with high yields. Elongating plants under intermittent flooding tend to lodge if they survive flooding and this results in severe yield reduction and poor grain quality.

The MAB approach outlined above is now being used to transfer *Sub1* into other mega varieties. A similar project has also begun for salt tolerance, which is largely controlled by a QTL on chromosome 1 (Bonilla et al 2002). This approach saves considerable time and effort in screening. The entire project can be carried out using a very small area of field or greenhouse space. Phenotyping can be avoided until a plant with the desired graphical genotype is obtained. This allows a detailed comparison between the enhanced, tolerant mega variety and the original mega variety. With conventional backcrossing, it is expected that linkage drag and random segments from noncarrier chromosomes will be transferred to the progeny. Although plants that look like the mega variety can be selected for, it is not possible to evaluate the true effects of these donor segments without extensive testing and farmers' evaluation. It is thus reassuring to concentrate testing on only the breeding lines with minimal introgression.

The transfer of *Sub1* into Swarna was made easier by the observation that the respective region on chromosome 9 appeared to have a japonica origin in variety Swarna. One of the parents of Swarna is Mahsuri, which derives from an indica/japonica cross, explaining this observation. Therefore, many polymorphic markers in this region could be used for foreground and recombinant selection. When transferring *Sub1* into indica varieties, a lack of polymorphism was observed in many of the markers used and it was necessary to develop new markers for these crosses. For the MAB strategy, markers flanking the gene at around 3–5 cM on either side of the desired locus are necessary for recombinant selection. For the region of chromosome 9, markers were obtained from the Nipponbare genome sequence and from markers already published (McCouch et al 2002). A set of markers was identified that was polymorphic in various parents (Table 2). With this set of markers, the locus can be transferred with minimal additional donor segments into a wide range of varieties and breeding lines. Because of the high frequency of SSR markers in rice, and the availability of the whole-genome sequence, this toolkit of flanking markers can be developed for introgression of any genes or QTLs into a range of rice germplasm.

Using gene markers for QTL introgression

Introducing QTLs through the MAB approach will be facilitated by using markers that are based on the gene(s) underlying the QTL. For the *Sub1* locus, two putative ethylene responsive transcription factors (ERF) have been identified that are associated with

Table 2. Markers being used for introgression of *Sub1* into a range of rice varieties.

Marker	Position[a] (Mb)	Varieties[b]					
		Swarna	Samba	IR64	BR11	CR1009	TDK1
RM3609	0.7		+	+	+	+	
RM8219	1.05		+		+	+	+
RM8206	1.1		+	+	+	+	
RM5899	4.3			+	+		
RM6920	5.8	+				+	
RM5515	5.95	+					
RM5526	6.1					+	+
RM1328	6.1					+	+
RM3855	8.05	+	+			+	
RM7364	8.25	+	+	+	+	+	+
RM5777	8.8	+	+				
RM7390	9.0		+	+	+		
RM3912	9.5		+				
RM3769	10.4		+	+	+	+	
RM1896	10.4		+	+	+	+	
RM7038	10.4	+	+	+	+	+	
RM6051	11.5	+			+		

[a]Position from the tip of the chromosome in megabases of DNA estimated from the Nipponbare genome sequence. [b]+ = primers generate polymorphic markers between the submergence source IR49830-7-1-2-1-3 and the variety indicated.

the tolerant phenotype (Xu et al, submitted). One of these genes, designated *Sub1A*, contains two single nucleotide polymorphisms (SNPs) that distinguish between tolerant (*Sub1A-1*) and intolerant alleles (*Sub1A-2*). From one SNP, a cleaved amplified polymorphic sequence (CAPS) marker was developed that takes advantage of an *Alu*I restriction site and can be assessed in agarose gels. This marker has greatly facilitated the identification of tolerant plants because it is polymorphic in all the populations used so far, and there is no recombination with the submergence trait. The SNP currently being used is silent and therefore not thought to be causal for the trait. It might not be useful in all situations, but we have not encountered any problem in currently used varieties. One of the two additional ERF genes present in the *Sub1* locus likewise possesses a tolerant (*Sub1C-1*) allele and various intolerant (*Sub1C-2* to *Sub1C-8*) alleles, and markers are also being developed for this gene.

Another advantage of having the sequence of the gene is that it can be used for allele mining of germplasm to identify varieties with different tolerance genes. Up to now, *Sub1* is the only major locus that has been detected for tolerance, and it is present in the most tolerant varieties analyzed so far (FR13A, Goda Heenati, Kurkaruppan). It is known that other QTLs can increase submergence tolerance in an additive manner

with *Sub1*, as seen with FR13A (Nandi et al 1997, Toojinda et al 2003). We should therefore search for other genes or QTLs that can be used in conjunction with *Sub1* or that might be superior to *Sub1*. Using *Sub1A* and *Sub1C* genes, we can develop markers to screen germplasm collections and identify new alleles. In addition, we can screen tolerant varieties and search for any that have an intolerant allele at *Sub1*. Those with the intolerant allele may possess a different major QTL for submergence tolerance, and further genetic analysis can be performed to map this QTL and pyramid it together with *Sub1*.

Implications for breeding for rainfed conditions

The approach outlined above represents a strategy for applying the tools of QTL mapping and MAS in applied breeding programs. There is a clear advantage of this approach over conventional breeding because the MAB strategy results in the development of an ideal genotype within a short period of time. Thus, it is not only cost-effective, but it results in a product that could not be realistically derived through conventional breeding. However, several complicating factors need to be examined before applying this approach more widely in rice breeding.

The first concern is that the approach is inherently conservative, and the end result would be to limit progress to the genetic background of a handful of mega varieties. This criticism is not persuasive for several reasons. First of all, with improvements in technology, this MAB approach can be used more widely with more diverse germplasm. As an example, our submergence project is transferring the *Sub1* gene into at least six varieties at IRRI, and we are collaborating with other institutes so that they can transfer it into their promising varieties. The use of a few varieties on large areas always raises questions about the loss of biodiversity and the potential for susceptibility to new pests. In practice, this loss of variability is nearly always associated with the evolution of superior cultivars, and it probably has more beneficial effects than negative ones.

Another concern is that the mega varieties are actually deficient in multiple traits and it is still rare that a single QTL can make a marked improvement in phenotype. This is undoubtedly true, and multiple introgressions will be important to further optimize mega varieties. Strategies for multiple traits are not as well developed, but certainly the approach will be even more relevant for this situation. As discussed above, the standard backcrossing approach will result in unacceptably high levels of introgression of donor segments for a single desired locus, and the disadvantages will be compounded for multiple loci. If linkage drag results in a five- to tenfold increase in introgression over recombinant selection, this must be multiplied by the number of loci being introduced. It seems that the best approach for introducing multiple traits will be to do it in parallel, with a final pyramiding using the individual components. This has the advantage that single introgressions may be preferable for some situations. Of course, once an enhanced mega variety is developed, it can be easily used to build upon the next one. We are following this approach to combine submergence tolerance with salt tolerance.

The advantages and appropriateness of MAB to enhance the mega varieties and other important varieties is straightforward, and it should certainly find wide application in rice breeding programs. This might be said to answer concerns that molecular marker technology has not provided breeders with practical tools, or that these techniques have had limited impact in rice breeding. However, the substantial portion of rice breeding programs will remain synthetic in nature, and in these activities the use of MAB is still more problematic. This will be gradually overcome with the development of lower-cost markers that are based on candidate genes. In the *breeding-by-design* approach (Peleman and van der Voort 2003), it is envisaged that breeders will be manipulating traits as genes and chromosomal fragments. Graphical genotypes in segregating populations will allow breeders to synthesize new genotypes before extensive field testing is required. One caution is that these new technologies will be available only in wealthier countries at first, and it will take a longer time before they can be employed in Asian rice breeding programs. Costs for marker assays need to be considerably reduced to apply MAS on larger populations. For the near-term future, the MAB approach is the technology that deserves widespread application in rice.

References

Andaya VC, Mackill DJ. 2003. Mapping of QTLs associated with cold tolerance during the vegetative stage in rice. J. Exp. Bot. 54:2579-2585.

Babu RC, Nguyen BD, Chamarerk V, Shanmugasundaram P, Chezhian P, Jeyaprakash P, Ganesh SK, Palchamy A, Sadasivam S, Sarkarung S, Wade LJ, Nguyen HT. 2003. Genetic analysis of drought resistance in rice by molecular markers: association between secondary traits and field performance. Crop Sci. 43:1457-1469.

Bonilla P, Dvorak J, Mackill DJ, Deal K, Gregorio G. 2002. RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (*Oryza sativa* L.) using recombinant inbred lines. Philipp. Agric. Sci. 85:68-76.

Buddenhagen IW. 1978. Rice ecosystems. In: Rice in Africa. Buddenhagen IW, Persley G, editors. N.Y. (USA): Academic Press. p 11-27.

Buddenhagen IW. 1986. Strategies and approaches to wetland rice improvement. In: Juo ASR, Lowe JA, editors. The wetlands and rice in subsaharan Africa. Ibadan (Nigeria): International Institute of Tropical Agriculture. p 97-106.

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.

Das KK, Sarkar RK, Ismail AM. 2005. Elongation ability and non-structural carbohydrate levels in relation to submergence tolerance in rice. Plant Sci. 168:131-136.

Evenson RE, Gollin D. 2003. Assessing the impact of the Green Revolution, 1960 to 2000. Science 300:758-762.

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.

Holland JB. 2004. Implementation of molecular markers for quantitative traits in breeding programs: challenges and opportunities. Paper presented at New Directions for a Diverse Planet, Proceedings of the 4th International Crop Science Congress. Brisbane, Australia.

Li Z. 2001. QTL mapping in rice: a few critical considerations. In: Khush GS, Brar DS, Hardy B, editors. Rice genetics IV. Los Baños (Philippines): International Rice Research Institute. p 153-171.

Mackill DJ. 2006. Breeding for resistance to abiotic stresses in rice: the value of QTLs. In: Lamkey K, Lee M, Hallauer AR, editors. Plant breeding. The Arnel R Hallauer International Symposium. Ames, Iowa (USA): Blackwell Pub. p 201-212.

Mackill DJ, Coffman WR, Garrity DP. 1996. Rainfed lowland rice improvement. Manila (Philippines): International Rice Research Institute. 242 p.

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Zing Y, et al. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res. 9:199-207.

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.

Nguyen BD, Brar DS, Bui BC, Nguyen TV, Pham LN, Nguyen HT. 2003. Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). Theor. Appl. Genet. 106:583-593.

Nguyen VT, Nguyen BD, Sarkarung S, Martinez C, Paterson AH, Nguyen HT. 2002. Mapping of genes controlling aluminum tolerance in rice: comparison of different genetic backgrounds. Mol. Genet. Genomics 267:772-780.

Ni JJ, Wu P, Senadhira D, Huang N. 1998. Mapping QTLs for phosphorus deficiency tolerance in rice (*Oryza sativa* L.). Theor. Appl. Genet. 97:1361-1369.

Peleman JD, van der Voort JR. 2003. Breeding by design. Trends Plant Sci. 8:330-334.

Samoto S. 1965. Rice breeding. In: Symposium on rice culture in Malaya. Center for Southeast Asian Studies, Kyoto University. p 24-35.

Takahashi M. 1964. Linkage groups and gene schemes of some striking morphological characters in Japanese rice. In: Proceedings of the Symposium on Rice Genetics and Cytogenetics. Manila (Philippines): International Rice Research Institute. p 215-236.

Toojinda T, Siangliw M, Tragoonrung S, Vanavichit A. 2003. Molecular genetics of submergence tolerance in rice: QTL analysis of key traits. Ann. Bot. 91:243-253.

Tripathy JN, Zhang J, Robin S, Nguyen TT, Nguyen HT. 2000. QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress. Theor. Appl. Genet. 100:1197-1202.

Wissuwa M, Yano M, Ae N. 1998. Mapping of QTLs for phosphorus-deficiency tolerance in rice (*Oryza sativa* L.). Theor. Appl. Genet. 97:777-783.

Wu P, Liao CY, Hu B, Yi KK, Jin WZ, Ni JJ, He C. 2000. QTLs and epistasis for aluminum tolerance in rice (*Oryza sativa* L.) at different seedling stages. Theor. Appl. Genet. 100:1295-1303.

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

Xu K, Xu X, Ronald PC, Mackill DJ. 2000. A high-resolution linkage map in the vicinity of the rice submergence tolerance locus *Sub1*. Mol. Gen. Genet. 263:681-689.

Xu Y. 2002. Global view of QTL: rice as a model. In: Kang MS, editor. Quantitative genetics, genomics and plant breeding. Wallingford (UK): CAB International. p 109-134.

Yano M, Sasaki T. 1997. Genetic and molecular dissection of quantitative traits in rice. Plant Mol. Biol. 35:145-153.
Young ND, Tanksley SD. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theor. Appl. Genet. 77:95-101.

Notes

Authors' address: International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines.

Isolation of a QTL gene controlling grain number and QTL pyramiding to combine loci for grain number and plant height in rice

M. Ashikari, S. Lin, T. Yamamoto, T. Takashi, A. Nishimura, E.R. Angeles, Q. Qian, H. Kitano, and M. Matsuoka

Many agronomically important traits, including yield, are expressed in continuous phenotypic variation. These complex traits usually are governed by a number of genes known as quantitative trait loci (QTLs) derived from natural variations. Now, QTL analysis has been employed as a powerful approach to discover agronomically useful genes. Grain number and plant height are important traits that directly contribute to grain productivity. We aimed to identify genes of QTLs for grain number and plant height, not only to elucidate molecular mechanisms that regulate grain productivity but also to use these genes for breeding. We first identified that a QTL that increases grain productivity in rice, *Gn1a*, is a gene for cytokinin oxidase/dehydrogenase (OsCKX2), an enzyme that degrades the phytohormone cytokinin. Reduced expression of OsCKX2 causes cytokinin accumulation in inflorescence meristems and increases the number of reproductive organs, resulting in enhanced grain yield. QTL pyramiding to combine loci for grain number and plant height in the same genetic background generated lines exhibiting both beneficial traits. These results provide a strategy for tailor-made crop improvement. Discovering useful genes, improving agricultural traits hidden in the plant genome, and applying these findings to crop breeding will pave the way for a new green revolution.

Keywords: cytokinin oxidase, grain number, plant height, QTL, QTL pyramiding

Grain number and plant height are important traits that directly contribute to grain productivity. In the Green Revolution in rice and wheat in the 1960s, the introduction of a dwarf phenotype by classical breeding resulted in higher yield by avoiding lodging with increased application of nitrogenous fertilizers (Peng et al 1999, Sasaki et al 2002, Ashikari et al 2002). During the last decade, many attempts were made to characterize QTLs for grain production and plant height. However, the genes involved in these QTLs have not been identified yet, and their chromosomal positions remain obscure. We aimed to identify genes of QTLs for grain number and plant height not

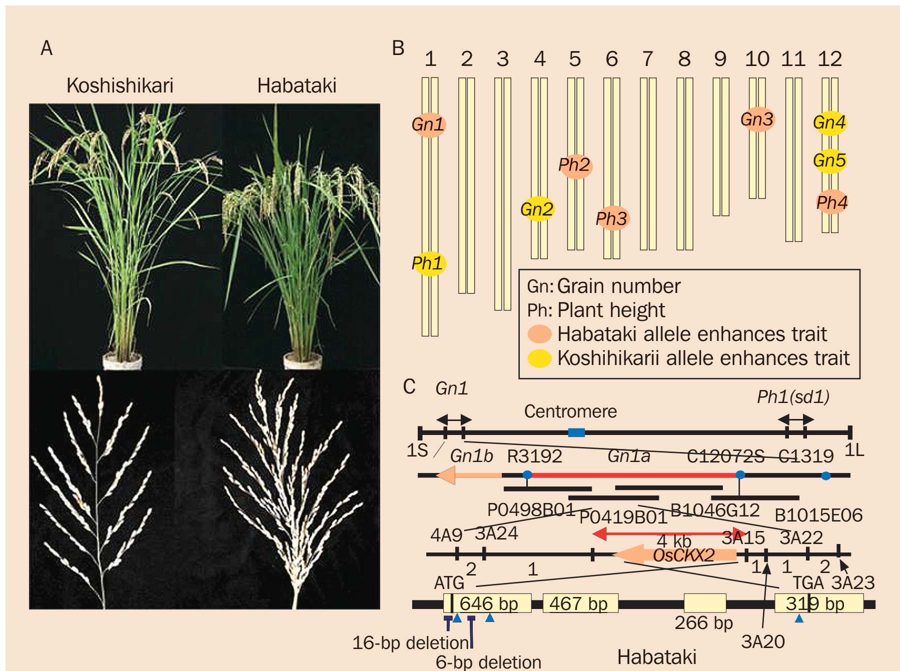


Fig. 1. QTL analysis and molecular cloning. (A) Gross morphology and panicle structure of Koshihikari and Habataki. (B) QTL map of grain number (Gn) and plant height (Ph) on rice chromosomes. (C) Location of *Gn1* and *Ph1* genes on chromosome 1 (the top line) and *OsCKX2* structure (the bottom line). Linkage map and physical map of *Gn1* (the second line from top). The number of recombinants between *Gn1a* and molecular markers is indicated below the high-resolution map (the third line from top). The mutation sites in Habataki are marked in blue. Blue arrowheads indicate the site of amino acid substitutions. (Adapted from Ashikari et al 2005.)

only to elucidate molecular mechanisms that regulate grain productivity but also to use these genes in rice molecular breeding.

QTL analysis

A choice of parental lines that show wide phenotypic variations in the targeted traits is necessary for QTL analysis because QTL detection is based on allelic differences between parental lines. An indica rice variety, Habataki, and a japonica variety, Koshihikari, were chosen in this study, since they not only exhibit large variations of agronomically important traits but also show high frequencies of DNA polymorphism (data not shown). On average, Habataki plants are significantly shorter than individuals of Koshihikari, but they produce more grains in their main panicle (Fig. 1A).

Primary-mapping populations of 96 backcross inbred lines (BILs) derived from the cross between Habataki and Koshihikari were developed. Both grain number and plant height seemed to be regulated by QTLs as these traits were approximately normally distributed in the mapping population. QTL analysis detected five QTLs for increasing grain number (*Gn*) and four QTLs for plant height (*Ph*; Fig. 1B). The most effective QTL for plant height, *Ph1*, was located close to the *semidwarf 1* gene (*sd1*), which encodes gibberellin 20 oxidase (Ashikari et al 2002). Comparison of *sd1* between Habataki and Koshihikari revealed that Habataki had a 383-bp deletion in the coding region, as is the case with the gibberellin 20 oxidase gene in Dee-geo-woo-gen and IR8. The resulting loss of function caused the reduced plant height in Habataki.

The most effective QTL for grain number, *Gn1* on chromosome 1, was selected for further analysis. A line carrying the Habataki *Gn1* is expected to produce about 92 grains more per main panicle than Koshihikari; *Gn1* explains 44% of the difference in grain number between Habataki and Koshihikari. So far, several QTLs associated with grain number have been reported in rice (Xiao et al 1996) and some of them are located near the *Gn1* region identified in our study. Therefore, it is possible that this *Gn1* locus contributes to increasing the crop yield in various rice varieties, although these genes for QTLs have not been identified and characterized yet.

QTL cloning

Ninety-six F_2 individuals derived from heterozygous (*Gn1/gn1*) plants of the near-isogenic line (NIL) *Gn1* were used for mapping of *Gn1*. We found that *Gn1* consisted of two loci, QTL-*Gn1a* and QTL-*Gn1b*. QTL-*Gn1a* was mapped within 2 cM between the molecular markers R3192 and C12072S, whereas QTL-*Gn1b* was mapped to the upper region of QTL-*Gn1a* (Fig. 1C). *Gn1a* was chosen as the target for positional cloning because the effects of *Gn1a* and *Gn1b* were almost identical and because the position of *Gn1a* between the two markers had been unambiguously determined. The *Gn1a* allele of Habataki was semidominant because the grain number of heterozygous plants (*Gn1a/gn1a*) was intermediate between those of homozygous plants (*gn1a/gn1a* and *Gn1a/Gn1a*). About 13,000 F_2 plants derived from heterozygotes (*Gn1a/gn1a*) were used for high-resolution mapping of *Gn1a*. The candidate region of *Gn1a* was narrowed down to the 4 kb between the markers 3A28 and 3A15 (Fig. 1C). In this region, the Rice GAAS (Rice Genome Automated Annotation System: http://RiceGAAS.dna.affrc.go.jp/) predicted one reading frame with high similarity to the cytokinin oxidase/dehydrogenase (CKX) previously designated *OsCKX2* (Fig. 1C, Schmülling et al 2003). The *OsCKX2* of Koshihikari and Habataki consists of four exons and three introns, and encodes proteins of 565 or 563 amino acids, respectively. Comparison of the DNA sequences between the cultivars revealed several nucleotide changes, including a 16-bp deletion in the 5′-untranslated region, a 6-bp deletion in the first exon, and three nucleotide changes resulting in amino acid variation in the first and fourth exon of the Habataki allele (Fig. 1C).

We also analyzed the *OsCKX2* sequence of three alleles of high-yielding varieties from China. An 11-bp deletion in the coding region was detected in one variety,

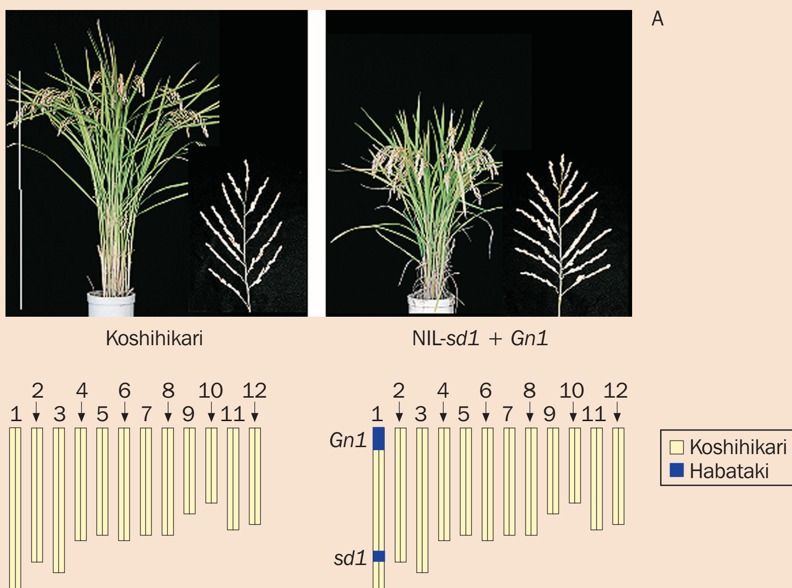
5150, which produced more than 400 grains in the main panicle in our experimental field. This deletion created a premature stop codon, suggesting that 5150 is null for *OsCKX2*. The other two varieties had sequences identical to those of the Habataki allele. The coincidence of the *OsCKX2* null allele and higher grain number suggested that a reduction in or loss of function of *OsCKX2* enhanced grain production. To confirm that *OsCKX2* corresponds to *Gn1a*, we produced transgenic plants expressing different levels of *OsCKX2* and examined their grain yield. As Koshihikari and Habataki fail to regenerate shoots from callus, the readily regenerating cultivar Taichung 65 (TC65), which possesses the Koshihikari allele of *OsCKX2*, was used. Transgenic plants carrying two copies of the sense strand of *OsCKX2* showed a lower grain number than TC65. On the contrary, transgenic plants with antisense strands had a higher grain number (data not shown). We concluded that the QTL for increased grain number, *Gn1a*, is identical to *OsCKX2*.

QTL pyramiding

Since the Habataki alleles of *Gn1a* and *Gn1b* increase grain number, they were targeted in a breeding program. Through the analysis of *Gn1*, we produced several NILs of Koshihikari, the leading variety in Japan, which carried the *Gn1* locus of Habataki and showed increased crop yield. However, because of the heavier panicle, these lines lodged frequently. To solve the problem, we employed a QTL pyramiding breeding strategy. In this approach, desirable QTLs are combined through crossing of NIL QTLs in a common genomic background. First, we developed a NIL carrying the Habataki *sd1* allele in the Koshihikari background (Fig. 2A). This NIL-*sd1* was about 20% shorter than Koshihikari, as expected because of the effect of the *sd1* allele (Fig. 2A and B). Simultaneously, a NIL-*Gn1* carrying the Habataki *Gn1a*+*Gn1b* chromosome fragment that produced about 45% more grains than Koshihikari (Fig. 2A and C) was also selected. The extent of increase in grain number (45%) and reduction in plant height (20%) in the NILs corresponded to the phenotypic variation effects of *Gn1* (44%) or *Ph1* (30%) as predicted by the QTL analysis. NIL-*Gn1*+*sd1* was generated by crossing NIL-*Gn1* and NIL-*sd1* (Fig. 2A). The grain number in the main panicle was 26% higher and plants were 18% shorter in this line than in Koshihikari (Fig. 2A and B). The reduction in grain number in NIL-*Gn1*+*sd1* (207 grains) compared with NIL-*Gn1* (237 grains) seemed to be caused by pleiotropic effects of the *sd1* allele. The same degree of grain number reduction was also found in NIL-*sd1* relative to Koshihikari (Fig. 2C). This negative effect of *sd1*, however, was outweighed in NIL-*Gn1*+*sd1* by the 45% increase in grain number induced by *Gn1* (Fig. 2C). When grain number per plant was compared instead of on a per main panicle basis, similar results were found (Fig. 2D).

Toward an application of QTLs for a new green revolution

We succeeded in cloning a QTL (*Gn1a*) that increased the grain number in rice *Gn1a* that encodes CKX, an enzyme that degrades bioactive CK. Actually, a null allele of



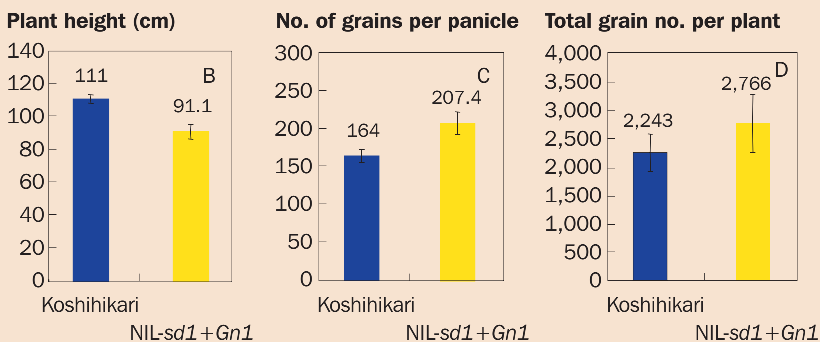


Fig. 2. Phenotypic characterization of NIL QTLs. (A) Plant morphologies and chromosome maps of Koshihikari and NIL-*sd1+Gn1*. Comparison of plant height (B), grain number in the main panicle (C), and grain number in whole plants (D) for Koshihikari and NIL-*sd1+Gn1*. Values in (B) to (D) are means with standard deviation (n = 10). (Adapted from Ashikari et al 2005.)

OsCKX2 had been selected for increasing crop yield in a conventional breeding program in China without molecular characterization. The syntenic relationship of the genomes of rice and other cereals such as maize, barley, and wheat (Liu et al 2002) should allow researchers to use orthologous CKX genes in other species to regulate grain number as in rice.

In molecular studies on the Green Revolution in rice and wheat, the phytohormone gibberellin was identified as a key player in controlling crop morphological architecture. Our study demonstrates that CK metabolism plays a similar role. Since CK controls cell division and lateral meristem activity (Mok and Mok 2001), CK accumulation in the inflorescence meristem can explain the effects described above.

Our study demonstrates that the identification of agronomically important QTLs and the pyramiding of these QTLs is a powerful strategy for efficient crop breeding. Using interspecific crosses between *Oryza sativa* and wild relatives could lead to the discovery of useful QTLs from a range of allelic variations much wider than that present in cultivated lines (Xiao et al 1996). Furthermore, wild rice species are likely to provide an experimental access to QTLs not only for yield but also for disease resistance and stress tolerance because these plants have adapted to unique geographic and environmental conditions. So far, QTL studies have been used for the prediction of chromosome location. However, fine-mapping technologies using molecular markers have enabled us to identify target QTLs while excluding linked but undesirable genes. We can now design genome arrangements for "tailor-made breeding" by marker-assisted selection. We face a global food shortage. FAO estimates that 852 million people worldwide were undernourished in 2000-02. The global population, now 6.4 billion, is still growing rapidly—it will reach 8.9 billion by 2050. Discovering useful genes that improve agricultural traits hidden in the plant genome and their application to crop breeding are essential to avoid the impending food crisis. This molecular breeding strategy will pave the way for a new green revolution.

References

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles E, Qian Q, Kitano H, Matsuoka M. 2005. Cytokinin oxidase regulates rice grain production. Science 309:741-745.

Ashikari M, Sasaki A, Ueguchi-Tanaka M, Itoh H, Nishimura A, Datta S, Ishiyama K, Saito T, Kobayashi M, Khush GS, et al. 2002. Loss-of-function of a rice gibberellin biosynthetic gene, *GA20 oxidase (GA20ox-2)*, led to the rice 'green revolution.' Breed. Sci. 52:143-150.

Liu JP, Van Eck J, Cong B, Tanksley SD. 2002. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc. Natl. Acad. Sci. USA 99:13302-13306.

Mok DWS, Mok MC. 2001. Cytokinin metabolism and action. Annu. Rev. Plant Physiol. Plant Mol. Biol. 52:89-118.

Peng J, Richards D, Hartley N, Murphy G, Devos K, Flintham J, Beales J, Fish L, Worland A, Pelica F, et al. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256-261.

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Datta S, Ishiyama K, Saito T, Kobayashi M, Khush G, Kitano H, Matsuoka M. 2002. Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416:701-702.

Schmülling T, Werner T, Riefler M, Krupkova E, Manns IBY. 2003. Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, *Arabidopsis* and other species. J. Plant Res. 116:241-252.

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

Notes

Authors' addresses: M. Ashikari, H. Kitano, and M. Matsuoka, Bioscience and Biotechnology Center, Nagoya University, Nagoya 464-8601, Japan; S. Lin, T. Yamamoto, T. Takashi, A. Nishimura, and E.R. Angeles, Honda Research Institute Japan Co., Ltd., Kisarazu, Chiba, Japan; Q. Qian, China National Rice Research Institute, Hangzhou, 310006, China.

Genetic and molecular dissection of flowering time in rice

M. Yano and T. Izawa

Flowering time (heading date) is a major determinant of regional and seasonal adaptation of cultivated rice. A large amount of variation is observed in heading date and photoperiodic response among rice cultivars and strains, including wild relatives. Quantitative trait locus (QTL) analyses of progeny derived from several cross combinations of rice cultivars suggest that more than 15 loci are involved in heading date. Map-based cloning has been performed on several QTLs for photoperiodic response. We have demonstrated that *Heading date 1* (*Hd1)* is an ortholog of *CONSTANS* (*CO*) in *Arabidopsis* and is involved in the promotion of heading under short-day (SD) conditions and inhibition under long-day (LD) conditions. *Hd6* is involved in inhibition under LD conditions and encodes the alpha-subunit of protein kinase CK2. *Hd3a* shows a high level of similarity to *Arabidopsis FT* (flowering time) and functions as a flowering inducer. *Early heading date 1* (*Ehd1*) is involved in promotion under SD conditions and encodes a B-type response regulator. *Hd5* is involved in inhibition under LD conditions and encodes a putative subunit of a CCAAT-box-binding protein. *Late heading date 4* (*Lhd4*) is involved in inhibition under LD conditions and encodes a protein with a CCT motif. The combining of information from genetic and sequencing analyses reveals that the combination of natural alleles with loss or gain of function at particular QTLs, such as *Hd1, Hd5, Hd6, Ehd1,* and *Lhd4*, seems to generate a wide range of continuous variation in photoperiodic flowering in rice. These genetic and molecular analyses have allowed us to propose a pathway for the genetic control of photoperiodic flowering in rice, and analysis of the mRNA levels of genes in near-isogenic lines has clearly revealed their hierarchical relationship in the genetic control pathway. Identification and expression analyses of genes suggest the conservation and divergence of various features in the photoperiodic control of flowering in rice, an SD plant, and *Arabidopsis*, an LD plant.

Keywords: *Oryza sativa* L., heading date, natural variation, QTL, map-based cloning, genetic control, pathway

Transition of the apical meristem from vegetative to reproductive growth is a critical event in a plant's life cycle, as it must occur for the plant to produce progeny. In rice, the timing of this transition affects the timing of flowering (often referred to as heading date). Heading date is a key determinant for the adaptation of plants to different cultivation areas and cropping seasons. Therefore, control of heading date is an important objective in rice breeding.

Rice is a short-day (SD) plant; its heading is promoted by short daylength. Response to daylength (referred to as photoperiod sensitivity, PS) and duration of basic vegetative growth determine heading date in rice. Several genes controlling PS in rice have been genetically identified (summarized in Kinoshita 1998). However, the detailed mechanism of the genetic and molecular control of heading date remains to be analyzed. In the past decade, by the enhancement of resources derived from rice genome analysis (Sasaki 2003), the genetic and molecular bases of heading date have been very well studied (reviewed by Yano et al 2001, Izawa et al 2003, Hayama and Coupland 2004). In particular, natural variation has become an efficient resource for the genetic and molecular repetition analysis of complex traits in rice (Yano and Sasaki 1997, Yano 2001).

In the last 10 years, quantitative trait locus (QTL) analyses have been conducted on heading date using several kinds of genetic mapping populations, and four QTLs have been identified at the molecular level (Yano et al 2000, Takahashi et al 2001, Kojima et al 2002, Doi et al 2004). Cloning of additional QTLs is now progressing. These analyses have revealed both conserved and divergent features in the genetic control mechanisms of rice, an SD plant, and *Arabidopsis*, a long-day (LD) plant (Yano et al 2001, Hayama et al 2003, Izawa et al 2003, Hayama and Coupland 2004, Doi et al 2004).

We summarize the current status of analyses of natural variation in flowering time in rice and what we have learned about the genetic and molecular mechanisms controlling heading in rice.

Genetic analysis of genes controlling heading date

Varietal variations in heading date and daylength response

A wide range of variation in heading date and daylength response has been observed among rice varieties. Variations in PS and heading date of six cultivars used in genetic analysis are shown in Figure 1, together with the geographic origin of each cultivar. We observed a wide range of differences in days to heading and degree of PS. The degree of PS varied among varieties adapted to Japanese conditions: for example, Hayamasari, which is adapted to the climate of Hokkaido, the northernmost part of Japan, showed almost the same number of days to heading (DTH) under both SD and LD conditions, exhibiting no PS. On the other hand, Koshihikari had a weak PS and Nipponbare showed a relatively strong PS, even though both cultivars were adapted in almost the same area in Japan. Taichung 65, which originated in Taiwan, was characterized by a large number of DTH under SD conditions and a very weak PS. On the other hand, indica cultivar Nona Bokra from India exhibited an extremely

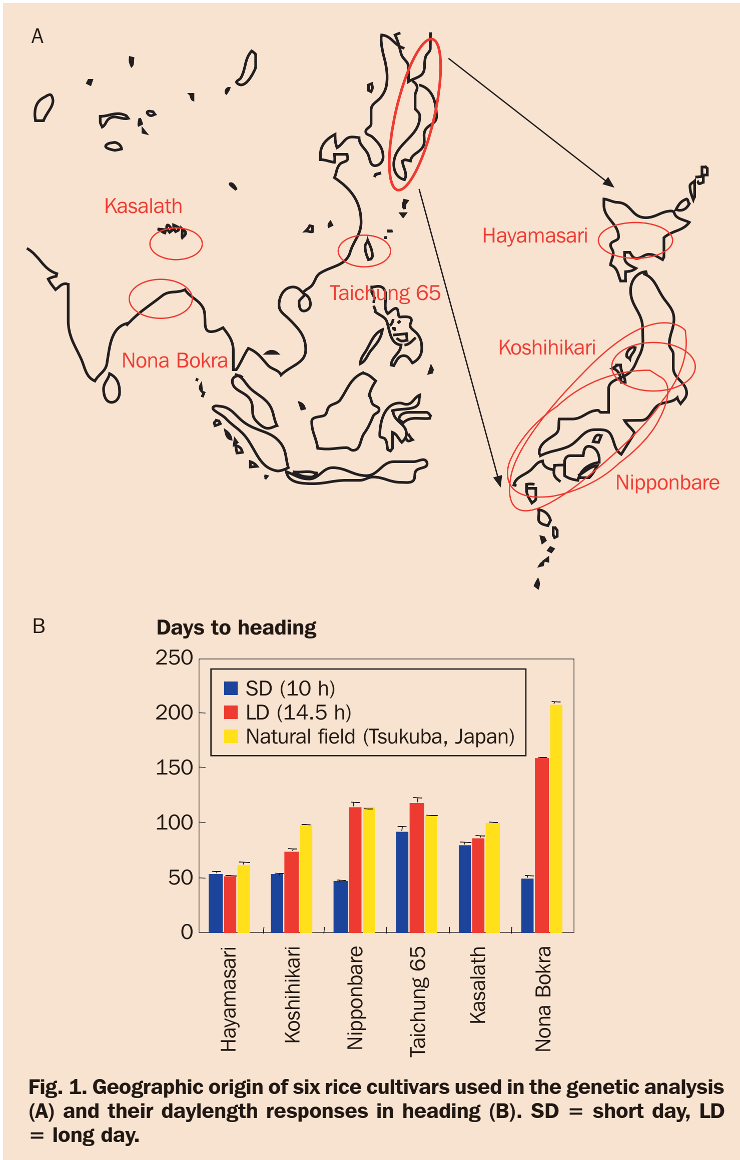


Fig. 1. Geographic origin of six rice cultivars used in the genetic analysis (A) and their daylength responses in heading (B). SD = short day, LD = long day.

strong PS. Although the DTH of Nona Bokra were almost the same as those of the extremely early-heading cultivar Hayamasari under SD conditions, no heading was observed in Nona Bokra under LD (14.5 h light) conditions. These observations clearly demonstrated that a wide range of naturally occurring variation exists in the genetic factors controlling heading date in rice.

Identification of QTLs for heading date

Genetic analysis of heading date in mutants and natural variants has so far allowed the identification of several genes involved in PS (Yokoo et al 1980, Yamagata et al 1986, Yokoo and Okuno 1993, Okumoto and Tanisaka 1997, Ichitani et al 1998). A series of tester lines for several PS genes has been developed to facilitate genetic analysis of flowering time in rice (Yamagata et al 1986). Nishida et al (2001) proposed a putative genetic model for control of heading in rice, based on genetic and physiological analyses of these tester lines. These analyses have contributed to our understanding of the genetic control of heading date in rice. However, the quantitative nature of the inheritance of heading date has hindered the performance of more detailed analyses, including the analysis of epistatic interactions of genes and the accurate determination of the chromosomal locations of these genes.

In the last decade, progress in the development of DNA markers has made it possible to conduct QTL analysis to clarify which genes control heading date in rice (Yano and Sasaki 1997). In our laboratory, we have been conducting QTL analyses of heading date using progeny derived from crosses among the six cultivars mentioned above. In particular, comprehensive QTL detection has been performed by using several types of progeny derived from a cross between Nipponbare (japonica) and Kasalath (indica). Five QTLs, *Heading date Hd 1–Hd5,* have been mapped on the basis of analysis of the F_2 population (Yano et al 1997), and an additional three QTLs, *Hd7, Hd8*, and *Hd11*, have been detected by using BC_1F_5 lines (Lin et al 1998, Takeuchi et al 2003). In addition, other loci—*Hd6*, *Hd9, Hd10*, *Hd12*, *Hd13*, and *Hd14*—have been detected only 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, Lin et al 2002, 2003; for review, see Yano et al 2001). In total, by using single cross combinations, we have identified 15 QTLs controlling heading date.

The detection of putative QTLs for heading date allowed us to embark on their further genetic analysis. The development of near-isogenic lines (NILs) by marker-assisted selection, in which a small chromosomal segment including the detected QTL of the donor cultivar Kasalath was substituted into the Nipponbare genetic background, has brought many advantages to the genetic analysis of heading date in rice (for review, see Yano and Sasaki 1997). For example, QTL NILs can be used in the characterization of the level of PS, in epistatic interaction analysis, and in fine genetic linkage mapping of target QTLs. Eight QTLs—*Hd1, Hd2, Hd3, Hd4, Hd5, Hd6, Hd8*, and *Hd9*—were mapped as single Mendelian factors, and five of them—*Hd1*, *Hd2*, *Hd3*, *Hd5*, and *Hd6*—were found to confer PS (Lin et al 2000, 2002, 2003, Yamamoto et al 2000). Recently, QTLs involved in extremely early and late heading have been identified (Nonoue, Uga, and Yano, unpublished data). One QTL involved

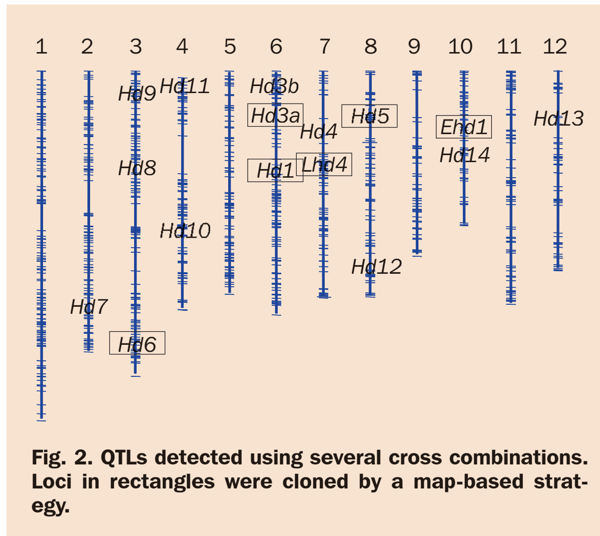


Fig. 2. QTLs detected using several cross combinations. Loci in rectangles were cloned by a map-based strategy.

in extremely early heading was detected on chromosome 7 and one on chromosome 8, and their chromosomal locations were coincident with those of *Hd4* and *Hd5*. On the other hand, six QTLs involved in extremely late heading were detected by QTL analysis using an F_2 population derived from a cross between the early-heading cultivar Koshihikari and the extremely late-heading cultivar Nona Bokra. The chromosomal locations of the QTLs detected were coincident with those of *Hd1*, *Hd2*, *Hd3a*, *Hd4*, *Hd5*, and *Hd6*. Although only six cultivars were used in the genetic analysis, these results suggest that most of the loci involved in natural variation in heading date are likely to be identifiable genetically.

High-resolution mapping also enabled us to dissect two tightly linked loci, *Hd3a* and *Hd3b*, in the *Hd3* region (Fig. 2) (Monna et al 2002). Analysis of photoperiodic response in NILs of *Hd3a* and *Hd3b* revealed that the Kasalath allele of *Hd3a* promotes flowering under SD conditions, and that the Kasalath allele of *Hd3b* delays flowering under LD and natural field conditions (Monna et al 2002). Together, these results clearly demonstrate that the genetic control mechanisms of flowering in rice can be dissected into individual components by a series of genetic analyses of flowering date based on QTL analysis.

QTL NILs can also be used in the analysis of epistatic interactions between QTLs. Genetic analysis using QTL NILs has been used to clarify the existence of an epistatic interaction between *Hd1* and *Hd3* (Lin et al 2000). It has also been suggested that the Kasalath allele of *Hd3* itself does not affect photoperiod sensitivity, but that it is involved in enhancement of the expression of the Nipponbare alleles of the PS QTLs *Hd1* and *Hd2* (Lin et al 2000). In addition, epistatic interaction between *Hd2* and *Hd6* has been clearly detected in analysis of advanced progeny. The effect of the Kasalath allele of *Hd6* was observed only in the presence of the Nipponbare allele of

Table 1. QTLs controlling heading date in rice.

QTLs	Daylength response	Structure	Biochemical function	Ortholog in *Arabidopsis*	Reference
Hd1	Promotion—SD Inhibition—LD	B-box (Zn finger domain) CCT motif	TF	*CO*	Yano et al (2000)
Hd3	Promotion—SD	?	?	*FT*	Kojima et al (2002)
Hd6	Inhibition—LD	Protein kinase CK2a	Phosphorylation	*CK2a*	Takahashi et al (2001)
Ehd1	Promotion—SD	B-type response regulator	TF	None	Doi et al (2004)
Lhd4	Inhibition—LD	CCT motif	?	None	Unpublished
Hd5	Inhibition—LD	CCAAT-box-binding protein subunit	TF	Unknown	Unpublished

Hd2 (Yamamoto et al 2000). The effect of the Kasalath allele of *Hd5* was observed only in the presence of the Nipponbare allele of *Hd1* (Lin et al 2003).

As mentioned above, marker-assisted genetic analyses have clearly contributed to our understanding of the genetics of heading date and PS. In addition, these works have allowed us to embark on the molecular analysis of heading-date genes in rice.

Molecular analysis of genes involved in heading date

Map-based cloning of QTLs

The development of NILs of the QTLs for heading date allowed us to embark on map-based cloning of genes identified genetically at the QTLs (Fig. 2 and Table 1).

A major QTL, *Hd1*, controlling response to photoperiod, was identified by means of a map-based cloning strategy (Yano et al 2000). *Hd1* was detected on the proximal region of chromosome 6 by QTL analysis using an F_2 population derived from a cross between Nipponbare and Kasalath (Yano et al 1997). Advanced back-cross progeny allowed us to map *Hd1* as a single Mendelian factor (Yamamoto et al 1998). High-resolution mapping using 1,505 F_2 plants enabled us to define a genomic region of about 12 kb as a candidate for *Hd1*. Further analysis revealed that the *Hd1* QTL corresponds to a gene that is a homolog of *CONSTANS* (*CO*) in *Arabidopsis* (Putterill et al 1995, Yano et al 2000). *Se1* was allelic to the *Hd1* QTL, as determined by analysis of two *se1* mutants, *HS66* and *HS110*. Genetic complementation analysis proved the function of the candidate gene. The level of *Hd1* mRNA was not greatly affected by changes in daylength. It is interesting to note that a 36-bp deletion occurred in the Nipponbare allele of *Hd1* compared with the wild-type Ginbouzu allele, suggesting that the Nipponbare *Hd1* may be leaky in terms of function. *Arabidopsis CO* promotes flowering under LD conditions, whereas *Hd1* has a dual function under

different daylength conditions: *Hd1* is involved in promotion of heading under SD conditions but inhibition under LD conditions.

Hd6 was detected as a QTL on the long arm of chromosome 3 by using backcross progeny derived from a cross between Nipponbare and Kasalath, and was found to be involved in PS (Yamamoto et al 2000). High-resolution linkage analysis of *Hd6* enabled us to delimit *Hd6* to a 26.4-kb genomic region. In this region, we found one probable candidate gene, an ortholog of *Arabidopsis* CK2a. The function of the candidate gene has been proved by genetic complementation analysis (Takahashi et al 2001). We identified a gene encoding a subunit of protein kinase CK2 (CK2a) in this region. The Nipponbare allele of CK2a contains a premature stop codon, and the resulting truncated product is undoubtedly nonfunctional. This result indicates that CK2 plays an important role in the photoperiod response of flowering in rice.

Hd3a has been detected as a heading-date-related QTL on the short arm of chromosome 6 (Yano et al 1997, Monna et al 2002). A previous study had revealed that the Kasalath allele of *Hd3a* promotes heading under SD conditions (Monna et al 2002). High-resolution linkage mapping located the *Hd3a* locus in an approximately 20-kb genomic region (Kojima et al 2002). In this region, we found a candidate gene that showed high similarity to the *FLOWERING LOCUS T* (*FT*) gene, which promotes flowering in *Arabidopsis* under LD conditions (Kobayashi et al 1999). Introduction of the candidate gene caused an early-heading phenotype in rice. From these results, it was concluded that *Hd3a* encodes a protein with high similarity to the *FT* gene in *Arabidopsis* (Kojima et al 2002).

Early heading date 1 (*Ehd1*) has been detected on chromosome 10 as a major QTL in the analysis of a BC_1F_1 population of Taichung 65 and *Oryza glaberrima* (Doi and Yoshimura 1998). With a map-based strategy, it was revealed that *Ehd1* encodes a B-type response regulator (Doi et al 2004). *Ehd1* promotes heading under SD and LD conditions in the absence of a functional allele of *Hd1*. Based on the BLAST search, there might not be an ortholog of *Ehd1* in the *Arabidopsis* genome. These results indicate that a novel two-component signaling cascade is integrated into the conserved pathway in the photoperiodic control of flowering in rice.

Hd5, a QTL detected on the short arm of chromosome 8, is involved in inhibition of flowering under LD conditions, but no effect has been observed under SD conditions (Lin et al 2003). We performed map-based cloning of *Hd5*. High-resolution linkage analysis of 2,308 F_2 plants defined a genomic region of 4.3 kb as a candidate for *Hd5* and allowed us to focus on one candidate gene, which showed high sequence similarity to a subunit of a CCAAT-box-binding transcription factor (Yamanouchi and Yano, unpublished data). Transgenic Nipponbare plants with a candidate genomic region of Kasalath exhibited late flowering under LD conditions but no differences in flowering under SD conditions. This result clearly demonstrated that a candidate genomic region for *Hd5* suppressed flowering under LD conditions.

Late heading date 4 (*Lhd4*) was detected on chromosome 7 by using a backcross population derived from a cross between the japonica cultivar Hoshinoyume (extremely early heading) and indica cultivar Kasalath (late heading) (Nonoue and Yano, unpublished data). High-resolution mapping revealed one probable candidate

gene for *Lhd4*. The putative protein predicted from the candidate gene contained the CCT (CO, CO-like, and TOC1) motif, known as a conserved region that has been suggested to be related to the regulation of flowering (Kitazawa, Izawa, Monna, and Yano, unpublished data). The function of the candidate gene was proved by genetic complementation analysis.

Cloning of other QTLs

Two additional QTLs, *Hd2* and *Hd9*, have been targeted for map-based cloning. Owing to the chromosomal location of *Hd2* (distal end of the long arm of chromosome 7), it has been very difficult to perform a high-resolution linkage analysis of *Hd2*. Currently, we have identified one probable candidate gene, and complementation analysis is now under way (Yamanouchi, Nakagawa, Izawa, and Yano, unpublished data). For *Hd9*, high-resolution mapping allowed us to define a genomic region of about 20 kb on chromosome 3 (Yamanouchi and Yano, unpublished data). Genetic complementation analysis is now in progress. Recently, we detected two QTLs for heading date in the progeny of two japonica cultivars, Nipponbare and Koshihikari. These two QTLs are not likely to be the same as those previously detected (*Hd1–Hd15*). In Japanese varieties, additional genetic factors might be involved in the generation of variation in heading date and daylength response. Map-based cloning of the two QTLs is in progress.

Naturally occurring allelic variations in genes for PS

Molecular identification of the genes controlling PS has allowed us to clarify the functional nucleotide polymorphisms (FNPs) underlying natural variations in flowering time. In *Hd1*, a 36-bp deletion occurred in the first exon, resulting in partial loss of function. A 2-bp deletion occurred in the Kasalath *Hd1*, and a 1,901-bp insertion was detected in the Taichung 65 *Hd1*, both resulting in complete loss of function (Yano et al 2000, Doi et al 2004). In the case of *Hd3a*, the Nipponbare allele was less functional than that of Kasalath. Comparison of both alleles in terms of expression levels and nucleotide changes revealed that FNP might occur in the intron or 3′ flanking region of *Hd3a* rather than in the exon region (Kojima et al 2002). However, this remains to be clarified. In Nipponbare *Hd6*, a single nucleotide polymorphism (SNP) occurred in the exon, and this generated a premature stop codon (Takahashi et al 2001). The SNP resulted in an amino acid substitution in the GARP domain of Taichung 65 *Ehd1* (Doi et al 2004), resulting in loss of function of *Ehd1* (Doi et al 2004).

Recently, FNPs were found to occur frequently in the regulatory regions, but not the coding regions, of several QTLs (Salvi and Tuberosa 2005). However, in the case of rice heading-date genes, many FNPs that result in loss of function are likely to occur in the coding regions. Together, the results of the QTL and sequencing analyses reveal that naturally occurring, continuous variations in DTH and PS among rice cultivars are likely to have been generated by a combination of gain and loss of function of a series of heading-date genes.

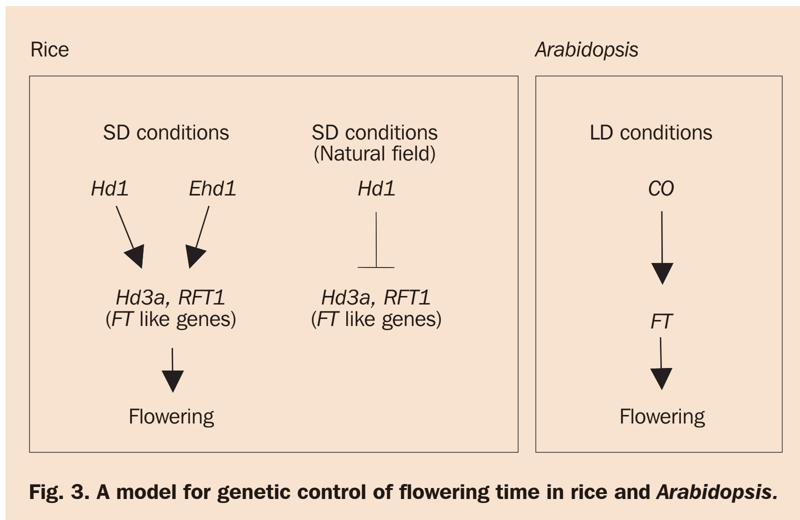


Fig. 3. A model for genetic control of flowering time in rice and *Arabidopsis*.

Model of genetic control of heading in rice

Recent progress in the molecular cloning of heading-date QTLs has allowed us to clarify the molecular mechanism of the genetic control of heading date. We surveyed the mRNA levels of genes identified in several NILs developed by marker-assisted selection. This type of analysis was often helpful to prove the molecular basis of the epistatic interactions among heading-date genes, as observed in genetic analyses.

We have monitored the mRNA levels of both *Hd1* and *Hd3a* in several NILs (Kojima et al 2002). *Hd3a* transcripts were detected early and gradually increased with time under SD conditions. The level of expression of *Hd3a* in a NIL, in which the chromosomal region of *Hd3a* is substituted with Kasalath in the genetic background of Nipponbare, was higher than that in Nipponbare under SD conditions. NIL *Hd3a* showed earlier heading than did Nipponbare under SD conditions. This result indicated that the higher expression levels of the Kasalath *Hd3a* allele resulted in the promotion of heading under SD conditions (Kojima et al 2002). To investigate whether *Hd1* regulates *Hd3a*, we also quantified the expression levels of *Hd3a* in NIL *Hd1*. *Hd1* has been found to be an ortholog of the *Arabidopsis CO*, which is involved in the promotion of flowering and up-regulates the expression of *FT1* under LD conditions. In NIL *Hd1*, the Nipponbare functional allele was replaced with a loss-of-function Kasalath allele. NIL *Hd1* showed later heading than did Nipponbare under SD conditions. The expression levels of *Hd3a* were reduced in NIL *Hd1*, indicating that the functional allele of *Hd1* up-regulates the expression of *Hd3a* (Kojima et al 2002) (Fig. 3). Similar observations have been made by Izawa et al (2002) in the analysis of the *se5* mutant, which shows extremely early and complete loss of PS in rice. The results obtained in both studies (Kojima et al 2002, Izawa et al 2002) suggest that the functions of *Hd3a* and *FT1* and the regulation of their expression by *Hd1* and *CO*, respectively, are

conserved between rice (an SD plant) and *Arabidopsis* (an LD plant). However, these results also demonstrated that there is a difference between rice and *Arabidopsis* in the expression profiles of the key flowering-time genes *Hd3a* and *FT* in response to daylength.

Doi et al (2004) have demonstrated that *Ehd1* was involved in up-regulation of an *FT*-like gene such as *Hd3a*, independently of the action of *Hd1*, resulting in the promotion of heading under SD conditions (Fig. 3). A BLAST search using *Arabidopsis* genome sequence data reveals that there might not be a gene in *Arabidopsis* orthologous to *Ehd1* (Doi et al 2004). Although it has been proved that similar genetic control mechanisms in flowering time (heading date) are involved in rice and *Arabidopsis*, these results clearly demonstrate that a divergent genetic control mechanism could be involved in the flowering of rice.

Currently, we are analyzing the mRNA levels of several PS genes, including *Hd1*, *Hd3a*, *Ehd1*, *Hd5*, and *Lhd4*, identified in several NILs. These analyses will contribute further to our understanding of the gene hierarchy in the mRNA transcriptional network of heading in rice.

Future prospects

To exploit a wide range of allelic variation in the genes controlling flowering in rice, other cultivated species (*O. glaberrima*) and wild relatives, which are adapted to specific environmental conditions, can be used as donor parents to develop mapping populations. It will be important to use novel mapping populations, such as chromosome segment substitution lines (CSSLs), to detect new loci for heading date. In fact, the use of CSSLs in mapping QTLs for heading date has allowed us to detect new QTLs, even in the same cross combinations previously analyzed (Ebitani et al 2005).

Currently, we are investigating the transcriptional gene network for heading date, but this will not be enough to clarify the genetic mechanisms controlling heading date. Post-transcriptional regulation might be involved in the control of heading date. Interaction at the protein level needs to be analyzed if we are to develop a comprehensive understanding of the mechanism.

Several genes have been identified through molecular analysis in rice. Although the biochemical functions of *Arabidopsis CO* and *FT* seem to be conserved in rice *Hd1* and *Hd3a*, respectively, the inductive photoperiod for flowering differs between rice and *Arabidopsis*. This raises a simple question: What kind of genes or mechanisms are involved in generating the completely opposite reaction to the photoperiod between SD and LD plants? The analysis of newly identified QTLs, such as *Ehd1*, *Hd5*, and *Lhd4*, may provide a trigger for our understanding of such mechanisms. Further comparative studies between *Arabidopsis* and rice will allow us to clarify the conserved and/or divergent features in the important and complex developmental system of flowering.

References

Doi K, Yoshimura A. 1998. RFLP mapping and QTL analysis of heading date and pollen sterility using backcross population between *Oryza sativa* L. and *Oryza glaberrima* Steud. Breed. Sci. 48:395-399.

Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A. 2004. *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT*-like gene expression independently of *Hd1*. Genes Dev. 18:926-936.

Ebitani T, Takeuchi Y, Nonoue Y, Yamamoto T, Takeuchi K, Yano M. 2005. Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of indica rice cultivar 'Kasalath' in a genetic background of japonica elite cultivar 'Koshihikari'. Breed. Sci. 55:65-73.

Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K. 2003. Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 422:719-722.

Hayama R, Coupland G. 2004. The molecular basis of diversity in the photoperiodic flowering responses of *Arabidopsis* and rice. Plant Physiol. 135(2):677-684.

Ichitani K, Okumoto Y, Tanisaka T. 1998. Genetic analyses of low photoperiod sensitivity of rice cultivars from the northernmost regions of Japan. Plant Breed. 117:543-547.

Izawa T, Oikawa T, Sugiyama N, Tanisaka T, Yano M, Shimamoto K. 2002. Phytochrome mediates the external light signal to repress *FT* orthologs in photoperiodic flowering of rice. Genes Dev. 16:2006-2020.

Izawa T, Takahashi Y, Yano M. 2003. Comparative biology comes to bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Curr. Opin. Plant Biol. 6:113-120.

Kinoshita T. 1998. Report of the committee on gene symbolization, nomenclature and linkage groups. II. Linkage mapping using mutant genes in rice. Rice Genet. Newsl. 15:13-74.

Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T. 1999. A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960-1962.

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M. 2002. *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol. 43:1096-1105.

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.

Lin HX, Ashikari M, Yamanouchi U, Sasaki T, Yano M. 2002. Identification and characterization of a quantitative trait locus, *Hd9*, controlling heading date in rice. Breed. Sci. 52:35-41.

Lin HX, Liang ZW, Sasaki T, Yano M. 2003. Fine mapping and characterization of quantitative trait loci *Hd4* and *Hd5* controlling heading date in rice. Breed. Sci. 53:51-59.

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.

Monna L, Lin HX, Kojima S, Sasaki T, Yano M. 2002. Genetic dissection of a genomic region for a quantitative trait locus, *Hd3*, into two loci, *Hd3a* and *Hd3b*, controlling heading date in rice. Theor. Appl. Genet. 104:772-778.

Nishida H, Okumoto Y, Nakagawa H, Ichitani K, Inoue H, Tanisaka T. 2001. Analysis of tester lines for rice (*Oryza sativa* L.) heading-time genes using reciprocal photoperiodic transfer treatments. Ann. Bot. 88:527-536.

Okumoto Y, Tanisaka T. 1997. Trisomic analysis of a strong photoperiod-sensitivity gene *E1* in rice (*Oryza sativa* L.). Euphytica 95:301-307.

Putterill J, Robson F, Lee K, Simon R, Coupland G. 1995. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80:847-857.

Salvi S, Tuberosa R. 2005. To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci. 10:297-304.

Sasaki T. 2003. Rice genome analysis: understanding the genetic secrets of the rice plant. Breed. Sci. 53:281-289.

Takahashi Y, Shomura A, Sasaki T, Yano M. 2001. *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc. Natl. Acad. Sci. USA 98:7922-7927.

Takeuchi Y, Lin SY, Sasaki T, Yano M. 2003. Fine-scale linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. Theor. Appl. Genet. 107:1174-1180.

Yamagata H, Okumoto Y, Tanisaka T. 1986. Analysis of genes controlling heading time in Japanese rice. In: Rice genetics. Manila (Philippines): International Rice Research Institute. p 351-359.

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 interactions 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 quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell 12:2473-2483.

Yano M. 2001. Genetic and molecular dissection of naturally occurring variation. Curr. Opin. Plant Biol. 4:130-135.

Yano M, Kojima S, Takahashi Y, Lin HX, Sasaki T. 2001. Genetic control of flowering time in rice, a short-day plant. Plant Physiol. 127:1425-1429.

Yokoo M, Okuno K. 1993. Genetic analysis of earliness mutations induced in the rice cultivar Norin 8. Jpn. J. Breed. 43:1-11.

Yokoo M, Kikuchi F, Nakane A, Fujimaki H. 1980. Genetical analysis of heading time by aid of close linkage with blast, *Pyricularia oryzae*, resistance in rice. Bull. Natl. Inst. Agric. Sci. D 31:95-126.

Notes

Authors' address: Molecular Genetics Department, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan, e-mail: myano@nias.affrc.go.jp.
Acknowledgments: This work was supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN) and by the Rice Genome Project (IP-1001) of the Ministry of Agriculture, Forestry, and Fisheries of Japan (MAFF).

Understanding broad-spectrum durable resistance in rice

J.E. Leach, R. Davidson, B. Liu, P. Manosalva, R. Mauleon, G. Carrillo, M. Bruce, J. Stephens, M.G. Diaz, R. Nelson, C. Vera Cruz, and H. Leung

A long-standing goal in rice disease control is to identify and incorporate broad-spectrum durable resistance (BSDR). Although quantitative resistance can potentially contribute to BSDR, neither the genes responsible for quantitative resistance nor the pathways or mechanisms by which they may function to contribute to BSDR are understood. Using varieties that show durable resistance historically, we have identified rice genes that are candidates for contributing to BSDR through co-localization with disease resistance QTLs in mapping studies. Several of these genes are known as disease defense response genes (e.g., oxalate oxidase, chitinase, PR1, etc.), whereas others are of unknown function. Genome-wide expression analyses at critical stages of host-pathogen interactions are also being used to reveal additional genes that may play a role in quantitative resistance. By combining chromosomal segments associated with five different candidate genes by marker-assisted selection, rice lines were produced that exhibited a high level of resistance to rice blast in multilocation trials. The current challenge is to understand if and how these candidate genes contribute to BSDR as well as the allelic variation that accounts for function in some lines but not in others. Targeted gene expression and functional analyses of candidate gene family members, for example, the oxalate oxidase gene families, are being used to focus on gene members involved in BSDR, and to determine what gene structural features are key to involvement. Sequence comparisons are providing clues as to critical allelic variation in rice germplasm. Finally, analysis of mutants exhibiting inappropriate activation of defense pathways is guiding the selection of candidate genes or genic regions. The integration of expression, mapping, and allelic diversity data is expected to unveil genes or gene interactions with significant phenotypic effects that can be used in breeding programs.

Keywords: Rice, broad-spectrum durable resistance, candidate genes, gene expression, QTL

Historically, many breeding programs have relied upon the introduction of genes for qualitative resistance to provide protection against the major crop diseases. Qualitative resistance, which confers specific resistance against some pathogen races, is the easiest to incorporate into breeding programs because it is controlled by one or a few genes with large effects. For many pathogens, this type of resistance is not durable because of changes to virulence in the pathogen population (reviewed in Leach et al 2004). As a result, increasing attention has focused on the accumulation of quantitative disease resistance traits in crop plants. Quantitative resistance is controlled by many genes, usually with small but additive effects, is considered relatively stable (durable), and is possibly effective against many types of pathogens (broad spectrum) (Johnson 1981). As of yet, breeding programs have not widely exploited quantitative traits for resistance, in part because the identity of genes that contribute to quantitative traits is unknown, and therefore specific markers for accumulating the genes are not available (Bonman et al 1992). A better understanding of the role of the particular genes involved will make it feasible to more reliably generate crop genotypes with broad-spectrum durable resistance (BSDR), which will contribute to sustainable crop productivity.

Our goal is to understand the basis of BSDR in rice, focusing on three distinct diseases, bacterial blight caused by *Xanthomonas oryzae* pv. *oryzae*, rice blast caused by *Magnaporthe grisea*, and sheath blight caused by *Rhizoctonia solani*. Genes for qualitative resistance are available for control of bacterial blight and blast, but are not available to control sheath blight (Bonman et al 1992, Leung et al 2003). We hypothesized that defense response (DR) genes, which are genes involved in the disease defense response pathways (Dixon and Harrison 1990, Young 1996), may be involved in quantitative durable resistance. The prediction in a candidate gene approach is that DR genes are potentially involved in the biochemical pathways responsible for the expression of resistance and, therefore, that these genes will be useful markers for quantitative resistance (Leach et al 2004). However, although the candidate gene approach is appealing, since no disease resistance QTL had been cloned, there was no experimental evidence that quantitative resistance was controlled by DR genes or, furthermore, that resistance controlled by genes with small effects was actually more durable (Correa-Victoria and Zeigler 1995). We review the evidence associating candidate genes with disease resistance phenotypes in advanced genetic materials, and progress in a set of collaborative studies aimed at determining whether DR genes are associated with disease resistance QTLs, and determining whether DR genes actually function in disease resistance and confer BSDR.

Genetic materials for dissecting "durable" resistance

BSDR to diverse types of pathogens or to multiple races of a pathogen can be found in natural germplasm and by examining historical breeding records (Leung et al 2003). Rice cultivars such as Sanhuangzhan 2 (SHZ-2), Moroberekan, Oryzica Llanos (OL5), and IR64 carry varying levels of resistance to diverse pathogens, much of which is non-race-specific and is observed to be durable. For example, rice cultivar SHZ-2 was shown to be resistant to a large and diverse collection of isolates of the rice blast

fungus, *M. grisea* (Liu et al 2004), confirming the broad resistance of the cultivar. The inheritance of BSDR has been characterized by using various rice mapping populations created using the lines exhibiting BSDR. These include the Moroberekan (Javanica, exhibits BSDR) × Co39 recombinant inbred (RI) population (Wang et al 1994), a doubled-haploid (DH) population derived from IR64 (indica, exhibits BSDR) × Azucena (japonica, susceptible), and RI and advanced backcross populations, such as those derived from other indica × japonica crosses, SHZ-2 × Lijiangxin-tuan-heigu (LTH) (Liu et al 2004) and Vandana × Moroberekan (Wu et al 2004). These populations allowed the localization of disease and insect resistance QTLs, and, later, association of the regions containing QTLs with candidate genes (Liu et al 2004, Ramalingam et al 2003, Wisser et al 2005, Wu et al 2004).

In addition to the studies mentioned above, in which the natural variation captured in rice cultivars was analyzed, rice lines with induced mutations have been identified that show BSDR to multiple diseases. Through intensive screens of a large (>30,000) collection of chemical- and irradiation-induced mutants in rice cultivar IR64, we have isolated mutants with altered responses to blast and bacterial blight. Two of these mutants show enhanced resistance to blast and blight, and three of them are highly blast-susceptible (Hirochika et al 2004, Leung et al 2001, Wang et al 2004, Wu et al 2005). In addition, more than 20 different lesion mimic mutants have been isolated from the deletion collection; a number of these show enhanced resistance to blast and to blight (C.J. Wu et al, unpublished; Leach et al 2004). These mutants are vehicles for the identification of genes contributing to BSDR.

What are the candidate genes?

Two classes of genes were considered as candidate genes in our published studies. The first of these are genes involved in the recognition process (R or resistance or major genes) and the second class are those involved in the defense response process (DR or minor genes, Dixon and Harrison 1990). A third class of genes, those involved in signal transduction and/or transcriptional regulation of defense response genes, was not included in the initial studies; however, these genes may emerge as targets as the molecular pathways to disease resistance are dissected.

A large number of R genes, which are involved in recognition of pathogen avirulence genes, have been cloned (Bent 1996, Hulbert 2002). These genes share common sequence motifs, reflecting related functions in signal transduction pathways (for review, see Hulbert et al 2001). Common motifs include those encoding a protein kinase (e.g., *Pto* and *Xa21*) and gene products with a leucine-rich repeat (LRR). Other conserved features include the nucleotide binding site (NBS) and coiled coil or TIR domains. Although these recognition-type resistance genes are most commonly associated with qualitative resistance, evidence suggests that some of them are also associated with quantitative resistance (Wang et al 1994). Many of these loci are complex, and carry multiple homologous genes. Further, mapping studies have shown that some putative QTLs are localized in chromosomal regions that harbor these R genes (Wang et al 1994), although it has not yet been determined whether the QTLs

are members of the R-gene family. The R genes or R-gene analogs used in our studies to determine co-localization with QTLs were NBS-LRR domains obtained by PCR amplification and cloning (Bai et al 2002).

The second class of genes contains those involved in the defense response process (DR genes, Dixon and Harrison 1990, Young 1996). DR genes are frequently recognized based on their increased expression during plant defense. The proteins encoded by these genes include (1) structural proteins that are incorporated into the extracellular matrix and participate in the confinement of the pathogen; (2) enzymes of secondary metabolism, for instance, members of the phenypropanoid pathway important in the synthesis of isoflavonoid and stilbene phytoalexins and lignin; and (3) enzymes that are implicated to be directly involved in the defense response, including chitinases, peroxidases, catalases, glucanases, sulfotransferases, and proteins that inactivate fungal ribosomes and bind chitin (Stintzi et al 1993). Overexpression of some DR genes in transgenic plants has been shown to increase plant resistance to pathogens (Epple et al 1997, Grison et al 1996, Livingstone et al 2005, Wu et al 1995, Zhu et al 1994), and suppression of expression of others by antisense technology has been shown to increase disease susceptibility (Maher et al 1994). The DR genes used in these studies were identified from (1) database and literature searches (Ramalingam et al 2003), (2) suppression subtractive hybridization (SSH, Diatchenko et al 1996) libraries of rice with the *Xa7* or *Xa10* genes for bacterial blight resistance that were undergoing compatible or incompatible interactions with *X. oryzae* pv. *oryzae* (unpublished results), and, more recently, (3) differential gene expression in microarray hybridization experiments (Liu et al, unpublished).

Are candidate DR genes associated with disease QTLs?

Using different rice mapping populations, an association was demonstrated for a number of candidate DR genes and quantitative traits for disease and pest resistance (Liu et al 2004, Ramalingam et al 2003, Wisser et al 2005, Wu et al 2004). In the first study, Ramalingam et al (2003) demonstrated that candidate DR genes co-localized with QTLs in rice for resistance to blast, bacterial blight, sheath blight, and brown planthopper. For mapping, a DH population derived from two highly diverged parents (IR64 and Azucena) that exhibit contrasting phenotypes for a variety of agronomic traits was used (Guiderdoni et al 1992). Several candidate RGA and DR gene markers were associated with QTLs for bacterial blight, blast, sheath blight, and brown planthopper (Ramalingam et al 2003). For example, the markers dihydrofolate reductase thymidylate, aldose reductase, oxalate oxidase, JAMyb, and peroxidase were significantly associated with QTLs for one or more pathogens or pests.

In a second study, an association of candidate genes was observed with blast resistance QTLs in an RI population of 215 lines derived from a blast-resistant indica cultivar (SHZ-2) crossed with a susceptible japonica cultivar (LTH) (Liu et al 2004). In particular, regions harboring five putative DR genes—oxalate oxidase, dehydrin, PR1, chitinase, and 14-3-3 protein genes—co-localized with blast resistance QTLs, and together accounted for 60% of the variation in quantitative disease resistance.

The contribution of the regions containing the candidate DR genes to disease resistance was determined by estimating the reduction in disease attributed to specific alleles of candidate genes across diverse environments (Liu et al 2004). RI lines with alleles of SHZ-2 (resistant) at the oxalate oxidase, dehydrin, and PR-1 loci showed significantly less disease (measured as diseased leaf area, DLA) in all test locations than lines with the LTH alleles. The SHZ-2 alleles of chitinase and 14-3-3 protein showed significant effects in some but not all test conditions. The most intriguing finding in these experiments was that the DLA of RI lines decreased with an increasing number of effective DR gene alleles, suggesting a significant cumulative effect of the DR genes on quantitative resistance. Furthermore, in preliminary screens, the lines with all five DR genes also exhibited enhanced resistance to sheath blight (F. Correa et al, unpublished). Since there is no known qualitative resistance to sheath blight (Zou et al 2000), these results are very encouraging.

Association of DR genes with blast resistance QTLs was also found in an advanced backcross population (n = 80 BC_3F_3 and BC_3F_4 lines) of a japonica cultivar, Moroberekan, known to exhibit durable resistance to rice blast in Asia (Wang et al 1994), crossed to a popular Indian upland cultivar, Vandana. The AB-QTL approach was used because it allows analysis of QTLs while developing elite germplasm for commercial use (Tanksley and Nelson 1996). Six candidate gene markers, NBS-LRR (*Pic19*), *Rp1* (a maize rust resistance gene), NBS (7-4F), PR5 (thaumatin), RGA 1-8 (resistance gene analog, LRR), and RGA6-7 (resistance gene analog, kinase), and one simple sequence repeat (SSR) marker (RM21) were significantly associated with quantitative blast resistance in rice ($P = 0.01$). When tested against a single isolate, PO6-6, four candidate genes (oxalate oxidase, 14-3-3, RGA8-4, and RGA1-10) and four SSR markers (RM21, RM168, RM215, and RM250) were associated with resistance to this single isolate. These markers accounted for observed phenotypic variation of up to 28.7%. Seedling reactions in the blast nursery and greenhouse tests in the Philippines were highly correlated.

Field tests of BC_3F_3 and BC_3F_4 lines in India across screening sites identified six lines showing ≤ 5% DLA and 13 lines showing quantitative resistance; however, only two (V4M-6-1-B and V4M-5-3-B) of these quantitatively resistant lines have good agronomic acceptability with a varying level of neck blast resistance.

Cluster analysis of DNA profiles of a BC_3 population derived from Vandana/Moroberekan showed genetic similarity of >85% to the recurrent parent Vandana (Wu et al 2004). To accumulate different mechanisms involved in quantitative resistance to blast, 15 BC_3F_3 or BC_3F_4 lines showing partial resistance and carrying positive candidate alleles were selected and crossed in all pairwise combinations. Among the 10 F_2 families, 14 F_2 plants showed a good to moderate level of agronomic acceptability and high level of seedling blast resistance (0.75–3.2% DLA). Resistant lines from selected families of the crosses VM5/VM14, VM6/VM14, and VM82/VM14 with acceptable agronomic traits resembling those of Vandana were advanced from F_3 to F_6. The field performance of the top 84 F_4 and F_5 lines (out of >600 lines) indicated that the major QTLs have been captured in the BC lines. Multilocation tests of the 84 promising lines reveal an enhanced level of blast resistance of selected lines in blast

hotspots in the Philippines and India. Interestingly, some also perform well under natural drought conditions. The effect of blast loci on yield was established based on an evaluation of pyramided blast lines under managed drought. These lines will be crossed with the most blast-tolerant F_5 intermated lines to combine blast resistance and drought tolerance.

Collectively, the above studies show that certain candidate genes are associated (co-localize on genetic maps) with quantitative disease resistance, and that using DR genes as molecular markers to accumulate chromosomal regions harboring them allowed the selection of rice with higher disease resistance.

The oxalate oxidase gene is a strong candidate for contributing to rice defense responses

In three rice mapping studies, putative oxalate oxidase (*OXO*) gene markers were mapped to QTLs for blast resistance, with two studies showing co-localization of *OXO* to blast QTLs on rice chromosome 8 (Liu et al 2004, Ramalingam et al 2003) and two studies showing co-localization of *OXO* to blast QTLs on chromosome 3 (Ramalingam et al 2003, Wu et al 2004). The cumulative positional data are consistent with a role for *OXO* as one of the largest contributors to resistance when using candidate DR genes as predictors of quantitative blast resistance in rice. Interestingly, *OXO* has also previously been associated with QTLs for tan spot and leaf spot resistance in wheat (Faris et al 1999).

OXO (also classified as a germin-like oxalate oxidase) is predicted to act both directly and indirectly in plant defense responses. The enzyme converts oxalate into CO_2 and H_2O_2, with the release of free Ca^{2+} from Ca-oxalate. Both Ca^{2+} and H_2O_2 have been implicated as signaling molecules for plant defense responses (Levine et al 1994, 1996, Price et al 2002) and H_2O_2 can be directly toxic to microbes (Grant and Loake 2000). In barley seedlings, oxalate oxidase activity was induced in response to an attack by the powdery mildew fungus, and its effect was correlated with apoplastic H_2O_2 generation and increased Ca^{2+} concentrations (Zhou et al 1998). Overexpression of a barley *OXO* gene in peanut resulted in enhanced resistance to *Sclerotinia minor* (Livingstone et al 2005). The *OXO* gene does not seem to control all types of pathogens, however, because while overexpression of the wheat *OXO* gene in sunflower conferred enhanced resistance to the oxalic acid–producing necrotic pathogen *Sclerotinia sclerotiorum* (Hu et al 2003), expression of the gene in wheat epidermal cells did not confer enhanced resistance to the biotrophic wheat powdery mildew fungus (Altpeter et al 2005). Intriguingly, overexpression of *OXO* in sunflower did not cost the plant in terms of fitness (Burke and Rieseberg 2003).

Is expression of *OXO* candidate genes correlated with increased disease resistance?

Genetic mapping studies allowed us to identify candidate DR genes that are located in regions that make large contributions to disease resistance. However, on close analysis of the rice genome sequence containing these genes, several candidates, including the *OXO* genes, were found to be members of tandemly arranged gene families. For

example, 12 putative oxalate oxidase-like (*OXO*) genes were predicted in the 2.8-Mb section on rice chromosome 8, whereas four oxalate oxidase genes, which are members of the germin-like family of proteins, were associated with a blast resistance QTL on chromosome 3 (R. Davidson et al, unpublished). The 12 genes on chromosome 8 are predicted to have two exons and one intron, and there is a high sequence similarity among them. The predicted gene sizes range from 823 to 963 nucleotides, and the predicted proteins range from 221 to 225 amino acids. For convenience, we refer to the members as *OXO1–OXO12* based on their positional order on chromosome 8.

It is possible that only one or a few (if any) of the *OXO* genes function in disease resistance. Thus, our first task was to determine which *OXO* genes might be expressed during the defense response in resistant and susceptible rice cultivars. Because of the conserved nature of the coding regions, gene-specific primers for each *OXO* member were designed in the nonconserved 5′ and 3′ untranslated regions (UTR) for these studies. The gene-specific primers were used for subsequent reverse transcriptase polymerase chain reactions (RT-PCR). We first analyzed the expression of the genes in different ages of unchallenged rice leaves to determine whether *OXO* gene expression is constitutively and/or developmentally regulated. Expression of the 12 genes in leaf tissue was analyzed in two resistant indica parents (IR64 and SHZ-2) and in a susceptible japonica parent (Azucena) by RT-PCR (Fig. 1). Our results showed that six *OXO* members were constitutively expressed in IR64 (*OXO3–5, 7, 11, 12*) and SHZ-2 (*OXO3–7, 12*), whereas eight were constitutive in Azucena (*OXO4–9, 11, 12*). Four members (*OXO6, 8, 9, 11*) exhibited developmental regulation in at least one line. *OXO1, 2,* and *10* were not expressed in unchallenged leaves in any of the lines.

Expression was also determined after challenge with *Magnaporthe grisea* at four time points in IR64, SHZ-2, Azucena, and an additional susceptible japonica parent, LTH (R. Davidson, unpublished). Plants were subjected to two treatments: (1) mock inoculation by spraying surfactant solution and (2) inoculation by spraying 10^5 *M. grisea* conidial spores per mL of surfactant solution. Phenotypic analyses of infected plants at 5 days postinoculation indicated that SHZ-2 exhibited few to no blast lesions and is more resistant than IR64. IR64 is more resistant than the two susceptible parents (Azucena and LTH). The three youngest leaves were pooled and harvested at 0, 12, 24, and 48 hours postinoculation, and qualitative expression analyses for all 12 *OXO* genes were performed by RT-PCR. Results from these studies revealed that some *OXO* members were induced in response to wounding and pathogen challenge. For example, *OXO8, 9,* and *10* were induced by both treatments at 12 and 24 h in all four lines. In SHZ-2, however, *OXO8* and *9* expression was reduced at 48 h in both treatments. *OXO6* was expressed at all time points in SHZ-2.

Among the *OXO* members that were constitutively expressed (expressed at 0 time), some showed interesting profiles. For instance, *OXO5* was expressed at 12 and 24 h in both treatments in all lines. The gene exhibited a reduction in expression at 48 h after wounding, but was still expressed at 48 h after pathogen challenge. Interestingly, *OXO7* showed expression patterns similar to those of *OXO5* in all lines except SHZ-2, where the reduction occurred at 48 h in both treatments. *OXO11* was expressed at all

time points in Azucena, LTH, and IR64; however, SHZ-2 did not show expression of *OXO11* in any condition tested in this study.

Our evidence is consistent with roles for some *OXO* family members not only in plant disease resistance but also in development and response to wounding. Gene expression profiling did not identify any one *OXO* gene as a candidate for primary control of resistance (a combination of family members may be involved), but the data allow the elimination of some family members from future analysis in these rice lines.

SHZ-2, the most resistant line, had a unique expression pattern after wounding and *M. grisea* challenge when compared with the other less resistant lines. Differential expression of some *OXO* members in this line suggests that the amount of transcript and timing of expression may be important factors in the response to our experimental conditions. For example, some *OXO* members in SHZ-2 seem to respond earlier than the other lines after wounding and pathogen challenge. Analysis of gene expression and comparison of *OXO* gene promoter sequences from published japonica and indica genomes indicate that differences in regulation of the *OXO* gene members may correlate with variation in disease resistance. Based on these findings, we are currently analyzing *OXO* gene promoters.

Although the predicted biological functions of the candidate *OXO* genes are consistent with the hypothesis that they are contributing factors to quantitative resistance, their actual contribution has not been demonstrated. To evaluate the contributions of these and other DR genes in resistance, we are using RNAi to suppress expression of the candidate genes (Manosalva et al, unpublished). Preliminary findings indicate that suppression of certain combinations of *OXO* genes renders plants more susceptible to attack by *M. grisea.*

Application of candidate gene–aided selection in breeding

Using sequence information and molecular markers from the rice genome, we designed candidate gene-based markers and identified SSRs within candidate genes. These markers were used to conduct a genome scan to determine the proportion of candidate gene alleles from Moroberekan in 108 introgression lines (84 selected for blast resistance and 24 for drought tolerance) derived from Vandana/Moroberekan populations. Single-marker analysis identified three candidate genes significantly correlated with blast resistance. Oxalate oxidase and peroxidase were correlated with yield under both leaf and panicle blast challenge at a blast hotspot in Almora, India. The presence of the Moroberekan chromosomal segment carrying the thaumatin gene is correlated with neck blast incidence in Cavinti, Philippines. Significant two-gene interactions (P<0.001) were observed in seedling blast (HSP90 × thaumatin and thaumatin × oxalate oxidase) and neck blast (chitinase × thaumatin and thaumatin × oxalate oxidase-like protein). Of the lines selected for blast resistance, 23 yielded more than Vandana under natural drought stress, suggesting that desirable Moroberekan alleles conditioning drought and blast resistance may be linked. These include two lines (IR78221-19-6-7-B-B and IR78221-19-6-99-B-B) that were also high-yield-

ing under high blast pressure in Almora. IR78221-19-6-7-B-B has alleles of oxalate oxidase, peroxidase, and HSP90 from Moroberekan. Of the candidate genes scored, IR78221-19-6-7-B-B has only the oxalate oxidase allele from Moroberekan. We are currently testing whether additional candidate defense genes are responsible for the observed resistance in these lines.

Through a combination of multilocation tests and molecular analysis, gene-based markers strongly associated with partial blast resistance have been identified. With candidate gene–aided selection, we are now in a position to combine blast resistance with drought tolerance in advanced breeding populations for a target environment in India. We are also validating the functional alleles for quantitative blast resistance in different genetic backgrounds and environments. This will allow us to design specific primers from candidate genes associated with quantitative blast resistance in Indonesia and India, and apply candidate gene–derived markers in advanced intercrosses pyramided for blast resistance in these environments.

The whole-genome approach to identify additional candidate genes

As suggested by analyses in several breeding and mapping populations, BSDR is most likely conditioned by multiple genes, some of which may function only in certain genetic backgrounds. Thus, there is a need for a generalized approach to quickly identify candidate genes in different genetic backgrounds or breeding pedigrees. So far, our approaches to identifying candidate genes have been based on sequence inference from rice or other species where genetic and biochemical information is available to suggest a putative defense function. This serves the purpose of investigating target genes with a suspected function, but fails to reveal genes previously unknown to play critical roles in host defense. Based on the analysis of SHZ-2 (Fig. 1), we suspect that additional genes, other than those identified by genetic mapping, could contribute to the total disease resistance phenotype. Yet, we are limited by (1) the set of candidate genes predefined by certain criteria and (2) a lack of genetic polymorphism of the candidate defense genes tested. This situation is not unique to SHZ-2 but a general drawback of studying a predefined list of candidate genes. The complete rice genome sequence together with associated genomic tools gave us new opportunities to apply genome-wide approaches to identify additional candidate genes. The advanced genetic materials developed for BSDR analysis provide an excellent platform to unveil useful genes through whole-genome expression analysis.

We first experimented with whole-genome chips using parental lines of contrasting phenotypes (Liu et al 2005). Expression analysis was conducted using cRNA from SHZ-2 and LTH over a time course most critical for rice-blast interaction (24, 48, and 96 h postinoculation) and hybridized against the Agilent 22K oligo chip. The differential gene expression patterns between SHZ-2 and LTH, under control and pathogen-induced conditions, revealed a new set of genes exhibiting "expression polymorphism" that can then be associated with chromosome regions where phenotypic QTLs are mapped. Figure 2 shows the experimental approach for the identification of additional candidate genes using a whole-genome approach. Preliminary results

OXO gene family member	Expression in leaf 1, youngest, to leaf 3, oldest, of rice cultivar								
	IR64			SHZ-2			Azucena		
	1	2	3	1	2	3	1	2	3
1									
2									
3	√	√	√	√	√	√			
4	√	√	√	√	√	√	√	√	√
5	√	√	√	√	√	√	√	√	√
6				√	√			√	√
7	√	√	√	√	√	√	√	√	√
8						√		√	√
9								√	√
10									
11			√				√	√	√
12	√	√	√	√	√	√	√	√	√

Fig. 1. Expression of rice chromosome 8 *OXO* gene family members in rice cultivars IR64 and SHZ-2 (which carry the QTL associated with blast resistance on chromosome 8) and Azucena (which does not carry the QTL associated with blast resistance) in three different leaf positions (leaf 1, youngest, to leaf 3, oldest). The expression of the gene family members (*OXO1–OXO12*) was determined using gene-specific primers in reverse transcriptase (RT) PCR. Leaves were harvested from plants at 21 days after sowing (R. Davidson et al, unpublished).

suggest that this relatively simple analysis (in terms of genotypes and treatments) provides an enrichment of genes that are associated with phenotypic effects. Of 38 differentially expressed genes identified at three time points, 18 (47%) were located in chromosomal regions previously not known to carry disease resistance QTLs (Liu et al 2005). Such a correlative analysis of expression patterns in the parental lines does not inform the causal effects of expression QTLs as in a full-scale "segregation genomics" analysis (Schadt et al 2003). However, we are able to identify new genes previously not known to be involved in disease resistance. Additional supporting evidence can be obtained by correlating the expression patterns with genotypic data in advanced backcross progeny.

To further determine the *cis*- and *trans*-acting relationships of expression QTLs, genome-wide expression analysis needs to be conducted with segregating progeny. The power of such analysis has been demonstrated in other systems (Doss et al 2005, Flaherty et al 2005, Schadt et al 2003), but a limiting factor is the high cost involved in analyzing all recombinant inbred progeny. Until the cost of expression analysis is

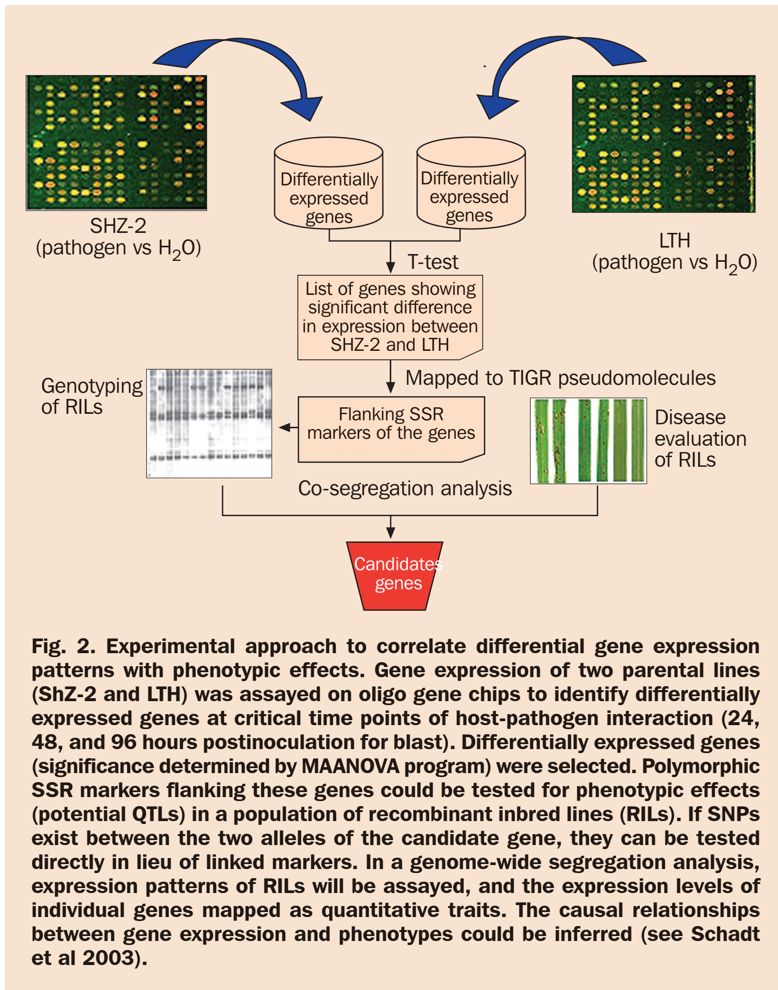


Fig. 2. Experimental approach to correlate differential gene expression patterns with phenotypic effects. Gene expression of two parental lines (ShZ-2 and LTH) was assayed on oligo gene chips to identify differentially expressed genes at critical time points of host-pathogen interaction (24, 48, and 96 hours postinoculation for blast). Differentially expressed genes (significance determined by MAANOVA program) were selected. Polymorphic SSR markers flanking these genes could be tested for phenotypic effects (potential QTLs) in a population of recombinant inbred lines (RILs). If SNPs exist between the two alleles of the candidate gene, they can be tested directly in lieu of linked markers. In a genome-wide segregation analysis, expression patterns of RILs will be assayed, and the expression levels of individual genes mapped as quantitative traits. The causal relationships between gene expression and phenotypes could be inferred (see Schadt et al 2003).

considerably reduced, expression analysis of parental lines coupled with information from selected advanced progeny (e.g., advanced BC lines or NILs) can be a cost-effective way to extract information from diverse donor parents.

QTL alignment and orthologous genes across cereals

The transfer of major R genes into cereal crops from related grasses has been an important tool for improving the disease and pest resistance of some cereals. To date, this approach has been limited to cereals and their very close relatives based on conventional genetic approaches. Recently, however, a few examples have emerged that demonstrate that resistance genes can be effectively transferred between distantly

related grasses in different tribes (e.g., rice, maize, and wheat), which may have a large impact on resistance breeding. For example, a "typical" NBS-LRR-type qualitative resistance gene, *Rxo1*, was identified in maize based on its ability to recognize and initiate a defensive hypersensitive response to the rice bacterial leaf streak pathogen, *Xanthomonas oryzae* pv. *oryzicola* (Zhao et al 2004). Transgenic rice lines that carried the maize *Rxo1* gene exhibited resistance when inoculated with pathogenic bacteria carrying the *avrRxo1* gene (Zhao et al 2005). Thus, in some cases, qualitative resistance genes can function to signal resistance responses, even in species that are evolutionarily quite diverged.

We hypothesize that across species, functioning quantitative resistance could be even more prevalent because it represents the basal defense mechanisms that commonly evolved. Such mechanisms may not depend on the activation of signaling pathways, but may be constitutive and/or directly interacting with the pathogen to contribute to resistance. Phenotypically, we observe broad-spectrum quantitative resistance among cereals against multiple pathogen races. For example, resistance to northern leaf blight in maize is known to be controlled by quantitative resistance that can be accumulated to a high level through recurrent selection (Ceballos et al 1991). It is also intriguing that some cereals are essentially "immune" to certain pathogen taxonomic groups. For example, there are no rust diseases in rice. Is this because of structural incompatibility between the host and the pathogen? Does this reflect an inherent ability of rice to resist rust through mechanisms underpinning broad-spectrum resistance? As shown in non-host resistance in *Arabidopsis*, the manifestation of broad-spectrum disease resistance could result from a combination of pre- and postinvasion genetic mechanisms (Lipka et al 2005). Testing the relationship between orthologous sequences of candidate genes and BSDR across species provides another means to identify conserved function and perhaps sequence motifs that are important for general host defense.

Because of genome synteny among cereals (Devos and Gale 2000) and because the complete genome sequence of rice is now known (Goff et al 2002, The International Rice Genome Sequencing Project 2005, Yuan et al 2001), rice is proving to be an excellent model for understanding genomes of other cereal crop species whose genomes are more complex or that are not so well characterized as rice. To test the utility of cross-genome comparisons for analysis of disease resistance, Wisser et al (2005, 2006) performed a comprehensive bioinformatics analysis of previously published studies that mapped loci governing quantitative disease resistance in rice and maize. Sixteen studies on disease QTLs for rice (Wisser et al 2005) and 33 for maize (Wisser et al 2006) were summarized. Although much of the cereal genome was implicated in conditioning quantitative disease resistance, distinct areas of the genome are associated with resistance to multiple diseases in multiple studies. Intriguingly, some of the most QTL-rich regions of the maize genome are syntenous with QTL-rich regions of the rice genome (Wisser et al 2006), which bodes well for a candidate gene approach. However, because the number of DR genes that have been implicated in the plant defense response is so large, and because the resolution of the QTLs is so poor (as some span tens of centi-Morgans and they sometimes overlap), hundreds of credible candidate genes could be associated with each of several disease QTL re-

gions analyzed in the rice genome. Thus, a more precise localization of disease QTLs and functional analysis of candidate DR genes are needed for the QTLs to be more effectively characterized and exploited. This task could be greatly facilitated by fine mapping and analysis of the large collection of rice mutants available for functional tests by forward and reverse genetics (Hirochika et al 2004 and reviewed in other chapters in this volume).

Conclusions

We have summarized our progress toward understanding the molecular basis of quantitative disease resistance in rice. Using mapping studies, we have demonstrated that DR genes co-localize with regions containing disease QTLs. Using molecular markers linked to these genes, we accumulated the regions harboring the DR genes in rice, which resulted in enhanced disease resistance. Taken together, these findings are consistent with the hypothesis that the DR genes function in quantitative disease resistance. However, the genomic regions underlying the QTLs are quite large, and the possibility remains that the DR genes of interest may co-localize but may not actually contribute to resistance, or that DR genes other than those initially considered are those actually responsible for conditioning resistance. Knowledge of whether or not the candidate DR gene actually functions in quantitative resistance is critical as we develop molecular markers based on the candidate genes for practical breeding programs. Thus, our current focus is to determine if indeed the DR genes are actually contributing to disease resistance. Preliminary data suggest that at least some oxalate oxidase gene family members do contribute to enhanced resistance.

As more is learned about defense response pathways in plants in general, it is becoming clear that some of the biochemical pathways that contribute to defense responses are common among plant species. Gene expression profiling experiments from multiple plant species undergoing defense responses have revealed the induction of many of the same genes, suggesting that plants have evolved common approaches to resist pathogens. This was the guiding principle behind the selection of the DR genes initially used as candidates for the studies reviewed here, that is, they were selected for screening because they were implicated in defense responses in rice or some other crop species (Ramalingam et al 2003). As mentioned earlier, the *OXO* gene family members were associated with defense responses in barley (Zhou et al 1998), and they also co-localize with QTLs for tan spot and leaf spot resistance in wheat (Faris et al 1999). Given the commonalities of some defense responses employed by plants, it is exciting to contemplate the broader implications for the candidate gene approach to improving disease resistance in other crop species whose genomes are not sequenced. Because of the rich genetic and genomic resources available, rice can serve as the model crop plant for testing candidate gene function in disease resistance.

References

Altpeter F, Varshney A, Abderhalden O, Douchkov D, Sautter C, Kumlehn J, Dudler R, Schweizer P. 2005. Stable expression of a defense-related gene in wheat epidermis under transcriptional control of a novel promoter confers pathogen resistance. Plant Mol. Biol. 57:271-283.

Bai J, Pennill L, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH. 2002. Diversity of nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res. 12:1871-1884.

Bent AF. 1996. Plant disease resistance genes: function meets structure. Plant Cell 8:1757-1771.

Bonman JM, Khush GS, Nelson RJ. 1992. Breeding rice for resistance to pests. Annu. Rev. Phytopathol. 30:507-528.

Burke JM, Rieseberg LH. 2003. Fitness effects of transgenic disease resistance in sunflowers. Science 300:1250.

Ceballos H, Deutsch JA, Gutierrez H. 1991. Recurrent selection for resistance to *Exserohilum turcicum* in eight subtropical maize populations. Crop Sci. 31:964-971.

Correa-Victoria FJ, Zeigler RS. 1995. Stability of partial and complete resistance in rice to *Pyricularia grisea* under rainfed upland conditions in eastern Colombia. Phytopathology 85:977-982.

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

Diatchenko L, Lau YC, Campbell A, Chenchik A, Moqadam F, Huang B, Lukyanov SA, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA 93:6025-6030.

Dixon RA, Harrison MJ. 1990. Activation, structure, and organization of genes involved in microbial defense in plants. Adv. Genet. 28:165-234.

Doss S, Schadt EE, Drake TA, Lusis AJ. 2005. *cis*-acting expression quantitative trait loci in mice. Genome Res. 15:681-691.

Epple P, Apel K, Bohlmann H. 1997. Overexpression of an endogenous thionin enhances resistance of *Arabidopsis* against *Fusarium oxysporum*. Plant Cell 9:509-520.

Faris JD, Li WL, Liu DJ, Chen PD, Gill BS. 1999. Candidate gene analysis of quantitative disease resistance in wheat. Theor. Appl. Genet. 98:219-225.

Flaherty L, Herron B, Symula D. 2005. Genomics of the future: identification of quantitative trait loci in the mouse. Genome Res. 15:1741-1745.

Goff S, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100.

Grant JJ, Loake GJ. 2000. Role of reactive oxygen intermediates and cognate redox signaling in disease resistance. Plant Physiol. 124:21-29.

Grison R, Grezes-Besset B, Schneider M, Lucante N, Olsen L, Leguay JJ, Toppan A. 1996. Field tolerance to fungal pathogens of *Brassica napus* constitutively expressing a chimeric chitinase gene. Nature Biotechnol. 14:643-646.

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.

Hirochika H, Guiderdoni E, An G, Hsing YI, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang Q, Pereira A, Sundaresan V, Leung H. 2004. Rice mutant resources for gene discovery. Plant Mol. Biol. 54:325-334.

Hu X, Bidney DL, Yalpani N, Duvick JP, Crasta O, Folkerts O, Lu G. 2003. Overexpression of a gene encoding hydrogen peroxide-generating oxalate oxidase evokes defense responses in sunflower. Plant Physiol. 133:170-181.

Hulbert S. 2002. Arming plants against pathogens: recombination in resistance loci. Trends Plant Sci. 3:121-122.

Hulbert SH, Webb CA, Smith SM, Sun Q. 2001. Resistance gene complexes: evolution and utilization. Annu. Rev. Phytopathol. 39:285-312.

The International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Johnson R. 1981. Durable resistance: definition of genetic control and attainment in plant breeding. Phytopathology 71:567-568.

Leach JE, Liu B, Manosalva P, Wu C, Wu J, Bordeos A, Bai J, Lee S, Ryba-White M, Bruce M, et al. 2004. Dissection of durable resistance in rice. In: Tsuyumu S, Leach JE, Shiraishi T, Wolpert T, editors. Genomic and genetic analysis of plant parasitism and defense. St. Paul, Minn. (USA): APS Press. p 164-173,

Leung H, Wu C, Baraoidan MR, Bordeos A, Ramos M, Madamba S, Cabauatan P, Vera C Cruz, Portugal A, Reves G, et al. 2001. Deletion mutants for functional genomics: progress in phenotyping, sequence assignment, and database development. Rice genetics IV. New Delhi (India): Science Publishers, Inc. p 239-251.

Leung H, Zhu Y, Revilla-Molina I, Fan JX, Chen H, Pangga I, Vera Cruz C, Mew TW. 2003. Using genetic diversity to achieve sustainable rice disease management. Plant Dis. 87:1156-1169.

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.

Levine A, Pennell RI, Alvarez ME, Palmer R, Lamb C. 1996. Calcium-mediated apoptosis in a plant hypersensitive disease resistance response. Curr. Biol. 6:427-437.

Lipka V, Dittgen J, Bednarek P, Bhat R, Wiermer M, Stein M, Landtag J, Brandt W, Rosahl S, Scheel D, et al. 2005. Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. Science 310:1180-1183.

Liu B, Zhang SH, Zhu XY, Yang Q, Wu S, Mei M, Mauleon R, Leach JE, Mew T, Leung H. 2004. Candidate defense genes as predictors of quantitative disease resistance in rice. Mol. Plant-Microbe Interact. 17:1146-1152.

Liu B, Satoh K, Mauleon R, Zhang S, Zhu X, Yang J, Yang Q, Wu S, Hulbert S, Leach J, Kikuchi S, Leung H. 2005. Associating differentially expressed genes with genomic variation in breeding lines reveals additional genes contributing to blast resistance. Abstract at Fifth Rice Genetics/Third Rice Functional Genomics Symposium, 19-23 Nov. 2005, Manila, Philippines. (Abstract.)

Livingstone DM, Hampton JL, Phipps PM, Grabau EA. 2005. Enhancing resistance to *Sclerotinia minor* in peanut by expressing a barley oxalate oxidase gene. Plant Physiol. 137:1354-1362.

Maher EA, Bate NJ, Ni W, Elkind Y, Dixon RA, Lamb CJ. 1994. Increased disease susceptibility of transgenic tobacco plants with suppressed levels of preformed phenylpropanoid products. Proc. Natl. Acad. Sci. USA 91:7802-7806.

Price A, Knight M, Knight H, Cuin T, Tomos D, Ashenden T. 2002. Cytosolic calcium and oxidative plant stress. Biochem. Soc. Trans. 24:479-483.

Ramalingam J, Vera Cruz CM, Kukreja K, Chittoor JM, Wu JL, Lee SW, Baraoidan MR, George ML, Cohen M, Hulbert S, Leach JE, Leung H. 2003. Candidate resistance genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. Mol. Plant-Microbe Interact. 16:14-24.

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, et al. 2003. Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297-302.

Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B. 1993. Plant 'pathogenesis-related' proteins and their role in defense against pathogens. Biochimie 75:687-706.

Tanksley SD, Nelson RC. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92:191-203.

Wang G, Mackill DJ, Bonman JM, McCouch SR, Nelson RJ. 1994. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136:1421-1434.

Wang GL, Wu C, Zeng L, He C, Baraoidan M, Goaes da Silva F, Williams C, Ronald P, Leung H. 2004. Isolation and characterization of rice mutants susceptible to *Xanthomonas oryzae* pv. *oryzae*. Theor. Appl. Genet. 108:379-384.

Wisser R, Balint-Kurti P, Nelson R. 2006. The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96:120-129.

Wisser RJ, Sun Q, Hulbert SH, Kresovich S, Nelson RJ. 2005. Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. Genetics 169:2277-2293.

Wu G, Shortt BJ, Lawrence EB, Levine EB, Fitzsimmons KC, Shah DM. 1995. Disease resistance conferred by expression of a gene encoding H_2O_2-generating glucose oxidase in transgenic potato plants. Plant Cell 7:1357-1368.

Wu J, Sinha P, Variar M, Zheng K, Leach JE, Courtois B, Leung H. 2004. Association between candidate genes and blast resistance in an advanced backcross population of rice. Theor. Appl. Genet. 108:1024-1032.

Wu J, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba RS, Ramos-Pamplona M, Mauleon R, Portugal A, Ulat V, et al. 2005. Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Mol. Biol. 59:85-97.

Young ND. 1996. QTL mapping and quantitative disease resistance in plants. Annu. Rev. Phytopathol. 34:479-501.

Yuan Q, Quackenbush J, Razvan S, Pertea M, Salzberg SL, Buell CR. 2001. Rice bioinformatics: analysis of rice sequence data and leveraging the data to other plant species. Plant Physiol. 125:1166-1174.

Zhao B, Lin X, Poland J, Trick H, Leach J, Hulbert S. 2005. A maize resistance gene functions against bacterial streak disease in rice. Proc. Natl. Acad. Sci. USA 102:15383-15388.

Zhao BY, Ardales E, Brasset E, Claflin LE, Leach JE, Hulbert S. 2004. The *Rxo1/Rba1* locus of maize controls resistance reactions to pathogenic and non-host bacteria. Theor. Appl. Genet. 109:71-79.

Zhou F, Zhang Z, Gregersen L, Mikkelsen JD, de Neergaard E, Collinge DB, Thordal-Christensen H. 1998. Molecular characterization of the oxalate oxidase involved in the response of barley to the powdery mildew fungus. Plant Physiol. 117:33-41.

Zhu Q, Maher EA, Masoud S, Dixon RA, Lamb CJ. 1994. Enhanced protection against fungal attack by constitutive co-expression of chitinase and glucanase genes in transgenic tobacco. Bio/Technology 12:807-812.

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

Notes

Authors' addresses: J.E. Leach, R. Davidson, P. Manosalva, M. Bruce, J. Stephens, and M.G. Diaz, Colorado State University, Fort Collins, CO, 80523-1177, USA, e-mail: Jan.Leach@colostate.edu; B. Liu, R. Mauleon, G. Carrillo, C. Vera Cruz, and H. Leung, International Rice Research Institute, Los Baños, Philippines; R. Nelson, Cornell University, Ithaca, NY, USA.

Acknowledgments: This work was supported by the USDA-CSREES-NRI, Rockefeller Foundation, USAID, Generation Challenge Program, and Colorado State University Agricultural Experiment Station.

Identification and transfer of trait-enhancing alleles from wild species

S.R. McCouch, M. Sweeney, J. Li, H. Jiang, M. Thomson, E. Septiningsih, P. Moncada, J. Xiao, J. Coburn, E. Fraker, A. Garris, T. Tai, C. Martínez, J. Tohme, M. Sugiono, A. McClung, L.P. Yuan, and S.-N. Ahn

The approach demonstrated by this collaborative breeding project has had important implications for the use of exotic germplasm in wide crosses of rice. We demonstrated that AB-QTL analysis is capable of (1) successfully uncovering positive alleles that were not obvious based on the phenotype of the parent, (2) offering an estimation of the value of crosses between *O. sativa* and exotic or genetically distant germplasm, and (3) identifying molecular markers for numerous alleles of interest to aid in their incorporation into elite cultivars with a minimum of linkage drag.

The project explored the distribution of diversity within and between subpopulations of rice, building on the unique evolutionary history of the species to explore the genetic architecture and combining ability of groups within and between species. The long-term potential of exploiting the well-partitioned gene space in rice depends on appropriate management of these gene pools and a sound intellectual framework within which the genetic variation of *Oryza* is explored and manipulated. It is of great interest to integrate knowledge about the evolution and natural population structure of this and other domesticated species to better manage and exploit natural variation for crop improvement.

Keywords: *Oryza rufipogon*, AB-QTL analysis, population structure, exotic germplasm, gene cloning

Rice (*Oryza sativa*) is a cultivated, inbreeding species that was domesticated in Asia approximately 10,000 years ago from a complex ancestral species, *O. rufipogon* (Oka 1988b, Vaughan et al 2005). Two major groups (subspecies) within *O. sativa* have been recognized since ancient times: indica (which includes Hsien rice from China, Aman rice from Bangladesh, and Tjereh rice from Indonesia) and japonica (which includes Keng rice from China and Bulu rice from Indonesia) (Matsuo et al 1997, Vaughan et al 2001). Recent studies estimate that the time since divergence of the indica and japonica groups is approximately 0.44 million years, based on sequence comparisons between the two fully sequenced rice genomes, cv. Nipponbare (japonica) and cv. 9311 (indica) (Ma and Bennetzen 2004). This time estimate suggests that the indica and japonica

groups predate the domestication of *O. sativa* by several hundred thousand years and is consistent with the belief that rice was independently domesticated on at least two occasions from a predifferentiated ancestral pool (Cheng et al 2003, Garris et al 2005, Roy 1921, Second 1982, 1991). The hypothesis of multiple independent domestication events in rice was originally proposed based on studies of hybrid sterility (Kato et al 1928) and was further supported by evidence from isozyme analysis (Second 1982). More recently, it has gained further credibility based on evidence from DNA markers (Cheng et al 2003, Garris et al 2005). The fact that different subpopulations of *O. sativa* appear to have been domesticated from a common ancestral gene pool with broad adaptation throughout Asia is in contrast to the single domestication event that has been documented for modern maize (Matsuoka et al 2002). This also raises interesting questions about the inherent genetic diversity and population structure of *O. rufipogon* as a progenitor of modern *O. sativa*.

Population structure of *O. sativa*

Differences between the indica and japonica subspecies have been recognized since ancient times but the two groups present overlapping ranges of variation and have been difficult to classify based on eco-geographical or phenotypic characters. In eco-geographical terms, indica is primarily known as lowland rice that is grown throughout tropical Asia, while japonica is typically found in temperate East Asia, upland areas of Southeast Asia, and high elevations in South Asia. The appearance of nonsticky (indica) and sticky (japonica) rice is documented in Chinese literature as early as AD 100 (Matsuo et al 1997) and other traits that have been used to classify indica and japonica include grain shape, phenol reaction, sensitivity to potassium chlorate, leaf color and pubescence, apiculus hair length, and glume color at heading (Glaszmann 1987, Chang 1976), but the spectra of variation for any of these individual traits do not clearly distinguish the two subspecies (Oka 1988b).

Molecular markers (both biochemical and DNA markers) provide an opportunity to assess genetic differences (detected as differences in allele frequencies) between subpopulations of rice that may not be clear based on geographical/ecological or phenotypic evaluation. Because of their inbreeding habit, subpopulations of rice are expected to show greater genetic differentiation than would be the case in an outcrossing species. This is because there are fewer opportunities for cross-pollination between lines within a landrace or between adjacent landraces (despite their geographical proximity) in autogamous species, leading to greater genetic compartmentalization and more distinctive genetic structure. The older the divergence time, the deeper the degree of differentiation between populations, and the greater the number of accumulated genetic differences. Thus, if the indica-japonica differentiation is really as old as 0.44 million years, it should be readily detectable with a small number of randomly selected molecular markers. If additional subpopulation structure has evolved since the divergence of the indica and japonica groups, it should also be detectable at the molecular level, though a larger number of molecular polymorphisms might be needed to identify more recently evolved subgroups.

Using a small number (10–30) of restriction fragment length polymorphism (RFLP) markers to evaluate *O. sativa* landraces and modern varieties, the indica-japonica division was very clear (Nakano et al 1992, Wang and Tanksley 1989, Zhang et al 1992), but no clear subpopulation structure was evident beyond this ancient divide. On the other hand, Glaszmann (1987) identified six varietal groups, indica, japonica, aus, aromatic, rayada, and ashina, using 15 isozyme loci when he analyzed a group of 1,688 *O. sativa* landraces and modern varieties. The aus, rayada, and ashina are minor groups that have generally been considered to be indica ecotypes, and all have a comparatively small geographic distribution along the Himalayan foothills. The early-maturing aus rice is widely grown in Bangladesh and along the Himalayas as far west as Iran, and rayada and ashina are floating rice of Bangladesh and India, respectively. Aromatic rice such as basmati from Pakistan, Nepal, and India and sadri from Iran is well known for its distinctive aroma and elongating grains as well as its crossing barriers, and the genetic origin of this group has been debated for years (Glaszmann 1987, Khush 2000, Nagaraju et al 2002). The temperate and tropical japonica rice can be difficult to distinguish genetically, though their different zones of adaptation make them recognizeable as separate gene pools.

In a recent study by Garris et al (2005), five subpopulations were clearly identified in *O. sativa* using simple sequence repeat (SSR) markers; the indica and aus subpopulations cluster within the indica subspecies and the tropical japonica, temperate japonica, and aromatic subpopulations cluster within the japonica subspecies (Fig. 1). This population structure appeared to be quite ancient, with overall estimates of divergence between subpopulations showing Fst values of 0.375 (approximately 37.5% of the observed genetic variation was characterized as "between-group" differences), compared with Fst values of 0.08–0.09 for heterotic groups in maize (Matsuoka et al 2002) or 0.08–0.11 for humans residing on different continents (Barbujani et al 1997). The basic number and identity of the subpopulations identified by Garris et al (2005) using 169 SSRs were consistent with those detected by Glaszmann (1987) using only 15 isozyme markers, or by single nucleotide polymorphisms (SNPs) assayed at 112 random genes (A. Caicedo, NCSU, pers. comm.). The SSRs offered higher resolution than the 15 isozyme markers, making it possible to distinguish the temperate and tropical japonica groups (Garris et al 2005) whereas the failure to identify the rayada and ashina subpopulations in the study with SSRs may be because different landrace accessions were used in the studies by Garris et al (2005) and Glaszmann (1987). A more complete study of the population substructure of *O. sativa* is currently under way as part of the Generation Challenge Program using a diverse set of molecular markers and several thousand accessions. Preliminary results suggest that the same groups identified by Garris et al (2005) and Glaszmann (1987) are emerging in this study (www.generationcp.org/vw/modules.php?name=News&file=article&sid=43), providing strong support for the underlying population substructure described to date.

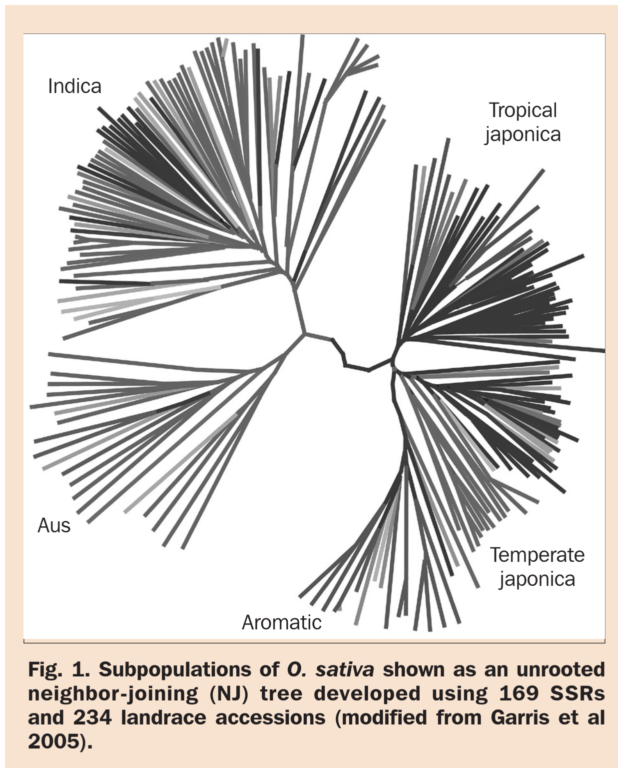


Fig. 1. Subpopulations of *O. sativa* shown as an unrooted neighbor-joining (NJ) tree developed using 169 SSRs and 234 landrace accessions (modified from Garris et al 2005).

Challenges of intraspecific breeding

The genetic isolation of the subpopulations of *O. sativa* has been maintained by their natural inbreeding habit and reinforced by partial sterility factors, making it problematic for breeders to generate fertile recombinant offspring from indica × japonica crosses (Harushima et al 2002, Oka 1988a, Sano 1993). Breeding programs based on indica-japonica hybridization boast a few success stories (*Tongil* rice in Korea, Choi 1974, 1978) but are inevitably plagued by genome-wide sterility factors that lead to hybrid breakdown and a persistent loss of specific recombinants in successive generations (Oka 1988b). The phenomenon is reminiscent of Dobzhansky-Muller incompatibilities (Dobzhansky 1936, Muller 1942), in which there is a gradual accumulation of mutations that cause incompatible epistatic interactions when two species (or subspecies) are hybridized.

Thus, inbred variety development in modern rice breeding programs has focused almost exclusively on crosses between members of the same subpopulation (indica-indica) (Marri et al 2005) or between related subpopulations (i.e., tropical japonica × temperate japonica) (Lu et al 2005, Ni et al 2002, Sano 1993). The long-term consequences for plant breeders of restricting crossing and population development

are that there is a limited pool of genetic variation available for identifying new and useful combinations of genes. Over time, this aggravates the "genetic bottlenecks" that were initially created during the process of domestication and leads to a cryptic form of genetic erosion that dramatically slows the rate of genetic gain possible by plant breeding programs. Yet there is no shortage of genetic variation available in natural populations of *Oryza*. Notably for rice breeders, a great deal of unexploited genetic variation is readily available in the numerous landraces of *O. sativa* and in the progenitor species, *O. rufipogon,* as well as in other sexually compatible AA genome species of *Oryza.*

Yet, rice breeders interested in hybrid variety development have turned their attention to intraspecific crosses (between subpopulations) with great success. In China, hybrid rice breeders have explored a much larger portion of the genetic variability of *O. sativa* than most other breeding programs, making hundreds of thousands of intra- and intersubpopulation crosses aimed at identifying productive heterotic combinations (Li and Yuan 2000). Although this has been very productive, hybrid rice breeding and seed delivery are a difficult, costly, and technology-intensive enterprise, in large part due to the enclosed floret morphology and the naturally inbreeding habit of the species. It is difficult to achieve reliable and economically viable levels of outcrossing during F_1 hybrid seed production and, as a result, hybrid rice breeding as a commercial enterprise is still in its infancy. Nonetheless, experience in hybrid rice breeding programs around the world highlights the value of crossing between subpopulations (particularly indica × japonica) as a way of maximizing levels of heterosis, with the caveat that wide-compatibility genes must be used to overcome the associated sterility barriers (Li and Yuan 2000, Zheng et al 1994).

Challenges of interspecific breeding

Interspecific crossing has been used to introduce several important qualitative characters into elite breeding material, such as disease resistance and male sterility (Brar and Khush 1997, Dalmacio et al 1995). However, when useful genes are transferred from genetically distant species, the process is often encumbered by anomalous patterns of recombination (Brar et al, this volume) and by linkage drag, whereby many undesirable genes are transferred along with desirable ones. This problem is especially serious in wide hybridization programs using evolutionarily divergent species because the effects of unwanted genes from the wild species can be dramatic. It has been shown both theoretically and empirically that a single gene transferred from a wild species will be associated with enough linked chromosomal DNA to contain more than 100 other potentially undesirable genes, even after 20 or more years of traditional breeding (Stam and Zeven 1981, Young and Tanksley 1989). In cases where linkage drag cannot be overcome, it is virtually impossible to obtain useful products from wide hybridization efforts. The use of molecular markers to identify selected recombinant individuals that retain only a very small piece of wild chromosome can greatly facilitate the transfer of specific "desirable" genes while avoiding the transfer of many "undesirable" genes also found in wild species.

A second major challenge in working with wild or exotic germplasm resources is the difficulty in identifying genes from unadapted materials that are likely to enhance the performance of elite cultivars without disrupting gene complexes that provide local adaptation and satisfy commercial grain quality requirements. It is understandable that breeders are reluctant to lose the immediate yield and quality advantages of elite germplasm in efforts to identify rare new genes that might ultimately help expand the genetic base and break an existing "yield barrier." Furthermore, because wild or exotic germplasm is likely to contain more genes that reduce yield and quality than genes that could improve these traits, investment in these kinds of crosses has generally been considered a poor risk.

Given the renewed interest in understanding and documenting subpopulation structure in rice, it is of interest to determine whether a better understanding of the genetic differences underlying the various subpopulations of *O. sativa* and its progenitor, *O. rufipogon,* will provide new insights into the combining ability and potential for generating heterotic combinations (for hybrids) or transgressive segregants (for inbred varieties) from crosses between well-defined subpopulations of both domesticated and wild species.

Population structure of *O. rufipogon*

Oryza rufipogon is considered to be the progenitor of *O. sativa* and it is a complex species that is genetically more diverse than *O. sativa* (Sun et al 2002). It is found dispersed throughout a wide range of habitats across most of tropical and subtropical Asia. Like *O. sativa, O. rufipogon* comprises genetically identifiable subpopulations that show strong geographical and ecological differentiation (Edwards et al 2005, Lu et al 2002, Morishima et al 1984, Oka and Morishima 1967). To date, there is no convincing evidence that modern *O. rufipogon* populations can be clearly identified with a predomesticated form of either indica or japonica, though genetic similarities between specific *O. rufipogon* populations and groups of indica or japonica cultivars have been reported (Cheng et al 2003, Edwards 2005, Londo et al 2006). Conclusions about the phylogeny of *O. sativa* subpopulations are complicated by the problematic issue of recent gene flow from cultivated to wild populations and movement of wild seed along with *O. sativa* that make it difficult to decipher ancestor-descendent relationships (Lu et al 2002).

A census of *O. rufipogon* populations around the world and a rigorous evaluation of their genetic composition using a set of common genetic markers would offer important insights into the genetic diversity, geographic dispersion, population structure, and evolutionary history of these ancient populations. It would also allow researchers to explore the evolutionary origins of the subpopulations of modern *O. sativa* with enhanced understanding of the domestication process. Further, such a census would provide a valuable organizational principle for investigating differences in the combining ability and heterotic potential of the different subpopulations of both wild and exotic germplasm and would enhance the possibilities for efficient use of wild relatives in plant improvement.

Exploring the genetic potential of wild and exotic germplasm for rice improvement

In this section, we will present a summary of results obtained from a decade of research exploring the potential of using low-performing wild and exotic germplasm as donors to enhance the performance of elite cultivars of *O. sativa*. We will ask what genes underlie some of the critical QTLs associated with interspecific breeding and how QTLs from the same (and different) wild or exotic donors contribute to phenotypic performance of different elite cultivars. We raise questions about the optimal genetic distance at which positive transgressive variation is detected in crosses with *O. sativa* and we discuss some perceived advantages of using *O. rufipogon* (or *O. nivara*) as donors in an advanced backcross (AB) breeding program.

Objectives

The objectives of the study were (1) to identify and use novel forms of genetic variation for rice improvement and (2) to understand the molecular basis of QTLs associated with domestication and transgressive variation in rice. We are also interested in using information about genetic diversity and population substructure in both *O. sativa* and *O. rufipogon* to understand the genetic basis of transgressive variation and to make predictions about optimal parental combinations for use in breeding.

Approach

Our approach aimed to explore the potential of *O. rufipogon* as a source of useful genetic variation for rice improvement. Using AB-QTL analysis, we selectively introduced useful components of quantitative trait variation from this wild/weedy gene pool into elite cultivars and evaluated the introgression lines in diverse environments around the world. Over the course of the project, an international network of rice breeders and geneticists independently tested the AB-QTL strategy in their own environments, using elite recurrent parents from diverse subpopulations that were adapted to vastly different growing environments. As a group, we sought to determine whether underused sources of natural genetic variation could serve as useful donors for both inbred and hybrid variety development. Coordination of breeding efforts and sharing of data enabled us to compare results from each group and to identify *O. rufipogon*-derived QTLs that contributed positively to performance in multiple environments and genetic backgrounds, while identifying other QTLs that appear to be useful in only a limited number of situations.

Having demonstrated that AB-QTL analysis was an efficient strategy for expanding the cultivated gene pool and harnessing the hidden potential of poor-performing wild and weedy accessions of *O. rufipogon*, we now look to better understand how to predict which combinations of parents are likely to be most productive in terms of genetic gain. We propose to use population substructure to provide a framework for guiding our efforts to enhance the efficiency and productivity of interspecific breeding efforts in the future.

Table 1. AB-QTL analysis and favorable QTLs detected in BC_2F_2 populations derived from crosses between diverse recurrent parents (RP) and *O. rufipogon* (IRGC 105691).

Parent	References	Participating institution	Origin	Percent favorable QTLs/(country where tested)
Donor				
O. rufipogon (acc. 105691)	Unpublished	IRRI[a]	Malaysia	–
RP				
Ce64/indica hybrid	Xiao et al (1998)	CNHRRDC[b]	China	51.0 (China)
Caiapo/tropical upland japonica	Moncada et al (2001)	CIAT,[c] Embrapa[d]	Brazil	56.0 (Colombia)
Hwacheong/temperate japonica	Cho et al (2003)	RDA,[e] Chungnam Nat'l Univ.	Korea	–
Jefferson/tropical japonica (irrigated)	Thomson et al (2003)	USDA-ARS[f]	USA	53.0 (USA)
IR64/indica inbred	Septiningsih et al (2003a,b)	ICABIOGRAD[g]	IRRI	33.0 (Indonesia)

[a]Internationational Rice Research Institute. [b]China National Hybrid Rice Research and Development Center. [c]Centro Internacional de Agricultura Tropical. [d]Empresa Brasileira de Pesquisa Agropecuária. [e]Rural Development Administration. [f]United States Department of Agriculture-Agricultural Research Service. [g]Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development.

Materials and methods

An interspecific crossing program started in the early 1990s with collaborating institutions in China, Korea, Indonesia, Colombia, Brazil, Côte d'Ivoire, and the U.S. (McCouch et al 2002). One or two elite cultivars from each collaborating national or international program were selected for hybridization with a wild/weedy accession of *O. rufipogon* from Malaysia (IRGC 105491) (Table 1). The interspecific hybrid plants were crossed back to the elite cultivated varieties for two generations to produce a series of 200–300 BC_2F_2 families or, in the case of hybrids, a series of BC_2F_1 testcross families. A flow diagram outlining the basic strategy is presented in Figure 2. The approach is a modification of older backcross introgression breeding strategies (Frey et al 1975, Harlan 1975, 1976, Hawkes 1958, Peloquin 1983, Rick 1967, 1983) in combination with molecular marker analysis (coined "advanced backcross (AB)-QTL analysis) as first described by Tanksley and Nelson (1996). All of the crossing and population development were accomplished by collaborating scientists in the country where each of the domesticated rice varieties originated, and all of the genotyping and QTL analysis were performed at Cornell University by students or visiting scientists from each participating institution.

BC_2F_2 families were grown in replicated field trials in each participating program and evaluated using common protocols for yield and yield components, grain

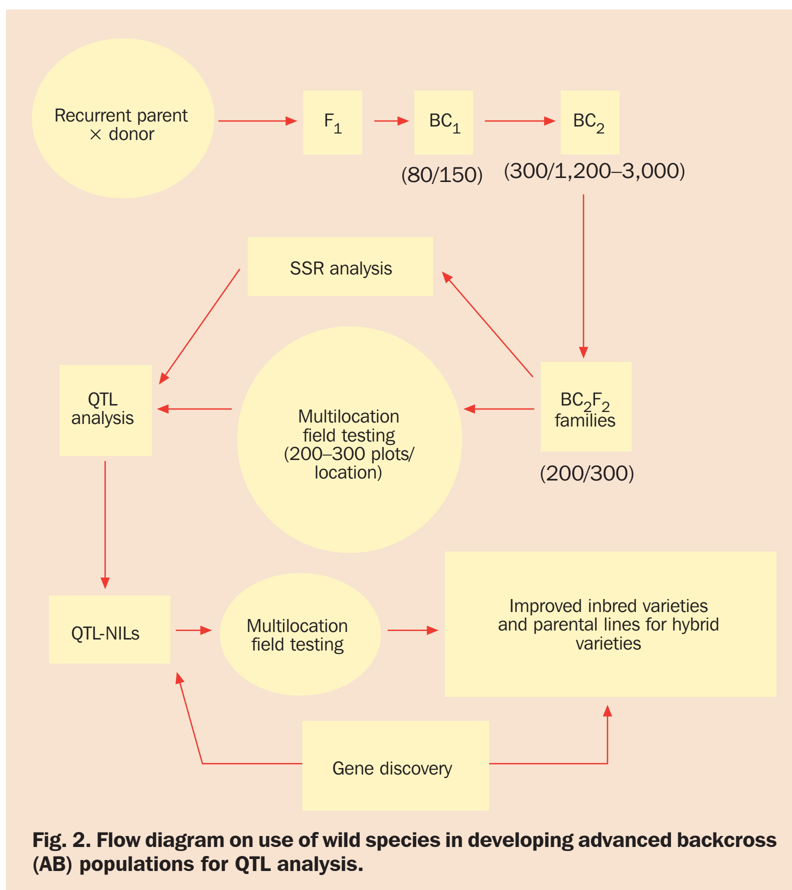


Fig. 2. Flow diagram on use of wild species in developing advanced backcross (AB) populations for QTL analysis.

quality, and other traits of importance to rice production. Remnant seed from each of the same families was sent to Cornell for DNA extraction, and the DNA was then subjected to analysis with a common set of molecular markers (McCouch et al 2002, Temnykh et al 2001). A molecular linkage map was constructed for each population. The segregation data from the molecular markers were used in combination with the phenotypic scores for each family to identify quantitative trait loci (QTL) associated with each of the traits that had been evaluated in the field (Cho et al 2003, Moncada et al 2001, Septiningsih et al 2003a,b, Thomson et al 2003, Xiao et al 1998).

AB-QTL analysis has been used with modifications by researchers working on a variety of crop species, for example, tomato (Bernacchi et al 1998), wheat (Huang et al 2003), barley (Pillen et al 2003), and pepper (Rao et al 2003). One of the main advantages of the approach is that QTL detection is concurrent with the production of improved varieties/inbreds. Negative selection is practiced in the BC_1 generation to eliminate plants with characteristics that are obviously unacceptable from an agro-

nomic perspective (i.e., sterility, photosensitivity or very late flowering, extreme plant height), but selection at this stage is minimal (i.e., not extreme). The value of waiting until the BC_2 generation to undertake replicated field trials is that BC_2F_2 families are skewed toward alleles from the recurrent parent, minimizing the possibility that favorable phenotypic effects resulting from epistatic interactions among donor QTLs will be detected, and providing a higher probability of detecting additive QTLs that will function as predicted when isolated in the nearly isogenic background of the recurrent parent. Finally, the AB population makes it possible to rapidly produce lines that are nearly isogenic for selected QTLs (QTL-NILs, near-isogenic lines). QTL-NILs (or hybrids produced with QTL-NILs) are candidates for new varieties with one or more enhanced attributes. Hence, unlike with traditional QTL analysis, which can require up to 5 years of breeding following QTL analysis (Steele et al 2005), AB analysis can lead to the creation of candidate varieties only a few years after the initial QTL detection.

Results

Over the course of these studies, QTLs associated with an array of agriculturally valuable traits were identified (Cho et al 2003, Moncada et al 2001, Septiningsih et al 2003a,b, Thomson et al 2003, Xiao et al 1996, 1998). Those of particular interest were (1) those associated with transgressive variation, where alleles from the low-performing *O. rufipogon* (donor) parent enhanced the performance of the elite *O. sativa* recurrent parent (RP), and (2) those associated with domestication-related traits (where either positive or negative selection is required to keep or eliminate the trait from useful breeding lines). Transgressive variation (both positive and negative) was a recurring theme in all interspecific populations examined in this study, and it is noteworthy that 30–50% of the QTLs identified in each study actually improved the key traits (Cho et al 2003, Moncada et al 2001, Septiningsih et al 2003a,b, Thomson et al 2003, Xiao et al 1996, 1998). In general, a 5–20% improvement was realized for most of the characters examined.

Although many of the genes identified from the wild rice accession decreased the target traits in these AB populations, the degree of positive transgressive variation was higher than expected. Because of the coordinated approach taken among research/breeding groups participating in this project, we were able to align the QTL-introgressions associated with enhanced performance (using the CMAP and QTL modules in the Gramene database, www.gramene.org) to show that specific *O. rufipogon* introgressions were associated with superior performance for a variety of agronomic and yield-related traits across several genetic backgrounds and environments (i.e., yield, panicle size, seed size and shape, number of seeds per plant) (Fig. 3A) (Lee et al 2005, Moncada 2001, Septiningsih et al 2003a,b, Thomson et al 2003, Xiao et al 1998). On the other hand, some yield and flowering time QTLs were associated with a positive effect in one genetic background and/or environment, but not in others (Fig. 3A & B). In other cases, *O. rufipogon*-derived QTLs with opposing effects (i.e., one contributing to large seed size and another contributing to small seed

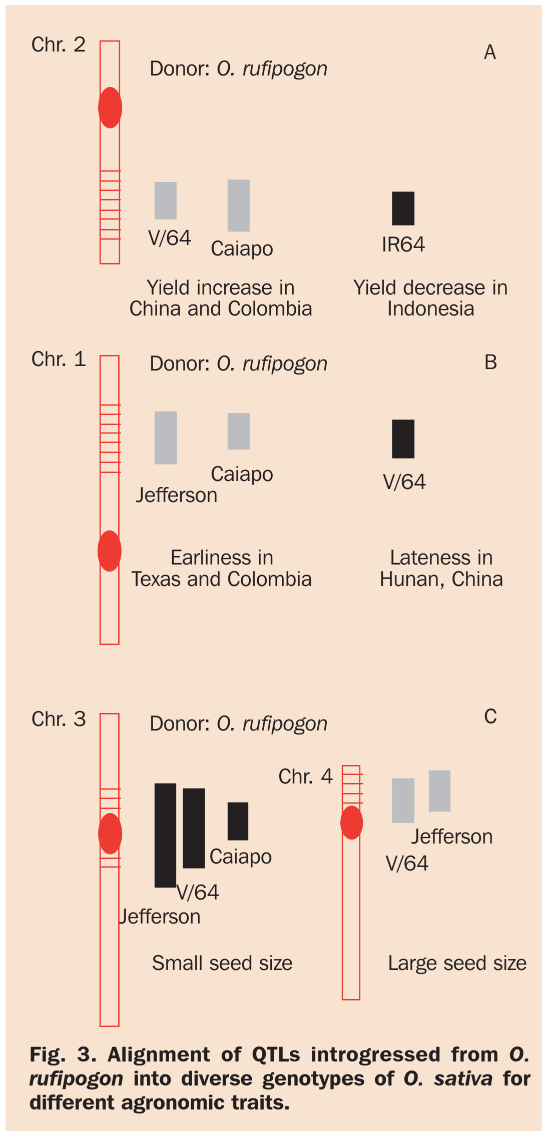


Fig. 3. Alignment of QTLs introgressed from *O. rufipogon* into diverse genotypes of *O. sativa* for different agronomic traits.

size) were identified in the same elite genetic background, with similar allele effects evident in multiple RPs (Fig. 3C).

Most of the AB-QTL analysis reported from the coordinated crossing program focused on a single donor accession of *O. rufipogon* (IRGC 105491) that was crossed with an array of different elite recurrent parents representing three of the major subpopulations of *O. sativa* (i.e., temperate japonica, tropical japonica, and indica) (Table 1). The donor was selected from a set of 25 *O. rufipogon* and *O. nivara* accessions because it crossed readily with both indica and japonica RPs as described by Xiao et al (1998). The structure of this project established a well-defined genetic system and clearly demonstrated that selected introgressions from one *O. rufipogon* accession could confer a consistent and agronomically relevant advantage in multiple genetic backgrounds grown in diverse environments. What was not clear from this work was whether these results were specific to the particular accession of *O. rufipogon* that we happened to select at the beginning of the project, or whether other, randomly selected accessions of *O. rufipogon* or other species would produce similar results.

To examine this question, we first looked at independent work undertaken by other researchers using different *O. rufipogon* donors. In parallel experiments using RPs from India, Korea, China, and the Philippines, with similar crossing and population development schemes, all reported essentially identical results, with similar levels of transgressive variation and comparative percentages of "favorable QTLs" coming from other wild/weedy accessions of *O. rufipogon* (Lee et al 2005, Marri et al 2005, Nguyen et al 2003, Xiong et al 1999). These results confirm that AB-QTL analysis using diverse *O. sativa* × *O. rufipogon* crosses offers an effective strategy for broadening the genetic base of *O. sativa* and enhancing the performance of a wide array of elite cultivars. We thus conclude that it was not simply a "lucky" choice of donor in the first instance, but that almost any accession of *O. rufipogon* can be used as a donor with similar results. We did, however, note that the accession of *O. rufipogon* used in our work (IRGC 105491, from Malaysia) was genetically more similar to the *Aus* subpopulation than to other *O. sativa* subpopulations. This may help explain why a lower proportion of positive transgressive QTLs were identified in the cross between indica cultivar IR64 and this accession of *O. rufipogon* (Septiningsih et al 2003a) than in crosses involving any of the temperate or tropical japonica RPs (Aus is genetically more similar to indica than to japonica). This suggests that knowing something about the population substructure of *O. rufipogon* might offer predictive value in selecting wild/weedy parents for crossing with particular elite *O. sativa* varieties.

The next set of questions we sought to answer were the following: At what genetic distance would it be most advantageous to look for useful variation in rice? Should we look in close *O. sativa* landrace relatives, in closely related species such as *O. rufipogon*, or in more distantly related species such as *O. glaberrima* or *O. barthii* from Africa, or *O. glumaepatula* from Brazil? Does the degree of genetic divergence provide any prediction about the likelihood of identifying positive transgressive variation in the offspring? Such questions are highly relevant to plant breeders who seek to identify the most productive materials to use as parents before investing in AB-QTL analysis.

To consider these questions, we studied the results of projects that involved crosses between elite *O. sativa* parents and the more diverse accessions of *O. glaberrima, O. barthii*, or *O. glumaepatula* using the AB-QTL strategy (Li et al 2005). In these crosses, there were considerably higher rates of sterility and a lower proportion of detected QTLs were associated with positive effects than in the crosses with *O. rufipogon*. This suggests that crosses between *O. sativa* and its immediate ancestor(s), *O. rufipogon* (and/or *O. nivara*), are more likely to generate useful variation for plant breeding than are more genetically divergent crosses.

Finally, we were interested in determining whether a breeder would be more likely to find useful genetic variability in domesticated cultivars that have undergone human selection for favorable alleles (i.e., landraces of *O. sativa* or *O. glaberrima*) or in genetically more distant relatives that are phenotypically inferior (i.e., wild species). Work initiated at IRRI has reported success using AB-QTL analysis with intraspecific populations derived from crosses between elite cultivars and divergent landraces of *O. sativa* (Li et al 2005). Numerous transgressive segregants with favorable phenotypes were identified in that study, but we have not been able to compare the frequency of detecting favorable transgressive QTLs or the magnitude of the allele effects associated with introgressions from diverse landraces compared to those from wild/weedy *O. rufipogon*.

It is interesting to speculate that there might be fewer crossing barriers encountered in *O. rufipogon* × *O. sativa* crosses than in crosses between subpopulations of *O. sativa*, particularly if the accumulation of Dobzhansky-Muller incompatibilities has been accentuated over the course of domestication. This would be consistent with the acute sterility problems associated with most indica × japonica crosses. The use of *O. rufipogon* as a donor can also be expected to increase the rate at which novel allelic variation (i.e., variation not formerly present in the cultivated gene pool) is discovered and would contribute more efficiently to expanding the genetic base of cultivated rice. *O. sativa* landraces have undergone a substantial genetic bottleneck associated with domestication (Sun et al 2002) and are inherently less genetically diverse than *O. rufipogon* populations. Landraces would tend to be genetically more similar to elite *O. sativa* varieties, making it less likely to identify transgressive segregants among the offspring. On the other hand, *O. sativa* landraces might contain fewer deleterious "wild" alleles and thus might generate less linkage drag. They also might be less disruptive of gene complexes that plant breeders seek to maintain. Opportunities to answer these questions and to quantify genetic differences between subpopulations of both *O. sativa* and *O. rufipogon* await further research.

In summary, the body of work using AB-QTL analysis with both interspecific and intersubpopulation crosses of rice has demonstrated that useful levels of positive transgressive variation are observed among the offspring of all of these divergent crosses and that QTLs associated with the transgressive segregants can be readily identified. Breeders and geneticists have been able to develop QTL-NILs that mimic the performance and grain quality of the initial elite cultivar coupled with the enhanced performance from the wild or exotic parent. Fine mapping to identify recombination events near the gene(s) of interest has allowed breeders to break linkage drag where it

is a problem, and the same recombinants also provide the basis for positional isolation of the genes underlying the target QTLs.

Identifying genes underlying QTLs for domestication and transgressive variation

Grain size

Rice grain length and shape are important to consumers because they determine the physical appearance and affect the cooking quality of the grain. Seed and grain weight are important to farmers because they are among the most stable components of yield in rice. Furthermore, seed size or weight is important in the evolution of cereal crops because humans tended to select for large seed size during the early domestication process, as evidenced by the fact that most cultivated species have larger seeds than their wild relatives (Doganlar et al 2000, Harlan 1992). Small seeds may be associated with reduced seedling vigor and difficult mechanical harvesting, which are problems for crop cultivation (Takeda 1991), but small seed is often favored under natural selection because it is frequently associated with a large number of seeds per plant, more rapid maturity, and wider geographic distribution.

Several independent studies in rice have identified QTLs associated with grain weight, shape (length/width), or kernel elongation after cooking. For example, a QTL associated with grain weight or length has been reported in the centromere region of rice chromosome 3 in at least ten different populations (Li et al 2005). These results suggest that the same QTL is expressed in different genetic backgrounds and environments, making it a valuable target for genetic analysis and also for further applications in rice breeding. Moreover, comparative mapping of this seed weight QTL in maize suggests that a homologous gene determining seed weight or size may be associated with domestication and subsequent selection in different species (Doebley et al 1994, Paterson et al 1995). While the positional convergence of these QTLs is suggestive, proof of structural and functional conservation of orthologs across the grasses awaits the cloning and characterization of the genes underlying these QTLs.

This study was undertaken to refine the position of a grain-weight QTL, *gw*3.1, mapped by Thomson et al (2003) to an interval of 31.8 cM in the peri-centromeric region of rice chromosome 3, and to develop a set of NILs that would provide the foundation for map-based isolation of the gene underlying this QTL. We developed large segregating populations to identify informative recombinants in the target region and we developed NILs to characterize the magnitude and behavior of the *O. rufipogon*-derived allele in a domesticated tropical japonica background. Further, because *gw*3.1 is adjacent to the centromere on rice chromosome 3, it was of interest to estimate the genetic:physical distance in the region and to determine whether a positional approach to gene isolation was likely to be successful.

Materials and methods. A BC_2F_2 population was constructed for QTL mapping as described by Thomson et al (2003) using an *O. sativa* subsp. tropical japonica cultivar (Jefferson) as the recurrent parent and a wild accession of *O. rufipogon* (IRGC105491 from Malaysia) as the donor parent. Jefferson is an elite tropical japonica cultivar developed for production in the southern U.S. It is semidwarf, early flowering, high-

yielding, resistant to multiple diseases, and has good grain quality (McClung et al 1997). The *O. rufipogon* donor is weedy and low yielding but crosses well to both indica and japonica varieties. From this population, one BC_2F_2 family, C126-3, was selected as the starting material for fine mapping and near isogenic line development of the QTL *gw*3.1 (Li et al 2005). The *O. rufipogon* allele was completely dominant over the Jefferson allele at the *gw*3.1 locus and the dominant allele conferred small grain weight. Selection of family C126-3 as the basis for fine mapping was based on the fact that it contained an *O. rufipogon* introgression in the target region and had significantly smaller grain weight than other BC_2F_2 individuals and relatively few nontarget background introgressions. NILs were developed by backcrossing to the Jefferson parent, followed by selfing to eliminate nontarget genomic regions.

QTL analysis in the BC_2F_2 population identified *O. rufipogon* introgressions (QTLs) that were associated with positive transgressive variation for days to heading (*dth*), seed weight, and number of grains per plant (i.e., yield) (Thomson et al 2003). QTLs associated with unwanted wild or weedy traits were also identified in the Jefferson and other RP backgrounds, including tall plant stature, red pericarp, dormancy, and shattering (Septiningsih 2002, Sweeney et al 2006). It was of immediate interest to evaluate NILs containing introgressions associated with earliness and higher yields and it was of equal interest to identify the gene(s) underlying the deleterious wild or weedy-related traits. Knowledge about the genetic identity of both positive and negative targets of selection provides the basis for more efficiently using *O. rufipogon* as a donor for rice improvement. Additional details regarding phenotypic and genotypic analysis are described in Li et al (2005).

Transgressive segregation for grain weight. As reported by Thomson et al (2003), *O. rufipogon* introgressions at six QTL locations on chromosomes 2, 3, 9, 10, and 12 decreased grain weight whereas those at two QTL locations on chromosomes 1 and 5 increased grain weight (Fig. 4). The fact that multiple QTLs with opposing effects originally resided in the same parental line explains why "Mendelizing" the trait using NILs can give rise to lines that show a more extreme phenotype than either parent (Fig. 5). Interestingly, the NIL containing a single *O. rufipogon* introgression at *gw*3.1 also had a higher R^2 value for grain weight and length associated with the most closely linked marker, JL109 (40–56%), compared to R^2 values of 10–15% associated with the same QTL in the primary QTL analysis (Thomson et al 2003). This is because JL109 was more closely linked to the gene underlying the QTL than were the markers used in the primary QTL analysis, and because NIL comparisons offer increased resolution (less genetic noise) than does primary QTL analysis, in which groups of individuals containing multiple "background" introgressions are grouped for means comparisons.

In Figure 5, an *O. rufipogon* introgression at the QTL *gw*3.1 on chromosome 3 is sufficient to confer small seed size in combination with Jefferson alleles at all other QTL loci, while transgressive variation for large seed size is associated with an *O. rufipogon* introgression on either chromosome 1 or 5. At these loci, the wild QTL alleles are associated with increased grain size compared with the Jefferson alleles. This example clearly illustrates one explanation for transgressive variation that is often

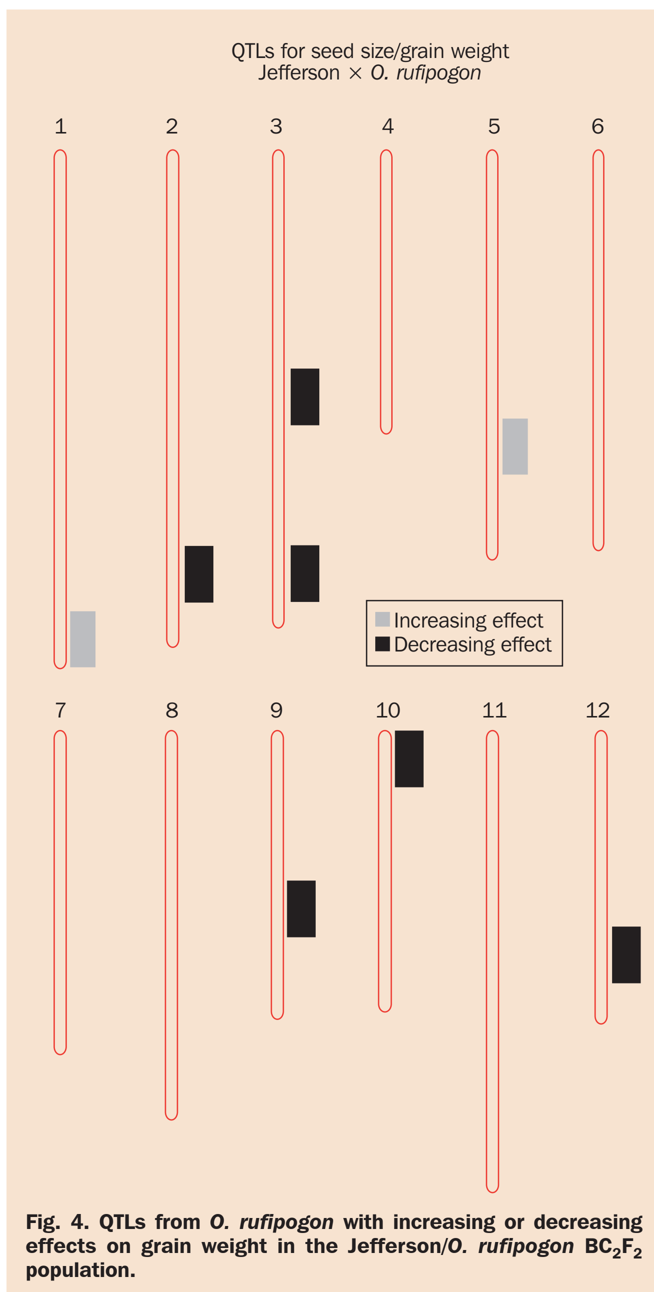


Fig. 4. QTLs from *O. rufipogon* with increasing or decreasing effects on grain weight in the Jefferson/*O. rufipogon* BC_2F_2 population.

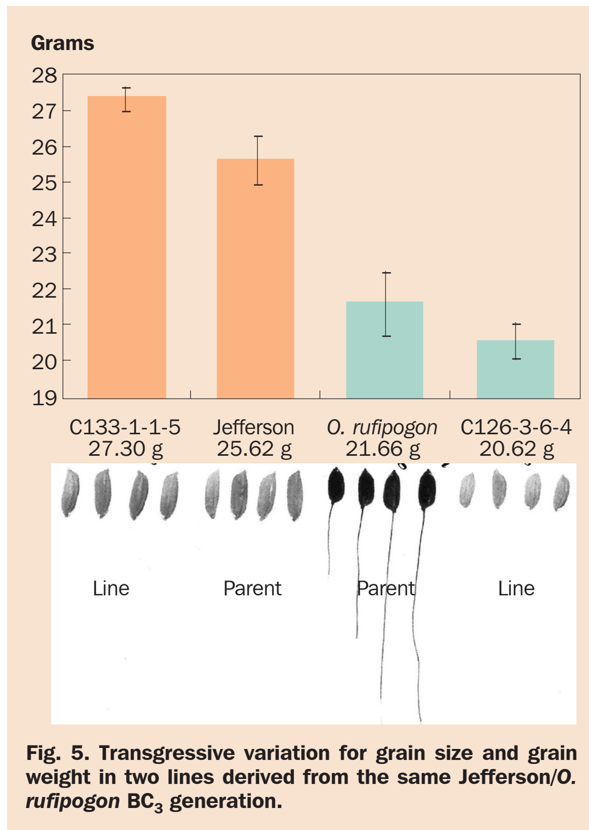


Fig. 5. Transgressive variation for grain size and grain weight in two lines derived from the same Jefferson/*O. rufipogon* BC_3 generation.

observed in divergent crosses: QTL alleles with similar effects that were originally dispersed among the parental lines are recombined into one genetic background in the progeny. Based on a review of the phenomenon in both plants and animals, this is the most common explanation for transgressive variation (Reiseberg et al 2003). The example also suggests that the gene underlying the *gw*3.1 QTL is likely to represent an early actor in the gene network controlling for seed size in rice, since it has the power to depress grain weight/seed size, even when combined with five QTL alleles from Jefferson that are associated with large grain weight/seed size.

Red pericarp

Most rice that is grown and consumed throughout the world has white pericarp, but rice can also produce grains with brown, red, and purple pericarp. The color is visible when the grains are dehulled, but can be removed by polishing to reveal the white endosperm. Red pericarp is ubiquitous among the wild ancestors of cultivated rice (*O. rufipogon*) and in some regions of the world red cultivars are preferred for their taste, texture, and ceremonial or medicinal value. Consumer interest in red and purple

rice represents a growing specialty market in the United States but, at the same time, the constant presence of weedy red rice in farmers' fields is the most economically important weed and grain quality problem faced by rice producers who use direct seeding (Gealy et al 2002). Weedy rice can be either *O. sativa* or *O. rufipogon* (Vaughan et al 2001). These red rice individuals interbreed freely with cultivated white-grained types and exhibit increased dormancy and shattering, allowing them to persist in rice fields despite vigorous attempts to remove them.

The red pigment in rice grains is proanthocyanidin, also called condensed tannins (Oki et al 2002). Proanthocyanidins are a branch off the anthocyanin pathway and share many of the same biosynthetic genes (Winkel-Shirley 2001). Proanthocyanidins have been shown to have important deterrent effects on pathogens and predators, so it is not surprising that spontaneous mutations that inhibit pigment production would be selected against in the wild (Shirley 1998). On the other hand, white grain is associated with the domestication syndrome and remains under strong selection in most rice breeding programs today.

Regardless of all the problems associated with red rice as a weed, the red pigment is of interest for nutritional reasons. It serves as a powerful antioxidant that has been demonstrated to reduce atherosclerotic plaque formation, a risk factor associated with cardiovascular disease (Ling et al 2001). On the negative side, proanthocyanidin pigments reduce the bioavailability of iron, protein, and carbohydrates (Carmona et al 1996, Eggum et al 1981, Glahn et al 2002), which has important implications for people with low nutritional status. A better understanding of the genetics and molecular biology of grains with red pericarp and the association of this character with other wild/weedy traits will provide important information for better management of both the negative and positive features associated with red rice.

Two loci have been identified using classical genetic analysis, *Rc* (brown pericarp and seed coat) and *Rd* (red pericarp and seed coat). When present together, these loci produce red seed color (Kato et al 1928). *Rc* in the absence of *Rd* produces brown seeds, while *Rd* alone has no phenotype. Although *Rc* is referred to as a mutant allele because its phenotype differs from common rice cultivars (which are considered to represent the "wild type"), the action of *Rc* is dominant over white pericarp (*rc*). This suggests that the modern cultivated (white) allele might actually be the mutant (nonfunctional) version of the ancestral *O. rufipogon* (red) allele. Both loci have been mapped using standard two-point analysis on the morphological map of rice: *Rc* on chromosome 7 and *Rd* on chromosome 1.

Proanthocyanidin biosynthesis is a branch of the anthocyanin pigment biosynthetic pathway, a well-studied system in multiple species due to its visible phenotype and lack of detrimental effects on the plant. These pathways are regulated by homologs with similar functions in different species. This provided us with an array of candidate genes and gene families that were useful in our attempts to clone the gene associated with red pericarp in rice.

We have recently reported the cloning of a basic domain helix-loop-helix (bHLH) gene underlying a QTL for rice pericarp color (Sweeney et al 2006). The QTL co-

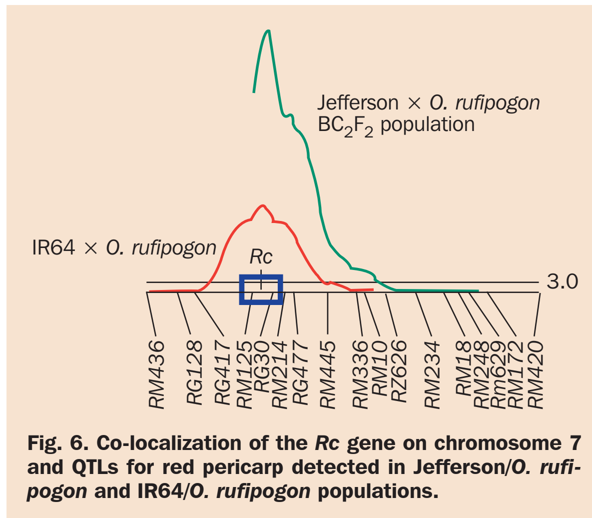


Fig. 6. Co-localization of the *Rc* gene on chromosome 7 and QTLs for red pericarp detected in Jefferson/*O. rufipogon* and IR64/*O. rufipogon* populations.

localizes with the mutant *Rc*. A frame shift deletion before the bHLH domain results in a knockout of proanthocyanidin production, leading to white rice.

Materials and methods. QTL mapping in several BC_2F_2 populations derived from crosses between tropical japonica cultivar Jefferson or indica cultivar IR64 and *O. rufipogon* (IRGC105491) identified a single, significant QTL associated with red grain (*rg* 7.1) near the centromere on chromosome 7 (Fig. 6). The LOD score associated with the *rg*7.1 QTL peaks was 99 and 33, respectively, in the two populations and the QTL was consistently detected in multiple environments (Septiningsih et al 2003a). The QTL peaks corresponded to the previously mapped position of the mutant locus, brown pericarp, *Rc* (Kinoshita 1998). All of the BC_2F_1 plants had red seeds, indicating that the *rg*7.1 locus was dominant for red color, with the dominant allele donated by the *O. rufipogon* parent.

Results. The gene underlying the QTL *rg*7.1 was fine-mapped using 1,410 BC_2F_3 plants. These plants were genotyped using markers flanking the QTL and recombinants were phenotyped for grain color. Several new SSR and indel markers were developed to help define breakpoints across the region and these were then used to look for recombinants in a population of 4,000 BC_2F_6 plants. Three classes of informative recombinants were identified, which narrowed the *rg*7.1 QTL to an 18.5-kb region. One nontransposable element gene, LOC_Os07g11020.1, was detected within the 18.5-kb target region. This was a single-copy gene, 668 aa in length, and it contained a predicted bHLH domain, which is commonly found among transcription factors known to regulate pigment synthesis. A recombination breakpoint eliminated the promoter region and the first two exons of the bHLH gene as the source of the functional nucleotide polymorphism, leaving only exons 3–7 within the 18.5-kb target region.

Sequencing of the bHLH locus in both mapping parents, *O. sativa* (cv. Jefferson) and *O. rufipogon,* allowed us to look for sequence changes that could explain the observed change in pericarp color. Six indels and 22 SNPs were detected across the genomic sequence. We also compared the sequences of the mapping parents with the publicly available cv. Nipponbare sequence. The Jefferson allele was identical to the Nipponbare sequence, both of which are japonica cultivars having white seeds. When the polymorphisms between Jefferson and *O. rufipogon* are aligned with the appropriate gene model, ten of the sequence polymorphisms fall within the coding sequence and five of those are expected to affect protein sequence. To help identify which of the sequence polymorphisms between the parents was responsible for the altered function of the gene, we also sequenced the bHLH locus in H75, an *Rc* mutant stock belonging to the japonica subspecies. H75, like *O. rufipogon,* carries a functional allele, but it is much more closely related to Jefferson than to *O. rufipogon.* Thus, a sequence comparison between H75 and Jefferson was expected to help eliminate some of the nonfunctional polymorphisms detected between the parents in the bHLH gene.

We found that the coding sequence of the bHLH allele in H75 was identical to the Jefferson sequence except for a 14-bp indel in exon 6. This 14-bp sequence was present in the H75 stock as well as in *O. rufipogon,* but it was deleted in Jefferson and Nipponbare. The deletion induces a frame shift in the sequence, resulting in two premature stop codons before the end of exon 6. The stop codons truncate the protein prior to exon 6, which contains the bHLH domain. Given that this deletion was the only difference between the alleles of LOC_Os07g11020.1 in the H75 mutant stock (pigmented seeds) and the japonica cultivars, Jefferson and Nipponbare (white seeds), that its location in exon 7 is consistent with the recombination data, and that it would have a clear and important impact on gene function, we conclude that the 14-bp deletion is the reason for the lack of pigment in the pericarps of Jefferson and Nipponbare seeds.

To examine the timing and localization of the *Rc* transcript, we used RT-PCR to amplify mRNA from leaves, young panicles (before fertilization), pericarps of young seeds (at the milk or dough stages of grain filling), and pericarp from mature seeds. The mRNA was collected from both Jefferson (white seeds) and *O. rufipogon* (red seeds) plants. RT-PCR showed no expression of *Rc* in leaf tissue, as expected for a gene associated with a seed phenotype; however, expression was seen in several stages of panicle development. Since the promoter of the bHLH gene had been eliminated as the source of polymorphism based on the recombination data, we anticipated that similar expression levels of *Rc* would be detected in red and white seeds. Our results confirmed this expectation and further demonstrated that the RNA transcript from Jefferson contained the 14-bp deletion predicted from the sequence information.

Implications of cloning the Rc *gene.* Red pericarp has long been used as a marker for the cluster of domestication traits associated with weedy rice, including dormancy and shattering (Gu et al 2005). Several studies have placed QTLs for dormancy and shattering in the pericentromeric region of chromosome 7 encompassing the *Rc* locus. With the cloning of *Rc*, it is now possible to ask whether this association is the result of

linkage or pleiotropy. Given the reduced rate of recombination within the *rg*7.1 QTL, it is logical that genes for shattering, dormancy, and pericarp color have simply hitch-hiked together in a linkage block. It is also possible that *Rc* acts pleiotropically, as do many of the other bHLH proteins associated with pigmentation. Using the recombinant lines generated in this work, we will be able to test these different hypotheses.

Rice and wheat (*Triticum aestivum*) are similar in that red pericarp in both species can be eliminated by one locus. Comparative mapping shows no homoeology between the position of the *Rc* gene in rice and the *R* gene controlling red pericarp in wheat. A reverse genetics approach also failed to locate any ESTs from wheat that map to the *Rc* locus, although this is not surprising, since no rice or maize ESTs have been found for this locus either. Our work confirms that the *Rc* transcript amplifies only with a high-sensitivity *Taq* polymerase and this suggests that low transcript abundance may also explain the lack of EST hits in wheat. The *R* locus in wheat may be orthologous to the *Rd* gene in rice, given their homoeologous positions on wheat chromosome 3 and rice chromosome 1. Although the systems look similar phenotypically, molecular data suggest that the mutation leading to white pericarp occurred at different points in the pathway.

The cloning of *Rc* will make possible new methods of fighting weedy rice infestations in rice paddies. Red rice is a noxious weed that is currently responsible for losses of as much as US$50 million per year in the U.S. (Gealy et al 2002). It is a perfect mimic of elite varieties, as the red pericarp is not visible until after harvest, when the grains are dehulled. Furthermore, the close association among red pericarp, seed shattering, and dormancy allows it to persist in fields for years despite vigorous attempts to remove it. The fact that the pericarp is maternal tissue, so that its color is dependent on the maternal genotype, means that seeds pollinated by red rice can be white (if the maternal plant carries the *rc* allele), but plants grown from these seeds will produce red seeds.

An immediate application of the work presented here involves the use of "perfect markers" that specifically target the 14-bp functional nucleotide polymorphism within the bHLH gene to screen for red rice contamination within certified seed lots. This will also facilitate the use of genes derived from crosses with wild relatives by allowing breeders to conclusively select against progeny carrying *Rc,* and to do so before the plants set seed.

References

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. 1997 An apportionment of human DNA diversity. Proc. Natl. Acad. Sci. USA 94:4516-4519.

Bernacchi D, Beck-Bunn T, Emmatty D, Inai S. 1998. Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desireable wild QTL alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor. Appl. Genet. 97:170-180.

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

Carmona A, Borgudd L, Borges G, Levy-Benshimol A. 1996. Effect of black bean tannins on in vitro carbohydrate digestion and absorption. J. Nutr. Biochem. 7:445-450.

Chang TT. 1976. The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25:435-441.

Cheng CY, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E. 2003. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol. Biol. Evol. 20:67-75.

Cho YC, Suh JP, Choi IS, Hong HC, Baek MK, et al. 2003. QTLs analysis of yield and its related traits in wild rice relative *Oryza rufipogon.* Treat. Crop Res. Korea 4.

Choi HO, Bae SH, Chung GS, Cho CY. 1974. A new short-statured rice variety "Tongil". Res. Rep. Rural Development Office, Suweon, Korea 16:1-12.

Dalmacio RD, Brar DS, Ishii T, Sitch TA, Virmani SS, Khush GS. 1995. Identification and transfer of a new cytoplasmic male sterility source from *Oryza perennis* into *indica* rice (*O. sativa*). Euphytica 82:221-225.

Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21:113-135.

Doebley J, Bacigalupo A, Stec A. 1994. Inheritance of kernel weight in two maize-teosinte hybrid populations: implications for crop evolution. J. Heredity 85:191-195.

Doganlar S, Frary A, Tanksley S-D. 2000. The genetic basis of seed-weight variation: tomato as a model system. Theor. Appl. Genet. 100:1267-1273.

Edwards JD. 2005. Origins and distribution of allelic diversity in populations of wild and cultivated rice and phenotypic consequences of admixture at a complex flowering time locus. PhD thesis. Cornell University, USA. 119 p.

Eggum B, Alabata E, Juliano B. 1981. Protein utilization of pigmented and nonpigmented brown and milled rice by rats. Qual. Plant Plant Foods Hum. Nutr. 31:175-179.

Frey KJ, Hammond EG, Lawrence PK. 1975 Inheritance of oil percentage in interspecific crosses of hexaploid oats. Crop Sci. 15:94-95.

Garris AJ, Tai TH, Coburn JR, Kresovich S, McCouch S. 2005 Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631-1638.

Gealy DR, Tai TH, Sneller CH. 2002. Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Sci. 50:333-339.

Glahn R, Cheng Z, Welch R. 2002. Comparison of iron bioavailability from 15 rice genotypes: studies using an in vitro digestion/CAco-2 cell culture model. J. Agric. Food Chem. 50:3586-3591.

Glaszmann JC. 1987. Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. 74:21-30.

Harlan J. 1992. Crops and man. Madison, Wis. (USA): American Society of Agronomy, Inc.; Crop Science Society of America, Inc. 284 p.

Harlan JR. 1975. Our vanishing genetic resources. Science 188:618-621.

Harlan JR. 1976. Genetic resources in wild relatives of crops. Crop Sci. 16:329-333.

Harushima Y, Nakagahra M, Yano M, Sasaki T, Kurata N. 2002. Diverse variation of reproductive barriers in three intraspecific rice crosses. Genetics 160:313-322.

Hawkes JG. 1958. Significance of wild species and primitive forms for potato breeding. Euphytica 7:257-270.

Huang XQ, Coster H, Ganal MW, Roder MS. 2003. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 106:1379-1389.

Kato S, Kosaka H, Hara S. 1928. On the affinity of rice varieties as shown by the fertility of rice plants. Central Agricultural Institute of Kyushu Imperial University 2:241-276.
Khush GS. 2000. Taxonomy and origin of rice. In: Singh RK, Singh US, Khush GS, editors. Aromatic rices. New Delhi (India): Oxford & IBH Publishing Co. Pvt. Ltd. p 5-13
Kinoshita T. 1998. Linkage mapping using mutant genes in rice. Rice Genet. Newsl. 15:13-74.
Lee SJ, Oh CS, Suh JP, McCouch SR, Ahn SN. 2005. Identification of QTLs for domestication-related and agronomic traits in an *Oryza sativa* x *O. rufipogon* BC_1F_7 population. Plant Breed. 124:209-219.
Li J, Yuan L. 2000. Hybrid rice: genetics, breeding, and seed production. Plant Breed. Rev. 17:150-158.
Li Z, Fu B, Gao Y, Xu J, Ali J, Lafitte H, Jiang Y, Rey J, Yijayakumar C, Maghirang R, Zheng T, Zhu L. 2005. Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). Plant Mol. Biol. 59:33-52.
Ling WH, Cheng QX, Ma J, Wang T. 2001. Red and black rice decrease atherosclerotic plaque formation and increase antioxidant status in rabbits. J. Nutr. 131:1421-1426.
Londo J, Chiang Y, Hung K, Chiang T, Schaal B. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc. Natl. Acad. Sci. USA 103:9578-9583.
Lu BR, Zheng KL, Qian HR, Zhuang JY. 2002. Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theor. Appl. Genet. 106:101-106.
Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR, Tai T. 2005. Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. Crop Sci. 45:66-76.
Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. USA 101:12404-12410.
Marri PR, Sarla N, Reddy LV, Siddiq EA. 2005. Identification and mapping of yield and yield related QTLs from an Indian accession of *Oryza rufipogon.* BMC Genetics 6:33.
Matsuo T, Futsuhara Y, Kikuchi F, Yamaguchi H. 1997. Science of the rice plant. Volume 3. Tokyo (Japan): Food and Agriculture Policy Research Center.
Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA 99:6080-6084.
McClung AM, Marchetti MA, Webb BD, Bollich CN. 1997. Registration of "Jefferson" rice. Crop Sci. 37:629-630.
McCouch S, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, et al. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199-207.
Moncada M, Martínez C, Tohme J, Guimaraes E, Chatel M, Borrero J, Gauch H, McCouch S. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. 102:41-52.
Morishima H, Sano Y, Oka HI. 1984. Differentiation of perennial and annual types due to habitat conditions in the wild rice *Oryza perennis*. Plant Syst. Evol. 144:119-135.
Muller HJ. 1942. Isolating mechanisms, evolution and temperature. Biol. Symp. 6:71-125.
Nagaraju J, Kathirvel M, Kumar RR, Siddiq EA, Hasnain SE. 2002. Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. Proc. Natl. Acad. Sci. USA 99:5836-5841.

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

Nguyen BD, Brar DS, Bui BC, Nguyen TV, Pham LN, Nguyen H. 2003. Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). Theor. Appl. Genet. 106:583-593.

Ni J, Colowit PM, Mackill DJ. 2002. Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci. 42:601-607.

Oka HI. 1988a. Functions and genetic base of reproductive barriers. In: Oka HI, editor. Origin of cultivated rice. Amsterdam (The Netherlands): Tokyo/Elsevier Science/Japan Scientific Societies Press. p 156-159, 181-209.

Oka HI. 1988b. Origin of cultivated rice. Tokyo (Japan): Japan Scientific Societies Press. 254 p.

Oka HI, Morishima H. 1967. Variations in the breeding systems of a wild rice, *Oryza perennis*. Evolution 21:249-258.

Oki T, Masuda M, Kobayashi M, Nishiba Y, Furuta S, Suda I, Sato T. 2002. Polymeric procyanidins as radical-scavenging components in red-hulled rice. J. Agric. Food Chem. 50:7524-7529.

Paterson A-H, Lin Y-R, Li Z, Schertz K-F, Doebley J-F, Puison S, Liu S, Stansel J, Irvine J. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714-1718.

Peloquin SJ. 1983. Utilization of exotic germplasm in potato breeding: germplasm transfer with haploids and 2n gametes. In: Brown WL, editor. Conservation and utilization of exotic germplasm to improve varieties. The 1983 Plant Breeding Research Forum, 9-11 August, Heber Springs, Arkansas. Des Moines (Iowa): Pioneer Hi-Bred International. p 147-167.

Pillen K, Zacharias A, Leon J. 2003. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 107:340-352.

Rao GU, Ben Chaim A, Borovsky Y, Paran I. 2003. Mapping of yield-related QTLs in pepper in an interspecific cross of *Capisicum annum* and *C. frutescens*. Theor. Appl. Genet. 106:1457-1466.

Reiseberg LH, Widmer A, Arntz AM, Burke JM. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Philos. Trans. Soc. Lond. 358:1141-1147.

Rick CM. 1967. Exploiting species hybrids for vegetable improvement. Proc. XVII Int. Hort. Congr. 3:217-226.

Rick CM. 1983. Conservation and use of exotic tomato germplasm. In: Brown WL, editor. Conservation and utilization of exotic germplasm to improve varieties. The 1983 Plant Breeding Research Forum, 9-11 August, Heber Springs, Arkansas. Des Moines (Iowa): Pioneer Hi-Bred International. p 147-167.

Roy SC. 1921. A preliminary classification of the wild rices of the Central Province and Berar. Agric. J. India 16:365-380.

Sano Y. 1993. Constraints in using wild relatives in breeding: lack of basic knowledge on crop gene pools. In: Buxton DR, editor. International crop science. Madison, Wis. (USA): Crop Science Society of America. p 437-443.

Second G. 1982. Origin of the genetic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Jpn. J. Genet. 57:25-57.

Second G. 1999. Molecular markers in rice systematics and the evaluation of genetic resources. In: Bajaj YPS, ed. Biotechnology in agriculture and forestry. New York (USA): Springer, Berlin Heidelberg. 14:468-494.
Septiningsih EM. 2002. Identification, near-isogenic line development and fine mapping of quantitative trait loci from the rice cultivar IR64 and its wild relative *Oryza rufipogon*. Ph.D. thesis. Cornell University, Ithaca, New York, USA.
Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch S. 2003a. Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theor. Appl. Genet. 107:1419-1432.
Septiningsih EM, Trijatmiko KR, Moeljopawiro S, McCouch SR. 2003b. Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theor. Appl. Genet. 107:1433-1441.
Shirley B. 1998. Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. Seed Sci. Res. 8:415-422.
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.
Steele KA, Price AH, Shashidhar HE, Witcombe JR. 2005. Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theor. Appl. Genet. 1-14.
Sun CQ, Wang XK, Yoshimura A, Doi K. 2002. Genetic differentiation for nuclear mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). Theor. Appl. Genet. 104:1335-1345.
Sweeney M, Pfeil B, McCouch SR. 2006. Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283-294.
Takeda K. 1991. Inheritance of grain size and its implications for rice breeding. In: Rice genetics II. Los Baños (Philippines): International Rice Research Institute. p 181-189.
Tanksley SD, Nelson JC. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92:191-203.
Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11:1441-1452.
Thomson MJ, Tai TH, McClung AM, Lai X, Hinga ME, Lobos KB, Xu Y, Martinez C, McCouch S. 2003. Mapping quantitative trait loci for yield, yield components, and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor. Appl. Genet. 107:479-493.
Vaughan DA, Kadowaki K-I, Kaga A, Tomooka N. 2005. On the phylogeny and biogeography of the genus *Oryza*. Breed. Sci. 55:113-122.
Vaughan LK, Ottis BV, Prazak-Havey AM, Bormans CA, Sneller C, Chandler J, Park W. 2001. Is all red rice found in commercial rice really *Oryza sativa*? Weed Sci. 49:468-476.
Wang ZY, Tanksley SD. 1989. Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32:1113-1118.
Winkel-Shirley B. 2001. Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 126:485-493.

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

Xiao J, Li J, Yuan L, Tanksley SD. 1996. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific cross. Theor. Appl. Genet. 92:230-244.

Xiong LZ, Liu KD, Dai XK, 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.

Young ND, Tanksley SD. 1989. RFLP analysis of the size of chromosomal segments retained around the *tm-2* locus of tomato during backcross breeding. Theor. Appl. Genet. 77:353-359.

Zhang Q, Maroof MAS, Lu TY, Shen BZ. 1992. Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. Theor. Appl. Genet. 83:495-499.

Zheng HG, Qian HR, Shen BZ, Zhuang JY, Lin HX, Lu J. 1994. RFLP-based phylogenetic analysis of wide compatibility varieties in *Oryza sativa* L. Theor. Appl. Genet. 88:65-69.

Notes

Authors' addresses: S.R. McCouch, M. Sweeney, J. Li, H. Jiang, M. Thomson, E. Septiningsih, J. Edwards, P. Moncada, J. Xiao, J. Coburn, A. Garris, and T. Tai, Department of Plant Breeding and Genetics, Cornell University, 162 Emerson Hall, Ithaca, NY 14853, USA, e-mail: srm4@cornell.edu; C. Martínez and J. Tohme, Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia; M. Sugiono, ICABIOGRAD, 16111 Bogor, Indonesia; L.P. Yuan, China National Hybrid Rice Research and Development Center, Changsha, Hunan 41025, China; S.N. Ahn, Department of Agronomy, College of Agriculture, Chungnam National University, Daejeon, Korea.

Genomics-based strategies for the development of "green super rice"

Qifa Zhang

Several challenges need to be met for sustainable rice production in China and to reduce the gap between potential yield and yield under large-scale production: (1) the increasingly severe occurrence of insects and diseases and the indiscriminate application of pesticides, (2) high pressure for a yield increase and overuse of fertilizers, and (3) the increasingly frequent occurrence of drought, resulting in water shortage. We have been using a combination of approaches based on recent advances in genomics research to address these challenges, with the long-term goal of developing rice cultivars referred to as "green super rice." To obtain a yield increase and improve quality, green super rice should possess resistance to multiple insects and diseases, high nutrient efficiency and drought tolerance, and potential to greatly reduce the use of pesticides, chemical fertilizers, and water. Most current efforts have focused on identifying germplasm and discovering genes for improving rice cultivars for the following traits: resistance to diseases and insects, N and P efficiency, and drought tolerance. Approaches adopted include (1) screening of germplasm collections, (2) mapping and identifying QTLs, (3) screening of mutant libraries, (4) microarray analysis of genes differentially regulated, and (5) functional tests of candidate genes by transgenic analysis. Progress toward the development of "green super rice" currently made in our group is presented.

Keywords: green super rice, genomics, candidate genes, gene identification, yield, biotic/abiotic stresses

China is the largest rice-producing country by quantity in the world. In the last 50 years, rice yield in China has gone through two cycles of breakthroughs, primarily as a result of genetic improvement: increasing the harvest index by reducing plant height through use of the semidwarf gene, and using heterosis by producing hybrids. Consequently, rice yield tripled from the 1960s to 1990s (http://faostat.fao.org/faostat/collections?subset=agriculture). To further increase the yield ceiling of rice cultivars, rice breeders in China have been working to develop "super hybrid rice" in the last decade, aiming to raise the yield potential, and have made significant progress.

However, many challenges need to be met in order to achieve sustainable rice production, and hence sustainable agriculture. The first is the increasingly severe occurrence of insects and diseases in almost all the rice-producing areas of the country, which causes a large yield loss. In general, three diseases, bacterial blight caused by *Xanthomonus oryzae* pv. *oryzae* (*Xoo*), blast caused by *Pyricularia grisea*, and sheath blight caused by *Rhizoctonia solani*, are considered as the most devastating diseases in most rice-producing areas of China. Similarly, three groups of insects, stem borers (yellow stem borer *Tryporyza incertula*s and striped stem borer *Chilo suppressalis*), leaffolders (*Marasmia patnalis* and *Cnaphalocrocis medinalis*), and planthoppers (mostly brown planthopper, *Nilaparvata lugens*), have been the most damaging pests. For a long time, control of the diseases and insects has depended heavily on indiscriminate applications of chemical pesticides. While such an intensive use of chemicals has created serious environmental pollution and hazards to the health of producers and consumers alike, chemical control is not very effective and crop loss still occurs frequently in almost all the rice-producing areas. For example, stem borers occur on approximately 15 million ha annually in China (Sheng et al 2003). Although on average 2 to 3 sprays were applied per year in the affected fields, which cost a total of approximately 5 billion yuan, damage to the crop by residual insects was still estimated to be 6.5 billion yuan. Thus, a total loss of approximately 11.5 billion yuan was estimated in the country, which is a considerable economic burden on farmers.

The second challenge results from fertilizer applications. The high pressure for a yield increase and the widespread cultivation of high-yielding cultivars have led to dramatic increases in fertilizer applications in the last 40 years. Statistics showed that, in 2002, China used approximately 30% of N and P fertilizers produced worldwide (Zhu et al 2005). Peng et al (2002) compared China with other major rice-producing countries for N fertilizer efficiency and found that the average rate of N application for rice production in China is high and fertilizer N-use efficiency is low. Overfertilization has greatly reduced the economic return of fertilizers, which places a heavy economic burden on farmers. Such overfertilization has resulted in widespread pollution of the environment, including soils, groundwater, lakes, rivers, and seas (Zhu et al 2005). Moreover, overfertilization also caused the deterioration of eating and cooking quality of rice grains. Thus, developing crops that are less dependent on the heavy application of fertilizers is essential for agricultural sustainability.

A third severe challenge is water shortage and drought stress. It is estimated that the total water usage in China is approximately 556.6 billion cubic meters per year, and agriculture uses 392 billion cubic meters, accounting for 70.4% of the total water consumption of the country (Tang et al 2000). There has also been a rough estimate that rice production consumes about 70% of the water used in agriculture. However, drought stress is still the single most important constraint to rice production in many rice-producing areas of China (Lin and Shen 1996) for several reasons, including variation in rainfall patterns from one year to another and the uneven distribution of rainfall in the rice-growing season, in addition to inadequate rainfall in many areas. For example, rain in central China is usually highly concentrated in the spring and early summer, corresponding to the vegetative growth period of the middle-season rice,

which represents the predominant cropping system in this region. But rain becomes scarce in late summer, corresponding to the reproductive stage of the rice plant, frequently causing dehydration stress at the anthesis stage when the plant is most sensitive to water stress, and also at the grain-filling stage, both of which can result in severe yield loss. In southern regions with two rice crops a year, water deficiency can cause damage to the rice crop at any stage, for example, seedling period, vegetative growth period, anthesis period, and grain-filling period.

Moreover, there is also an urgent need for grain quality improvement as most of the popular cultivars and hybrids have relatively poor quality. In addition, demand for a yield increase also puts constant pressure on rice production.

The definition of "green super rice"

To provide fundamental solutions to the problems identified above, crop production in China has to reach a stage in which yield increase and quality improvement are in harmony with a sustainable environment to achieve a goal referred to as a "second green revolution." The second green revolution aims at more production with fewer inputs in an environment-friendly manner. Since rice is the most important crop in Chinese agriculture, the goal of the second green revolution requires rice to be produced with a much lower use of pesticides, fertilizers, and water, with the goal of a continuous yield increase and quality improvement. We refer to rice cultivars that meet the requirement of the second green revolution as "green super rice."

The only way to realize the goal of green super rice is through genetic improvement by developing cultivars with the desirable traits. In addition to high yield and good quality, green super rice should possess the following characteristics: (1) adequate resistance to major diseases and insects, (2) tolerance of field conditions with lowered concentrations of major nutrients, and (3) drought tolerance. Together, these define our goals and the traits needed for developing green super rice.

Progress in research toward development of green super rice

A combination of strategies has been adopted for developing green super rice by integrating germplasm, genomics, and breeding with insect and disease resistance, N and P nutrition efficiency, and drought tolerance as the target traits. The following approaches have been adopted to identify genes and germplasm for the defined traits: (1) screening of germplasm collection, (2) mapping and identifying of QTLs, (3) screening of mutant libraries, (4) microarray analysis of genes differentially regulated, and (5) functional test of candidate genes by transgenic analysis. Genes identified would be incorporated into breeding lines by either transformation or molecular marker-assisted selection. Substantial progress has been made in all respects. We will introduce the results mostly from our own work and some relevant results from other groups.

Identifying genes for developing insect-resistant rice

Resistance to stem borers and leaffolders. In the last two decades, considerable research efforts have been invested to introduce insecticidal crystal protein genes from *Bacillus thuringiensis* (*Bt*) into crops, including rice. Use of these crops has benefited growers and the environment by greatly reducing the use of chemical insecticides (James 2005). However, there is also increasing concern that widespread adoption of *Bt* crops may lead to the development of resistance to the insecticidal genes in the pest populations (Tabashnik et al 1994, Gould 1998, Frutos et al 1999). For resistance management, high-dose with refuge and gene stacking have been proposed as two effective strategies to prevent or delay the emergence of pest resistance to *Bt* toxins (Frutos et al 1999, Ferre and Van Rie 2002, Shelton et al 2002). However, the high-dose with refuge strategy does seem not applicable in most rice-producing countries, especially in China, as the majority of the rice growers work on a small scale and thus it is not feasible to designate certain areas of nontransgenic crops as refuge. The gene stacking strategy suggests that plants containing two or more dissimilar *Bt* toxins have the potential to delay resistance more effectively than ones having only a single toxin, as the insects have to develop resistance to two or more insecticides in order to survive. One requirement for this strategy to work is that the "stacked" toxins have different modes of action.

Although a large number of *Bt* toxins are known, only a small fraction of them are currently used in developing transgenic crops to control Lepidopteran pests. The most commonly used *Bt* genes in transgenic crops, including rice, are *Cry1Ab*, *Cry1Ac*, and a fusion gene of *Cry1Ac/Cry1Ab* (Perlak et al 1990, Fujimoto et al 1993, Koziel et al 1993, Nayak et al 1996, Stewart et al 1996, Wünn et al 1996, Ghareyazie et al 1997, Wu et al 1997, Cheng et al 1998, Datta et al 1998, Alam et al 1999, Tu et al 2000, Wang et al 2002). The results of binding tests of midgut brush border membrane vesicles indicated that *Cry1Aa*, *Cry1Ab*, and *Cry1Ac* toxins share a common binding site (Escriche et al 1997, Ballester et al 1999, Karim and Dean 2000); thus, a mutant that is able to overcome one of the *Cry1A* genes is also likely resistant to other *Cry1A* genes. Therefore, combinations of *Cry1A* genes with other groups of *Bt* genes should be explored to prevent or delay the emergence of pest resistance. Based on the assay of δ-endotoxin binding to brush border membrane vesicles of rice stem borers, it was proposed that *Cry1Ab* or *Cry1Ac* could be combined with *Cry1C*, *Cry2A*, or *Cry9C* for more durable resistance in transgenic plants (Alcantara et al 2004).

In collaboration with IRRI, we previously developed a transgenic line harboring a *Cry1Ab/c* fusion gene in Minghui 63, an elite restorer line for hybrid rice (Tu et al 2000). This line and its hybrid with Zhenshan 97A, the male sterile line of the elite hybrid Shanyou 63, have now completed the production demonstration stage according to the regulatory procedures of the Chinese government. The results obtained from large-scale field tests promised to significantly reduce insecticides, labor, and related costs, while at the same time increasing yield and benefiting the health of farmers and consumers (Huang et al 2005).

To improve resistance management, we recently developed transgenic rice individually harboring the *Cry1C* and *Cry2A* genes that were codon-optimized for optimal

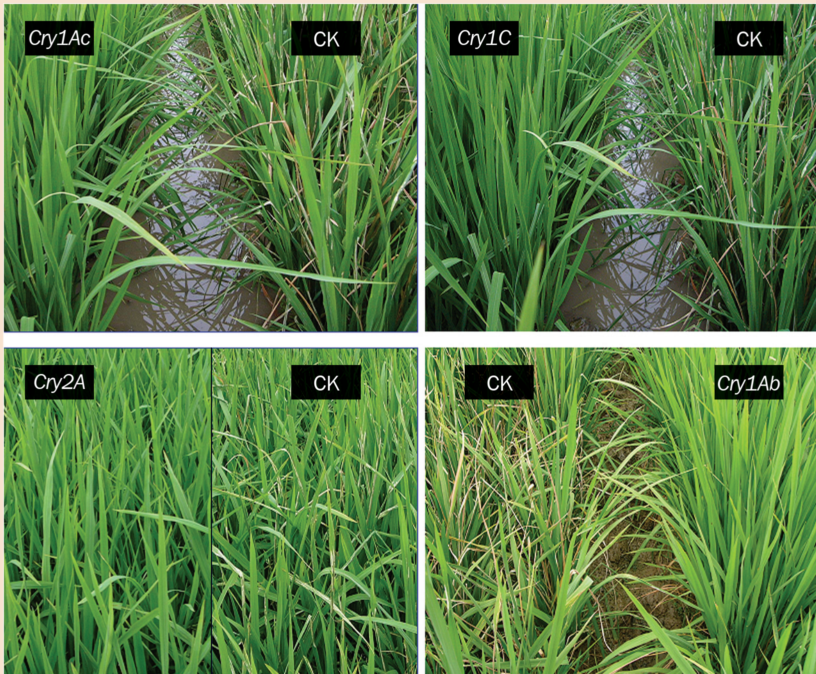


Fig. 1. Transgenic rice plants individually harboring four different *Bt* genes (*Cry1Ab*, *Cry1Ac*, *Cry1C,* and *Cry2A*) in the Minghui 63 background showing strong resistance to natural infestation of leaffolders under field conditions. CK (untransformed control).

expression in plants (Chen et al 2005, Tang et al 2006). Both genes were introduced into Minghui 63, and the transgenic plants with a single-copy transgene were also crossed with Zhenshan 97A. Field tests showed that both the transgenic lines and their hybrids with Zhenshan 97A exhibited strong resistance to natural infestation of leaffolders and stem borers (Fig. 1). Together with the transgenic line harboring the *Cry1Ab/c* fusion gene in Minghui 63, availability of the transgenic lines with the new *Bt* genes has provided the critical materials and flexibility for developing rice lines with various combinations of multiple resistance by gene stacking.

Resistance to brown planthopper (BPH). A total of more than 18 BPH resistance genes have currently been characterized according to their reactions to different BPH biotypes (Athwal et al 1971, Lakshminarayana and Khush 1977, Sidhu and Khush 1979, Khush and Brar 1991, Ishii et al 1994, Huang et al 2001, Jena et al 2006). Making use of a molecular marker linkage map, we identified two genes (or major QTLs) for BPH resistance from B5, a highly resistant line that derived its resistance genes from an accession of the wild rice *Oryza officinalis* (Huang et al 2001). These two loci were designated *Qbp1* and *Qbp2* and were located on chromosomes 3 and 4,

respectively. Both genes had large effects on BPH resistance and the two loci acted essentially independent of each other in determining the resistance. Comparison of the chromosomal locations and reactions to BPH biotypes indicated that these two genes are different from at least nine of the ten previously identified BPH resistance genes.

These two genes have been subsequently incorporated into several important parental lines in hybrid rice breeding programs in China. Lines containing either or both of the genes developed by marker-assisted selection showed enhanced resistance to BPH with artificial infestation (He YQ, unpublished data). Together with the BPH resistance genes reported by other groups, the resistant lines may provide effective protection of the rice crop against BPH under field conditions.

Identifying genes for developing disease-resistant rice

A number of genes for disease resistance have been cloned in rice, including *Xa1* (Yoshimura et al 1998), *xa13* (Chu et al 2006), *xa5* (Iyer and McCouch 2004), *Xa21* (Song et al 1995), *Xa26* (Sun et al 2004), and *Xa27* (Gu et al 2005) for bacterial blight resistance, and *Pib* (Wang et al 1999) and *Pi-ta* (Bryan et al 2000) for blast resistance.

Two of the genes in the above list, *xa13* and *Xa26*, were isolated by our group using a map-based cloning strategy. R gene *xa13* is completely recessive and confers resistance to *Xoo* strain PXO99 only in homozygous status. Molecular characterization of *xa13* showed that the resistance conferred by this gene resulted from mutations in the promoter region of a gene essential for pollen development (Chu et al 2006). Resistant (recessive) and susceptible (dominant) alleles of *xa13* can encode identical proteins located in the plasma membrane, but have crucial sequence differences in their promoter regions. Suppressing expression of either the dominant or recessive allele of *xa13* enhanced the resistance but caused male sterility of the plant. Thus, promoter mutations of an essential gene for pollen development and bacterial growth that cause down-regulation of expression during host-pathogen interaction created a fully recessive allele conferring race-specific resistance to bacterial blight, which represents a new type of plant disease resistance.

Xa26 was isolated from rice cultivar Minghui 63 (Sun et al 2004). It confers resistance against *Xoo* at both seedling and adult stages. Molecular analysis showed that it belongs to a multigene family consisting of four members. It encodes a leucine-rich repeat (LRR) receptor kinase-like protein and is constitutively expressed. Transgenic plants carrying *Xa26* showed enhanced resistance in the background of japonica cultivar Mudanjiang 8 compared with Minghui 63 at both seedling and adult stages, suggesting that the resistance conferred by *Xa26* is also influenced by genetic background. Further analysis showed that *Xa3* identified many years ago is the same gene as *Xa26* (Xiang Y et al, unpublished data).

With the progress in map-based cloning, several genes with strong field resistance have been fine-mapped, including *Xa4* and *Xa7* for bacterial blight resistance

and *Pi2* for blast resistance. Identification of closely linked markers has also enabled pyramiding of these genes using marker-assisted selection.

Chen et al (2001) assayed pathotypes of 792 single-spore isolates of *Pyricularia grisea* using samples collected from 13 major rice-growing provinces of central and southern China. These strains were tested by inoculation with 13 host differentials consisting of six indica and seven japonica near-isogenic lines (NILs), collectively carrying most of the genes for blast resistance identified previously. The results showed that there was a large difference in the frequencies of the isolates producing compatible reactions on the NILs. For example, very small proportions (10% or less) of the isolates could cause compatible reactions on NILs carrying *Pi1* or *Pi2*, but a large proportion (41.5%) of the isolates could overcome the resistance of the NIL carrying *Pi3*. Moreover, a combination of *Pi1* and *Pi2* would be susceptible to only 2.0% of the isolates and adding *Pi4* would make this proportion even smaller. The data provided very useful information for formulating strategies for improving blast resistance in rice breeding programs, which have been put into practice by our group.

Molecular marker-assisted selection has been successfully applied for pyramiding genes for bacterial blight resistance, which appeared to be very effective (Huang et al 1997a). *Xa21* was introgressed into hybrid rice parent Minghui 63 by marker-assisted selection (Chen et al 2000). Field examination of a number of agronomic traits showed that Minghui 63(*Xa21*) and Shanyou 63(*Xa21*) were identical to Minghui 63 and Shanyou 63 when there was no disease stress. Under heavily diseased conditions, Minghui 63(*Xa21*) showed significantly higher grain weight and spikelet fertility than Minghui 63, and Shanyou 63(*Xa21*) was significantly higher than Shanyou 63 in grains per panicle, grain weight, and yield. The improved hybrids have now been used in rice production in China.

To further enhance the resistance of hybrid rice, a strategy was developed such that bacterial blight resistance genes such as *Xa4*, *Xa7*, *Xa21*, and *Xa26* were pyramided in the restorer lines, and blast resistance genes such as *Pi1*, *Pi2*, and *Pi3* were transferred to the maintainer lines, based on which male sterile lines were developed by backcrossing. Field tests under disease-prone conditions showed that these lines and hybrids were highly resistant to both diseases (He YQ et al, unpublished data). Hybrids with various combinations of resistance genes can be produced by crossing these lines in order to meet the need of rice production.

Identifying genes for nutrient efficiency (N and P)

In recent years, we have adopted a combination of genomic approaches to discover genes that are involved in N and P metabolism, with the ultimate goal to improve N and P efficiency of the rice crop.

Identifying genes for low-N tolerance. Two approaches were taken by our group to identify genes that may be involved in response to low-N stress. As the first approach, Lian et al (2006) analyzed the expression profiles of indica rice cultivar Minghui 63 at the seedling stage at 20 min, 1 h, and 2 h after low-N stress, with normal N as the control, using a microarray of 11,494 rice ESTs representing 10,422 unique genes. Although no significant difference was detected in the leaf tissue, a total of 473 ESTs

were detected as responsive to low-N stress in the root tissue, with 115 ESTs showing up-regulation and 358 ESTs showing down-regulation. The analysis of expression profiles after low-N stress identified the following patterns: (1) the genes involved in photosynthesis and energy metabolism were down-regulated rapidly; (2) many of the genes involved in early responses to biotic and abiotic stresses were up-regulated, whereas many other stress-responsive genes were down-regulated; (3) regulatory genes, including transcription factors and ones involved in signal transduction, were both up- and down-regulated; and (4) the genes known to be involved in N uptake and assimilation showed little response to the low-N stress. Many of the genes have been selected for functional analysis by transformation in order to assess their potential usefulness in improving N-efficiency of the rice plants.

In the second approach, we analyzed the genetic components associated with low-N tolerance in rice at the seedling stage using a population of 239 recombinant inbred lines (RILs) from a cross between Zhenshan 97 and Minghui 63, the parents of elite hybrid Shanyou 63 (Lian et al 2005). A genetic linkage map with 253 DNA marker loci was constructed. Seedlings of RILs were cultivated in low-N and normal-N solutions. Root, shoot, and plant weight in the two N treatments was measured and the relative weight of the two treatments for each trait was considered as measurements for low-N tolerance. Four to eight QTLs with main effects were detected for each of the nine traits. Very few QTLs were detected in both low- and normal-N conditions, and most QTLs for the relative measurements were different from those for traits under the two N treatments, indicating very little commonality in the genetic basis of the traits and their relative performance under low- and normal-N conditions. Although many studies are necessary for fully understanding the biological mechanisms of N uptake and use efficiency under relative low-N conditions, these QTLs may provide the starting point for identifying the genes and could also be exploited for improving low-N tolerance in rice breeding programs.

Combining genetic analysis with gene expression profiles may provide an effective strategy for identifying genes and pathways involved in specific physiological processes. However, based on the above results, it is obvious that more efforts are needed to identify genes to be practically useful for improving N-use efficiency in rice breeding programs.

Identifying genes for P efficiency. From a cDNA library constructed by the suppression subtractive hybridization (SSH) method, a transcription factor was identified with the bHLH domain corresponding to a QTL for P-use efficiency and found to be P-deficiency-responsive (Yi et al 2005). The gene was cloned from a P-efficient indica landrace (Kasalath) and designated as *OsPTF1* (*O. sativa* L. phosphate transcription factor). When the transcription factor was introduced into a low-P-sensitive rice variety (Nipponbare) by transformation, the transgenic rice overexpressing *OsPTF1* showed enhanced P-use efficiency in both solution and soil cultures. Tillering ability, root and shoot biomass, and phosphorus content of transgenic rice plants were about 30% higher than those of the wild-type plants in P-deficient culture solution. In soil pot and field experiments, more than a 20% increase in tiller number, panicle weight,

and P content was observed in transgenic plants compared to wild-type plants at low-P levels (Yi et al 2005).

There was another major QTL, designated *Pup1*, for P uptake from P-fixing soils identified on chromosome 12 (Wissuwa et al 2005). A NIL carrying an introgressed allele from the donor parent Kasalath showed 4-fold increase in P uptake in a P-fixing soil compared with the recipient parent Nipponbare, with a large effect on plant growth and grain yield. The *Pup1* locus has now been fine-mapped to a DNA fragment of 150 kb in length, and a number of genes that were differentially expressed between the NIL and Nipponbare under P-deficient conditions were identified as likely candidates for this gene (Wissuwa et al 2005).

Both of the genes appear to be promising for improving P uptake and use efficiency of the rice crop, although further study is needed to evaluate their effectiveness in the genetic backgrounds of elite cultivars under diverse field conditions.

Identifying genes for drought resistance

Our group also invested a large effort in identifying genes for drought resistance following a combination of approaches.

Drought tolerance (DT) and drought avoidance (DA) are two major mechanisms in drought resistance of higher plants. We recently conducted a study on the genetic basis of drought resistance at the reproductive stage in rice, which separated the genetic components of drought tolerance from those of drought avoidance (Yue et al 2006). The study made use of a RIL population from a cross between lowland indica cultivar Zhenshan 97 and tropical japonica upland cultivar IRAT109. Sowing time was staggered among the lines to synchronize flowering based on the heading dates of the RILs obtained previously. The plants were grown individually in PVC pipes and two cycles of drought stress were applied to individual plants starting at the booting stage, with unstressed plants as the control. This experimental design would ensure that all the plants received a similar level of stress treatment such that genotypes with a deep root system or small plant size did not have an advantage in avoiding drought damage. A total of 21 traits measuring fitness, yield, and the root system were investigated. Little correlation was detected of relative yield traits with potential yield, plant size, and root traits, suggesting that DT and DA were well separated in the experiment. Thus, the relative performance of fitness and yield-related traits under drought stress and control conditions unambiguously provided measurements for DT. The root traits provided the measurements for DA, although the contribution of this component to drought resistance at the whole-plant level was eliminated by the experimental design (Table 1).

A genetic linkage map consisting of 245 SSR markers was constructed for mapping QTLs for these traits. A total of 30 QTLs were resolved for seven traits of relative performance of fitness and yield, 36 QTLs for five root traits under control conditions, and 38 for seven root traits under drought-stress conditions, suggesting the complexity of the genetic bases of both DT and DA. Only a small portion of QTLs for fitness and yield-related traits overlapped with QTLs for root traits, in which most of the positive alleles for fitness and yield-related traits and root traits were from different parents,

Table 1. Some examples of candidate genes for drought resistance.

Gene	Category	Function	Reference
CBF3	AP2 type transcription factor	Transcription factor	Jaglo-Ottosen et al (1998), Kasuga et al (1999)
LOS5	Enzyme involved in ABA biosynthesis	Water status regulation	Xiong L and Zhu J, unpublished
SOS2	Protein kinase	Ion homeostasis	Guo Y and Zhu J, unpublished
TPS	Trehalose-6-phosphate synthase	Novel osmoprotectant	Hölmstrom et al (1996), Yeo et al (2000)
HVA1	Late embryogenesis abundant protein gene	Tolerance to water deficit and salt stress	Xu et al (1996)
NPK1	MAPKKK	Detoxification of ROS	Kovtun et al (2000)
ZAT10	Zinc finger transcription factor	Transcription factor	Lee H and Zhu J, unpublished
NCED3	9-*cis*-epoxycarotenoid dioxygenase	Water status regulation	Qin and Zeevaart (2002), Iuchi et al (2001), Thompson et al (2000)
NHX1/2	Vacuolar Na+/H+ antiporter	Ion homeostasis	Apse et al (1999)
CodA	Choline oxidase	Novel osmoprotectant	Sakamoto et al (1998, 2000), Prasad et al (2000)

indicating that DT and DA had distinct genetic bases. The development of NILs for some of the QTLs with relatively large effects in the genetic backgrounds of breeding lines is now in progress.

In a different study, we identified a transcription factor, *SNAC1*, based on data from a DNA chip analysis, which may hold promising utility in improving drought and salinity tolerance in rice (Hu HH et al, unpublished data). Based on the expression profile of rice under drought stress using a cDNA microarray containing 9,216 unique cDNA sequences (Zhang Z et al, unpublished data), an EST showing a 5.6-fold increase in expression level in an upland rice cultivar (IRAT109) after drought stress was identified. This EST showed homology to known *NAC* genes. Expression of this gene could be induced by drought, salt, and cold stresses as well as ABA treatment. This gene is induced specifically in guard cells in leaves by dehydration stress. Transgenic rice overexpressing the *SNAC1* gene showed significantly enhanced drought resistance (22–34% higher seed setting than the control) in the field under severe drought-stress conditions at the reproductive stage, while showing no phenotypic changes or yield penalty. The transgenic rice lost water slowly by closing more stomatal pores and maintained turgor pressure at a significantly lower level of relative water content than the wild type. The transgenic rice also showed significantly improved drought and salt tolerance (80% higher survival rate than the control) at the vegetative stage.

We are also testing several genes for their effects on drought resistance in transgenic rice by introducing each of them to rice cultivar Zhonghua 11, driven by both constitutive and inducible promoters. These are genes reported in the last decade by various groups as having significant effects on drought resistance, which were recommended by a group of experts working in the area of plant abiotic stress, assembled by the Rockefeller Foundation. T_1 families from over 1,000 transgenic plants were tested both in a field equipped with a rain shelter and in PVC pipes with different levels of stress applied at the reproductive stage. Data gathered so far indicate that transgenic plants overexpressing some of the genes indeed produced a significantly increased level of spikelet fertility under severe drought stress and elevated relative yield compared with the wild-type plants.

We anticipate that some of the genes may have important roles to play in improving rice for drought resistance.

Identifying genes for yield and quality

For a long time, yield has been generally regarded as being controlled by multiple genes, each with a small effect. Dissection of QTLs making use of high-density molecular marker linkage maps and development and analyses of NILs in recent years revealed that many of the genes could individually have large effects on the trait, which makes it possible to clone the QTLs for yield traits (Frary et al 2000). In rice, grain yield is determined by three components: number of panicles per plant, number of grains per panicle, and grain weight. QTLs for number of tillers per plant and number of grains per panicle were recently cloned (Li et al 2003, Ashikari et al 2005).

Grain size is a major determinant of grain weight, one of the three components of grain yield. Grain size is also a highly important quality trait in rice, as long and slender rice grain is generally preferred for indica rice by the majority of consumers in China, the United States, and most Asian countries (Unnevehr et al 1992, Juliano and Villareal 1993). Many QTLs associated with rice grain size were identified in the last decade. Among them, *GS3*, a QTL with a major effect on grain size, was consistently detected around the centromeric region of rice chromosome 3 in numerous studies across different genetic backgrounds and environments (Huang et al 1997b, Yu et al 1997, Redoña and Mackill 1998, Tan et al 2000, Xing et al 2002, Aluko et al 2004, Li et al 2004). We recently cloned *GS3* using the map-based cloning method (Fan et al 2006). The gene has a major effect on grain length and weight and also a minor effect on grain width and thickness. Molecular analysis showed that *GS3* consists of five exons and encodes 232 amino acids with a putative PEBP-like domain, a transmembrane region, a putative TNFR/NGFR family cysteine-rich domain, and a VWFC module. Comparative sequencing analysis identified a non-sense mutation, shared among all the large-grain varieties tested in comparison with the small-grain varieties, in the second exon of the putative *GS3* gene. This mutation causes a 178-aa truncation in the C-terminus of the predicted protein, suggesting that *GS3* may function as a negative regulator for grain size. Cloning of such a gene provided an opportunity for fully characterizing the regulatory mechanism and related processes during grain

development. The molecular analysis also suggested useful strategies for improving grain size and shape in rice breeding programs.

Prospects

Development of green super rice as defined in this paper is critical for sustainable rice production. As rice is the most important crop in many countries, green super rice may also play a key role in sustainable agriculture at least for those countries.

Progress is summarized on the identification of genes and development of rice germplasm for improving the traits that make green super rice, such as insect and disease resistance, N and P nutrition efficiency, drought resistance, and grain yield and quality. Globally, the genes for most of the traits have now been isolated and have become available for breeding applications, although further evaluations of genes for some traits are still necessary, especially those for N-use efficiency and drought resistance. Moreover, molecular marker-based germplasm evaluation and genetic analyses have identified a large variety of genes and tightly linked markers for the target traits. With the completion of the rice genome sequencing and rapid advances in international efforts on rice functional genomic studies, hundreds of genes for various traits could be isolated in the next few years.

The availability of these genes would greatly facilitate rice cultivar improvement. It is also anticipated that combinations of genomic tools, which are already available nowadays and will also be generated and optimized for breeding use in future development, will be efficiently applied in rice breeding programs. This development will lead to crop breeding by design, an ideal situation that may be particularly feasible in rice. The goals of green super rice may be achieved as a two-stage process. In the first stage, which has been partly realized, different germplasm carrying single genes of interest is developed and thoroughly evaluated, which by itself is useful for varietal release. In the second stage, the genes introduced into this germplasm will be combined in various ways to develop cultivars with the desired traits for green super rice. Use of these cultivars will result in increased rice productivity, with many fewer inputs, to ensure sustainability of rice production and agriculture in general.

References

Alam M, Datta K, Abrigo E, Oliva N, Tu J, Virmani SS, Datta SK. 1999. Transgenic insect-resistant maintainer line (IR68899B) for improvement of hybrid rice. Plant Cell Rep. 18:7-8.

Alcantara EP, Aguda RM, Curtiss A, Dean DH, Cohen MB. 2004. *Bacillus thuringiensis* δ-endotoxin binding to brush border membrane vesicles of rice stem borers. Arch. Insect Biochem. Physiol. 55:169-177.

Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard HJ. 2004. QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. Theor. Appl. Genet. 109:630-639.

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

Ballester V, Granero F, Tabashnik BE, Malvar T, Ferre J. 1999. Integrative model for binding of *Bacillus thuringiensis* toxins in susceptible and resistant larvae of the diamondback moth (*Plutella xylostella*). Appl. Environ. Microbiol. 65:1413-1419.

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 12:2033-2046.

Chen HL, Chen BT, Zhang DP, Xie YF, Zhang Q. 2001. Pathotypes of *Pyricularia grisea* in rice fields of central and southern China. Plant Dis. 85:843-850.

Chen H, Tang W, Xu C, Li X, Lin Y, Zhang Q. 2005. Transgenic indica rice plants harboring a synthetic *cry2A** gene of *Bacillus thuringiensis* exhibit enhanced resistance against lepidopteran rice pests. Theor. Appl. Genet. 111:1330-1337.

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.

Cheng X, Sardana R, Kaplan H, Altosaar I. 1998. *Agrobacterium*-transformed rice plants expressing synthetic *cry1Ab* and *cry1Ac* genes are highly toxic to striped stem borer and yellow stem borer. Proc. Natl. Acad. Sci. USA 95:2767-2772.

Chu Z, Yuan M, Yao J, Ge X, Yuan B, Xu C, Li X, Fu B, Li Z, Bennetzen JL, Zhang Q, Wang S. 2006. Promoter mutations of an essential gene for pollen development result in disease resistance in rice. Genes Dev. 20:1250-1255.

Datta K, Vasquez A, Tu J, Torrizo L, Alam MF, Oliva N, Abrigo E, Khush GS, Datta SK. 1998. Constitutive and tissue specific differential expression of the gene in transgenic rice plants conferring resistance to rice insect pest. Theor. Appl. Genet. 97:20-30.

Escriche B, Ferre J, Silva FJ. 1997. Occurrence of a common binding site in *Mamestra brassicae*, *Phthorimaea operculella* and *Spodoptera exigua* for insecticidal crystal proteins Cry1A from *Bacillus thuringiensis*. Insect Biochem. Mol. Biol. 27:651-656.

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Zhang Q. 2006. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor. Appl. Genet. 112(6):1164-1171.

Ferre J, Van Rie J. 2002. Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*. Annu. Rev. Entomol. 47:501-533.

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85-88.

Frutos R, Rang C, Royer M. 1999. Managing insect resistance to plants producing *Bacillus thuringiensis* toxins. Crit. Rev. Biotech. 19:227-276.

Fujimoto H, Itoh K, Yamamoto M, Kyozuka J, Shimamoto K. 1993. Insect resistant rice generated by introduction of a modified δ-endotoxin gene of *Bacillus thuringiensis*. Bio/Technology 11:1151-1155.

Ghareyazie B, Alinia F, Menguito CA, Rubia LG, Palma JM, Liwanag EA, Cohen MB, Khush GS, Bennett J. 1997. Enhanced resistance to two stem borers in an aromatic rice containing a synthetic *cry1A(b)* gene. Mol. Breed. 3:401-414.

Gould F. 1998. Sustainability of transgenic insecticidal cultivars: integrating pest genetics and ecology. Annu. Rev. Entomol. 43:701-726.

Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C, Yang F, Chu Z, Wang GL, White FF, Yin Z. 2005. R gene expression induced by a type-III effector triggers disease resistance in rice. Nature 435:1122-1125.

Huang J, Hu R, Rozelle S, Pray C. 2005. Insect-resistant GM rice in farmers' fields: assessing productivity and health effects in China. Science 308:688-690.

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

Huang N, Parco A, Mew T, Magpantay G, McCouch S, Guiderdoni E, Xu JC, Subudhi P, Angeles ER, Khush GS. 1997b. RFLP mapping of isozymes, RAPD, and QTLs for grain shape, brown planthopper resistance in a doubled-haploid rice population. Mol. Breed. 3:105-113.

Huang Z, He G, Shu L, Li X, Zhang Q. 2001. Identification and mapping of two brown planthopper resistance genes in rice. Theor. Appl. Genet. 102:929-934.

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.

Iyer AS, McCouch SR. 2004. The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. Mol. Plant-Microbe Interact. 17:1348-1354.

Jena K, Jeung J, Lee J, Choi H, Brar D. 2006. High-resolution mapping of a new brown planthopper (BPH) resistance gene, *Bph18*(t), and marker-assisted selection for BPH resistance in rice (*Oryza sativa* L.). Theor. Appl. Genet. 112:288-297.

Karim S, Dean DH. 2000. Toxicity and receptor binding properties of *Bacillus thuringiensis* delta-endotoxins to the midgut brush border membrane vesicles of the rice leaf folders, *Cnaphalocrocis medinalis* and *Marasmia patnalis*. Curr. Microbiol. 41:276-283.

Koziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, Crossland L, Dawson J, Desai N, Hill M, Kadwell S, et al. 1993. Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. Bio/Technology 11:194-200.

Khush GS, Brar DS. 1991. Genetics of resistance to insects in crop plants. Adv. Agron. 45:223-274.

Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N. 2004. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). Plant J. 37:315-325.

Lakshminarayana A, Khush GS. 1977. New genes for resistance to the brown planthopper in rice. Crop Sci. 17:96-100.

Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q, Yuan L, McCouch SR. 2004. QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. Genome 47:697-704.

Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, et al. 2003. Control of tillering in rice. Nature 422:618-621.

Lian X, Xing Y, Yan H, Xu C, Li X, Zhang Q. 2005. QTLs for low nitrogen tolerance at seedling stage identified using a recombinant inbred line population derived from an elite rice hybrid. Theor. Appl. Genet. 112:85-96.

Lian X, Wang S, Zhang J, Feng Q, Zhang L, Fan D, Li X, Yuan D, Han B, Zhang Q. 2006. Expression profiles of 10,422 genes at early stage of low nitrogen stress in rice assayed using a cDNA microarray. Plant Mol. Biol. 60:617-631.

Lin JY, Shen M. 1996. Rice production constraints in China. In: Evenson RE, Herdt RW, Hossain M, editors. Rice research in Asia: progress and priorities. CAB International, Wallingford, UK, and International Rice Research Institute, Los Baños, Philippines. p 161-178.

Nayak P, Basu D, Das S, Basu A, Ghosh D, Ramakrishnan NA, Ghosh M, Sen SK. 1996. Transgenic elite *indica* plants expressing *cryA(c)* δ-endotoxin of *Bacillus thuringiensis* are resistant against yellow stem borer (*Scirpophaga incertulas*). Proc. Natl. Acad. Sci. USA 94:2111-2116.

Peng S, Huang J, Zhong X, Yang J, Wang G, Zou Y, Zhang F, Zhu Q, Buresh R, Witt C. 2002. Research strategy in improving fertilizer nitrogen use efficiency of irrigated rice in China. Sci. Agric. Sin. 35:1095-1103.

Perlak FJ, Deaton RW, Armstrong TA, Fuchs RL, Sims SR, Greenplate JT, Fischhoff DA. 1990. Insect resistant cotton plants. Bio/Technology 8:939-943.

Redoña E, Mackill D. 1998. Quantitative trait locus analysis for rice panicle and grain characteristics. Theor. Appl. Genet. 96:957-963.

Shelton AM, Zhao JZ, Roush RT. 2002. Economic, ecological, food, safety and social consequence of the deployment of *Bt* transgenic plants. Annu. Rev. Entomol. 47:845-881.

Sheng CF, Wang HT, Sheng SY, Gao LD, Xuan WJ. 2003. Pest status and loss assessment of crop damage caused by the rice borers, *Chilo suppressalis* and *Tryporyza incertulas*, in China. Entomol. Knowledge 40:289-294. (In Chinese with English abstract.)

Sidhu GS, Khush GS. 1979. Linkage relationships of some genes for disease and insect resistance and semidwarf stature in rice. Euphytica 28:233-237.

Stewart CN, Adang MJ, All JN, Boerma HR, Cardineau G, Tucker D, Parrot WA. 1996. Genetic transformation, recovery and characterization of fertile soybean transgenic for a synthetic *Bacillus thuringiensis cryA(c)* gene. Plant Physiol. 112:121-129.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804-1806.

Sun X, Cao Y, Yang Z, Xu C, Li X, Wang S, Zhang Q. 2004. *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes a LRR receptor kinase-like protein. Plant J. 37:517-527.

Tabashnik BE, Finson N, Groeters FR, Moar WJ, Johnson MW, Luo K, Adang MJ. 1994. Reversal of resistance to *Bacillus thuringiensis* in *Plutella xylostella*. Proc. Natl. Acad. Sci. USA 91:4120-4124.

Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang Q. 2000. Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. Theor. Appl. Genet. 101:823-829.

Tang D, Luo Y, Yu Q. 2000. Fundamentals of agricultural water saving. Guangai Paishui 19:1-9. (In Chinese.)

Tang W, Chen H, Xu C, Li X, Lin Y, Zhang Q. 2006. Development of insect resistant transgenic indica rice with a synthetic *Cry1C** gene. Mol. Breed. 18:1-10.

Tu JM, Zhang GA, Data K, Xu CG, He YQ, Zhang QF, Khush GS, Datta SK. 2000. Field performance of transgenic elite commercial hybrid rice expressing *Bacillus thuringiensis* δ–endotoxin. Nature Biotechnol. 18:1101-1104.

Unnevehr LJ, Duff B, Juliano BO. 1992. Consumer demand for rice grain quality. International Rice Research Institute, Manila, Philippines, and International Development Research Center, Ottawa, Canada. 248 p.

Wang Z, Shu Q, Ye G, Cui H, Wu D, Altosaar I, Xia Y. 2002. Genetic analysis of resistance of *Bt* rice to striped stem borer (*Chilo suppressalis*). Euphytica 123:379-386.

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.

Wissuwa M, Gatdula K, Heuer S, Lu X, Ismail AM, Yano M. 2005. Characterization of candidate genes at the *Pup1* locus: a major QTL associated with tolerance to P-deficiency. 5th International Rice Genetics Symposium and 3rd International Rice Functional Genomics Symposium, 19-23 November 2005, Manila, Philippines. Abstracts, p 30.

Wu C, Fan Y, Zhang C, Oliva N, Datta SK. 1997. Transgenic fertile *japonica* rice plants expressing a modified *cry1A(b)* gene resistant to yellow stem borer. Plant Cell Rep. 17:129-132.

Wünn J, Kloti A, Burkhardt PK, Biswas GCG, Launis K, Iglesias VA, Potrykus I. 1996. Transgenic indica rice breeding line IR58 expressing a synthetic *cryA(b)* gene from *Bacillus thuringiensis* provides effective insect pest control. Bio/Technology 14:171-176.

Xing YZ, Tan YF, Hua JP, Sun XL, Xu CG, Zhang Q. 2002. Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor. Appl. Genet. 105:248-257.

Yi K, Wu Z, Zhou J, Du L, Guo L, Wu Y, Wu P. 2005. *OsPTF1*, a novel transcription factor involved in tolerance to phosphate starvation in rice. Plant Physiol. 138:2087-2096.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang Z, 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.

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.

Yue B, Xue W, Xiong L, Yu X, Luo L, Cui K, Jin D, Xing Y, Zhang Q. 2006. Genetic basis of drought resistance at reproductive stage in rice: separation of drought tolerance from drought avoidance. Genetics 172:1213-1228.

Zhu Z, Sun B, Yang L, Zhang L. 2005. Policy and counter-measures to control non-point pollution of agriculture in China. Sci. Technol. Rev. (Agric.) 23:47-51.

Notes

Author's address: National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan) and National Center of Crop Molecular Breeding, Huazhong Agricultural University, Wuhan 430070, China.

Acknowledgments: The work reported here has been supported by grants from the National Program on the Development of Basic Research, the National Special Key Project of Functional Genomics and Biochips, the National Natural Science Foundation of China, and the Rockefeller Foundation.

From gene to adaptation in rice

K. Onishi and Y. Sano

The recent accumulation of information on plant genomes has enabled us to study adaptive traits at both the phenotypic and molecular levels. Genetic diversification is a consequence of the existence of a diverse set of environments. Plant breeding will accelerate the rate of micro-evolution in our changing world. To understand ongoing micro-evolutionary processes, genetic alterations in response to temperature, photoperiod, and biotic environments were investigated in wild and cultivated rice. These adaptive mechanisms were not well explained by a few major genes, suggesting that epistasis, genotype × environment (GE) interaction, and linked genes were involved in addition to genes with a small additive effect. Genetic diversity is affected both by current patterns of micro-evolutionary forces, such as gene flow and selection, and by phylogenetic history. Genealogies of agronomic genes provided insight into their history. Unexpectedly, the "Geen Revolution" gene (*sd1*) preexisted in the wild ancestor, showing that farmers selected it to obtain a high yield in response to altered practices in agriculture. In contrast, in the case of *C*, *A*, and *wx* genes, variants were generated from landraces through natural or artificial selection, suggesting that each of the genes may have its own history.

Keywords: Genetic diversity, genealogy, adaptation, agronomic genes, Asian rice, wild rice

Recent molecular studies on plants have greatly contributed to our understanding of various developmental processes forming the phenotypic diversity of rice; however, our knowledge is still limited as to how such genes contribute to the adaptive evolution of rice. It became increasingly important for breeders to understand the nature of adaptation at both the phenotypic and molecular levels since breeding will artificially accelerate micro-evolutionary processes. Little is known about the number of loci segregating for adaptability and their epistatic effects even by modern quantitative trait loci (QTL) mapping studies. Because of a large number of genotypes in segregating populations, each genotype is unique, hardly recurring without selection. Correlations among morphological traits and between morphology and fitness are often a conse-

quence of linkage disequilibrium, making it difficult to find hidden variations within a population. Genetic diversification is a consequence of the existence of a diverse set of environments to which different individuals have become fit by natural selection. The process of that fitting is the process of adaptation, implying that genetic variation reflects a shape of the underlying world. Furthermore, the phenotype of plants with a particular genotype is regarded as a function of the environment, which is called a norm of reaction. A norm of reaction is the examination of environment-phenotype interactions that is characteristic of a particular genetic constitution (Schlichting and Pigliucci 1998, Lewontin 2000). These problems are a long-standing issue in crop evolution and breeding: however, we need to revisit this based on recent molecular information in order to understand micro-evolutionary processes in a changing world.

Adaptation to temperature and photoperiod

Temperature

Tolerance of low temperatures plays a role for a wide range of geographic distribution in rice. The response is a complex phenomenon that includes primary and secondary injury by exposure to low (or chilling) temperature, depending on the stage of development (Kratsch and Wise 2000, Allen and Ort 2001). Genetic variation in responses to low temperature has been reported to be associated in rice with its geographic distribution as well as its taxonomic groups (Oka 1988, Nagamine and Nakagahra 1990). Among various responses to low temperature, acclimation has not yet been substantially evaluated in rice. Acclimation, which is a response to cold temperature by switching to a more cold-tolerant physiological state, plays a role in freezing tolerance of wheat and barley (Thomashow 1990). A total of 61 cultivated and wild strains, which originate from tropical to the northernmost (43°N) areas, were investigated for responses to low temperature at the three early stages (germination, plumule, and seedling), and after acclimation.

A wide range of variation was observed in all four index values for tolerance (Table 1). Cultivated strains were more tolerant for germinability under low temperature than wild strains, suggesting that this divergence might be related to rice domestication. No difference in tolerance was found at the seedling stage among these taxa, which suggests that variation at the seedling stage is equally required within these taxa. On the other hand, at the plumule stage, the japonica type (including the javanica type) was more tolerant than the indica type and wild strains, suggesting an association with indica-japonica differentiation, as previously mentioned (Oka 1988). A similar tendency was also found in acclimation ability. To examine the adaptive significance of indicas and japonicas, a latitudinal cline was investigated based on the correlation between the index value for tolerance and the latitude of their habitats. Since geographic distribution differs between indica and japonica types, it is difficult to distinguish their contemporary and historical processes. However, wild rice strains might more or less reflect the ongoing processes responding to environment. For plumule tolerance and acclimation ability, a distinct cline was detected among wild strains. The highest correlation of acclimation with latitude suggested that ac-

Table 1. Means of index values for low-temperature tolerance in japonica (J), indica (I), and wild (W) rice strains.

Tolerance	Type[a]	Mean[b]	Range
At germination	J	2.9 a	(2.1–4.0)
	I	2.6 a	(1.7–3.1)
	W	0.8 b	(0.0–2.4)
At plumule stage	J	2.2 a	(0.2–4.0)
	I	1.0 b	(0.0–4.0)
	W	0.9 b	(0.0–2.5)
At seedling stage	J	3.5 a	(1.3–4.0)
	I	2.3 a	(0.0–4.0)
	W	2.7 a	(0.0–4.0)
After acclimation	J	3.8 a	(2.0–4.0)
	I	1.9 b	(0.0–4.0)
	W	1.6 b	(0.0–4.0)

[a]Numbers of strains examined are 27, 13, and 21 in japonica (J), indica (I), and wild (W) rice strains, respectively. [b]Values followed by different letters are significantly different at the 1% level (Scheffe's test).

climation has a significant role for wide distribution (Fig. 1B) although it has not been recognized as an adaptive factor in rice. A clinal distribution was also found in cultivated strains; however, the tendency disappeared within the indica or japonica type (Fig. 1A). Using the recombinant inbred lines (RILs) from a cross between A58 (japonica type from Hokkaido) and W107 (annual type of wild rice from India), QTL analysis was used to elucidate the genetic basis of low-temperature tolerance. The results showed that low-temperature tolerance is controlled by multiple QTLs with a small effect distributed across the genome, as expected from their clinal distribution. Therefore, for low-temperature tolerance, it is hard to speculate that a particular gene with a large effect enabled rice plants to widen distribution toward the north.

Photoperiod

Photosensitivity is a major factor influencing the adaptation of rice plants to different localities by determining flowering time (Vergara and Chang 1985). As expected, several genes have been identified leading to the identification of causal genes (Izawa et al 2003). However, it is not easy to understand how the degree of photosensitivity is regulated by allelic differences and/or interacting genes to adjust flowering time to survive in natural habitats. Responses to different photoperiods were examined by curvilinear regression equations (Itoh and Sano 1999). The two alleles of *Se-1* (= *Hd-1*) from cultivated and wild strains were compared using near-isogenic lines (NILs) under growth regimes of different photoperiods. The equation is expressed by $y = a_0 + a_1x + a_2x^2$, where x and y stand for daylength and days to heading, respectively. The

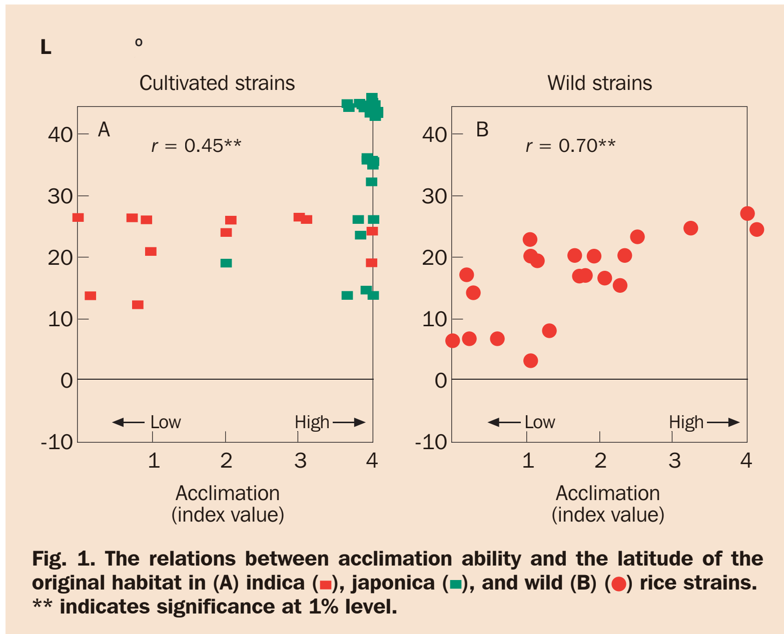


Fig. 1. The relations between acclimation ability and the latitude of the original habitat in (A) indica (■), japonica (■), and wild (B) (●) rice strains. ** indicates significance at 1% level.

two alleles differently affected both basic vegetative growth and optimum daylength, suggesting that the alleles affect the norm of reaction as inferred by the presence of replacement substitutions in the zinc finger domain.

Another example is whether or not multiple linked genes (or QTLs) play a role in hidden variation of flowering time within and between populations, which provides a mechanism for preserving genetic diversity in subdivided populations. It was previously reported that a recessive photosensitivity gene (*se-pat*) on chromosome 6 shows a unique expression of age-dependent manner (Dung et al 1998). Since this gene was recessive, the allelism test was possible using eight late-heading NILs with the segment around *se-pat* from wild and cultivated strains. The results indicated that their late-heading traits are all controlled by the same locus, showing a wide distribution of *se-pat*. Fine mapping, however, revealed that the *se-pat* gene was composed of two components, *se-pat-a* and *se-pat-b*. The allelism test was successful because *se-pat-a* is recessive, possessing a major effect within the compound locus, implying that an allelism test does not necessarily indicate an involvement of only a single gene. Based on fine mapping, *se-pat-a* was different from the previously reported genes (*Hd3a* and *Hd3b*), but *se-pat-b* could be the same as *Hd3a* (Fig. 2).

Among the presumed alleles at the *se-pat* locus, an allele from *Oryza rufipogon* (W593) caused an extreme delay in heading date. The gene was formerly designated *En-Se1*; however, recent genetic dissection showed that it behaves as a recessive photoperiod-sensitivity gene (*en-Se1*). Fine mapping also showed that *en-Se1* was composed of two components, *en-Se1-a* and *en-Se1-b*, like the case of *se-pat* (Fig.

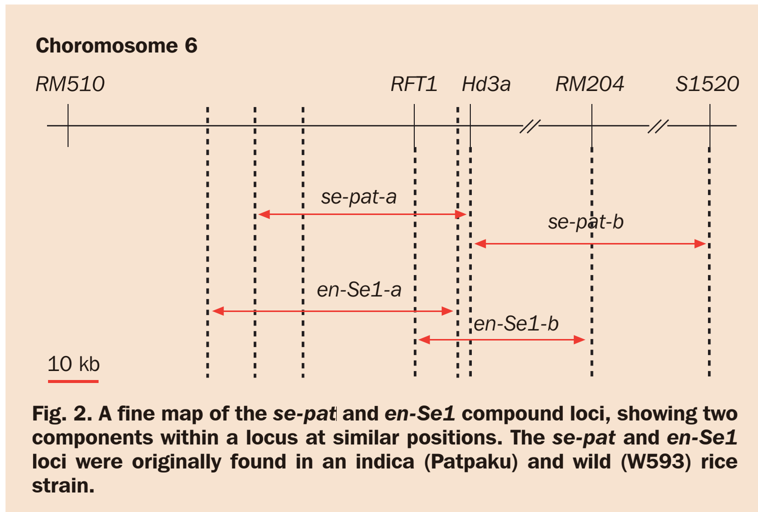


Fig. 2. A fine map of the *se-pat* and *en-Se1* compound loci, showing two components within a locus at similar positions. The *se-pat* and *en-Se1* loci were originally found in an indica (Patpaku) and wild (W593) rice strain.

2). The former component seemed to be allelic to *se-pat-a*, whereas the latter seemed to be allelic to *se-pat-b*, acting as a semidominant allele. Thus, the compound locus behaves like multiple alleles because of the differentiated effects of the components, although transgressive segregation will result from a breakdown of linkage.

Epistasis and G × E interaction

Crossing plants generates offspring that bear novel combinations of alleles, producing more extreme phenotypes than those of either parent. This phenomenon (transgression) is shown to be common (Rieseberg et al 2003), indicating the importance of gene flow for the emergence of novelty in crops. It is possibly related to a rapid change in their genetic architecture responding to the environment. Recent interest in evolution concerns to what extent epistasis is involved in obtaining the novelty as well as additive effects. Early-heading behavior in the northernmost regions of rice cultivation was investigated regarding epistasis and G × E interaction.

Heading time is determined not only by photoperiod sensitivity but also by the basic vegetative growth period, which is defined as the minimum duration of vegetative growth under optimum photoperiod for panicle initiation. Two genes, *Ef1* and *m-Ef1*, were found in two landraces from the northernmost region of rice cultivation (Tsai and Oka 1970, Tsai 1986). *Ef1* shortens the duration to floral initiation from germination without changing photoperiod sensitivity, and *m-Ef1* intensifies the earliness due to *Ef1*. In addition, *m-Ef1* was reported to suppress the effect of a photoperiod-sensitivity gene of *Se1* (Okumoto et al 1996, Ichitani et al 1998), indicating distinct epistasis among the three genes. We examined eight NILs carrying all possible combinations of the three genes under different regimes of two temperatures (28 and 22 °C) and four photoperiods (9, 10, 11, and 12 h) in growth cabinets. Analysis of variance (ANOVA) indicated that the effects due to the three genes and their epistasis are highly significant

Table 2. F values (degrees of freedom) based on ANOVA for detecting interactions between genotype and environment.[a]

Factors due to genotypes	Factors due to environments		
	Photoperiod (P)	Temperature (T)	P × T
Se1	183.47**(3)	11.29**(1)	1.89^{ns}(3)
Ef1	22.68**(3)	237.19**(1)	17.71**(3)
m-Ef1	169.42**(3)	298.00**(1)	7.74**(3)
Se1 × *Ef1*	113.99**(3)	81.10**(1)	6.84**(3)
Se1 × *m-Ef1*	76.87**(3)	155.92**(1)	6.70**(3)
Ef1 × *m-Ef1*	31.28**(3)	104.86**(1)	7.33**(3)
Se1 × *Ef1* × *m-Ef1*	38.06**(3)	140.64**(1)	3.67*(3)

[a]All plants were grown in growth cabinets under different regimes of photoperiod (9, 10, 11, and 12 h) and temperature (28 and 22 °C). * and ** show significance at 5% and 1%, respectively. ^{ns}indicates nonsignificance.

and that the effects due to photoperiod, temperature, and their interaction were also highly significant. Furthermore, almost all of the possible interactions were significant between genotype and environment (Table 2). Interestingly, the epistasis of *Ef1* × *m-Ef1* disappeared under natural paddy fields at Sapporo (43°N) although distinct epistasis was observed in the southern areas (Itoh et al, unpublished). Therefore, it is strongly suggested that both epistasis and G × E interaction need to be studied to understand regional adaptability.

Competitive interaction with weed species

Competition between crops and weeds has been one of the most serious problems influencing crop yield. The use of herbicides has been successful to control weeds in intensified management; however, sustainable agriculture urgently requires reducing their heavy application. The genetic basis of competitive abilities and the adaptive significance of weeds are little understood so far. Weed species invade rice fields and their abundance depends on locality and type of crop management. To examine competitive interactions with weeds, QTL analysis was carried out using RILs between cultivated (A58) and wild rice (W107). Three weed species (*Echinochloa oryzicola*, *Monochoria korsakowii*, and *Scirpus juncoides*) that are major weeds in rice fields of Hokkaido were used as competitors to rice. Competitive abilities were evaluated as the values of competitive response (CR) and competitive effect (CE) for each weed species. CR is the ability to withstand the negative effects of weeds and CE is the ability to depress weed growth. Values were estimated based on the ratio of dry weight (DW) as follows:

Table 3. Putative QTLs for competitive response (CR) and competitive effect (CE) against weed species detected in recombinant inbred lines (RILs) between a cultivated (A58, japonica type) and annual form (W107) of wild rice.

Competition and competitor[a]	Chromosome number	LOD	PVE[b]	Effect[c]
CR to weed species				
Echinochloa oryzicola	1	4.7	24.7	A58
Monochoria korsakowii	7	4.5	23.8	A58
CE against weed species				
M. korsakowii	2	3.5	16.6	W107
M. korsakowii	7	3.9	17.4	A58

[a]CR and CE were computed by *ln* (DW of a RIL in mixed stands/DW of a RIL in pure stands) and –*ln* (DW of weeds in mixed stands/DW of weeds in pure stands), respectively, where DW stands for dry weight of aboveground parts. [b]Percentage of variance explained. [c]The allele contributing to increased competitive ability.

CR = *ln* (DW of a RIL in mixed stands/DW of a RIL in pure stands).

CE = –*ln* (DW of weeds in mixed stands/DW of weeds in pure stands).

The higher values in CR and CE indicate higher competitive abilities. A58 showed higher values of CR to and CE against all three weed species than did W107. The CR and CE values for the three weed species tended to be distributed between the parental values; however, the CR and CE values often showed a transgressive segregation except for the CR for *E. oryzicola* and *S. juncoides*. Based on QTL analysis, one and three putative QTLs were detected for competition with *E. oryzicola* and *M. korsakowii*, respectively (Table 3). Three of the four QTLs increased CR or CE by the A58-derived loci, and a putative QTL increased CE for *M. korsakowii* by the W107-derived locus, suggesting a possible cause of the transgressive segregation observed. The location of the detected QTLs differed depending on the competitor, suggesting that rice plants could interfere with different kinds of weeds in different manners. The results confirm that competitive ability could be used for rice to adjust in different biotic environments. Further studies are under way to examine whether higher competitive ability is selected by accumulating the QTLs detected.

Molecular evolution of agronomic genes

Genealogy of the "Green Revolution" gene

High-yielding varieties (HYVs) were developed in rice by altering the architecture of plants to produce a short stature that results in lodging resistance and high seed production under heavy application of nitrogen fertilizer. A recessive *semi-dwarf 1* (*sd1*

or *OsGA20ox2*) gene detected in Dee-geo-woo-gen (DGWG) was used for an HYV (IR8). Recent identification of the gene proved that it controls the gibberellin (GA) biosynthetic pathway (Ashikari et al 2002). The DGWG allele has a 382-bp deletion in exon 1 causing a loss-of-function mutation. It is known, however, that semidwarfing traits due to the DGWG allele are slightly deleterious under traditional cultivation in the tropics, suggesting how such a deleterious gene has been preserved in landraces (Jennings and Aquino 1968). We investigated the distribution of the DGWG allele using 256 accessions, including cultivated strains and their wild relatives chosen from worldwide collections (Nagano et al 2005). By means of PCR amplification, the DGWG allele was detected in Liu-t'ou-tu (an indica rice from Taiwan) and two *O. rufipogon* accessions from China, showing its limited distribution. The nucleotide sequence of exon 1, exon 2, and the 5′-flanking region (alignment length of 1,824 bp) was compared in 66 accessions. Based on the nucleotide sequences, 20 haplotypes were discriminated in Asian cultivated and wild rice. The neighbor-joining (NJ) tree showed that the Asian rice complex contained *sd1* genes from two different lineages, one of which included all the cultivated strains examined. Among the 20 haplotypes, only three (RS1–RS3) were shared between wild strains and indica rice.

The finding that the DGWG carriers have the identical sequence indicates their common origin, showing that the mutant allele is distributed not only in landraces but also in the wild progenitor. The haplotype network using a maximum parsimony analysis showed that the DGWG allele was recently derived from haplotype S1 (Fig. 3). Regarding the origin of the DGWG allele, it was driven by a new mutation that arose during the domestication process or preexisted as a rare variant in the wild progenitor. The fixation of a beneficial allele eliminates DNA variation in the surrounding region. To investigate polymorphisms around the *sd1* locus, three molecular marker loci, whose positions are separated from the *sd1* locus by 6–7 cM on chromosome 1, were examined in the five RS2 carriers. A wild strain (W1944) was different from the other carriers in the three loci, suggesting that the DGWG allele preexisted in the wild progenitor, rather than that the DGWG allele was recently introgressed from HYVs to wild strains.

High planting density in productive or fertilized fields increases competition, which selects against semidwarf plants. A negative relationship between competitive ability and seed production is often observed as a trade-off in various plant species in natural fields. In rice, this is explained as follows: an increase in plant height leads to mutual shading and a decrease in leaf area index, causing low seed production; conversely, short-statured plants minimize mutual shading, giving rise to a high yielding capacity per unit area (Tanaka et al 1966). This complex relationship means that the selective value of the dwarfing phenotype differs depending on the neighboring plants or habitat.

How could such a slightly deleterious gene be preserved in the wild progenitor? Phenotypic effects of the DGWG alleles were examined in the F_5 populations of a cross between A58 (a landrace) and W1944 (the DGWG carrier). The crossing experiment revealed a continuous and transgressive segregation in culm length, implying that the DGWG allele of W1944 functions in combination with other interacting genes.

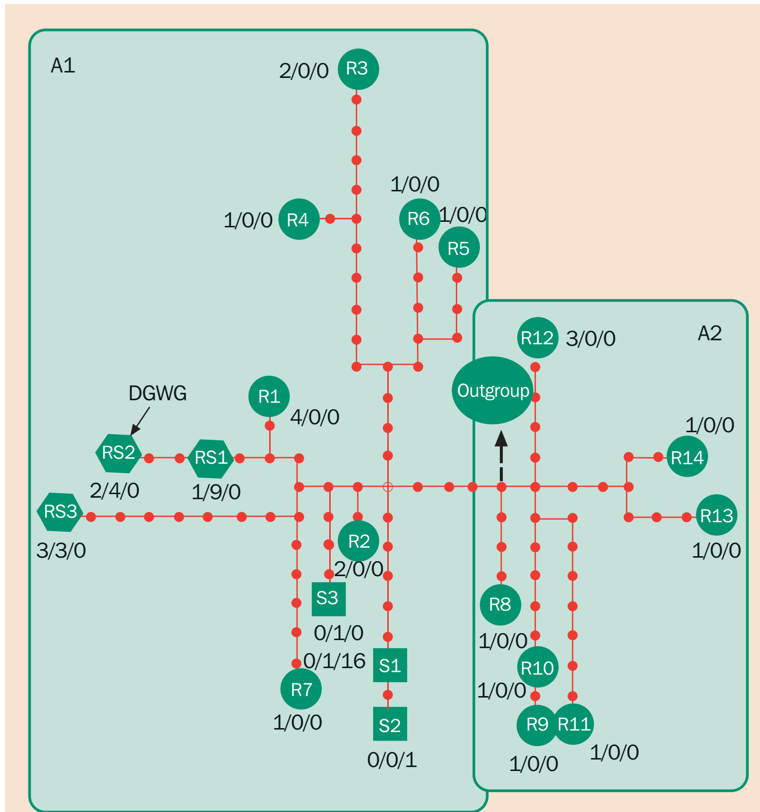


Fig. 3. The *sd1* (*OsGA20ox2*) haplotype tree based on the maximum-parsimony method. The tree represents one of three most-parsimonious arrangements. Mutational steps (substitutions and indels) are shown by small circles. The 20 haplotypes found in wild and cultivated rice are shown in circles with R and squares with S, respectively, while shared haplotypes are shown in hexagons with RS. The semidwarf mutation is shown by DGWG. Numbers near each haplotype indicate its occurrence in the three taxa (wild/indica/japonica), including 62 rice accessions.

Therefore, the DGWG variant might have been preserved as a hidden variation in the genetic background of wild rice, without expressing a short stature.

The results confirm that farmers have selected the DGWG allele together with its interacting genes to obtain high yield under various practices in agriculture and that the identical variant was maintained in different landraces. This view is controversial regarding the assumption that some landraces gained agronomic genes from a new variant in traditional practices.

The *Waxy* gene

The *Wx* gene encodes ADP glucose starch glycosyl transferase and controls amylose content (AC), a major determinant of eating quality. The diverse AC phenotypes are unique in cultivated rice, suggesting that this diversification may be related to domestication. The simple inheritance contrasts with the continuous variation of AC observed among nonwaxy (glutinous) cultivars. This is explained by a series of alleles at the *Wx* locus (Sano 1984). Naturally occurring *Wx* allleles are classified into five types—Wx^a, Wx^b, Wx^{in}, Wx^{op}, and *wx*—in terms of the amount of the gene product as well as AC. A mutation in the intron 1 splice donor site of the *Wx* gene is responsible for the difference between the two alleles of Wx^a and Wx^b, which are predominantly distributed in indica and japonica types, respectively (Hirano et al 1998).

The genealogy of the *Wx* gene was reported in 105 waxy and nonwaxy landraces using sequences of about 2.7 kb, including the intron 1 splice donor site and a variable region of *pSINE-r1* (Olsen and Purugganan 2002). We also examined the same region (987 bp) in 57 cultivated and wild relatives by the NJ tree. Although different taxa in Asian rice tended to be differentiated, their topology was supported only with low bootstrap values. Therefore, the entire coding sequences were examined in 22 cultivated and wild strains, including African rice species as an outgroup. The NJ tree supported the presence of three groups corresponding to indica, japonica, and African rice with high bootstrap values (above 90%), although an annual type of wild rice was included in a group of indica type. In addition, the detected replacement substitutions were supposed to be responsible for the allelic differentiation of Wx^{in} and Wx^{op}. An exception was that *wx* in the indica type was included in a group of japonica type, giving the two possibilities that one of the two lineages is fixed only within the japonica type or that the dimorphism in the indica type is a result of gene flow.

The *C* (= *OsC1*) gene

Divergent phenotypes are often detected in domesticated plants despite invariant phenotypes in their wild forms. An example is the varying degree of apiculus coloration due to anthocyanin pigmentation, which is regulated by a series of alleles at the *C* and *A* loci in rice (Takahashi 1982). Flavonoid derivatives, including anthocyanins, are responsible for a wide range of biological functions such as plant-microbe interactions and plant defense responses. The biosynthesis of anthocyanin pigments requires complex interactions between genes with both structural and regulatory roles. The importance of regulatory genes in phenotypic diversity has been proposed (Purugganan and Wessler 1994).

The *C* gene was recently proved to be the rice homolog (*OsC1*) of maize *C1* belonging to the group of R2R3-Myb factors (Saitoh et al 2004). Two different types of deletions causing a frame shift were detected in exon 3, and both of the deleted nucleotides corresponded to the positions of putative base-contacting residues, suggesting that the indica and japonica types carry loss-of-function mutations with independent origins. Replacement substitutions were frequently detected in *C* genes of strains carrying the different *C* alleles previously defined. The genealogy of the *C* gene suggested that allelic diversification causing phenotypic change might have

resulted from mutations in the coding region rather than from recombination between preexisting alleles, showing diversification within cultivated forms.

The *A* (= *DFR*) gene

The possible candidate of the *A* gene of rice was reported to be the *Dihydroflavonol 4-reductase* (*DFR*) gene, which is the rice homolog of maize *A1* (Nakai et al 1998). Our recent results supported this because the *DFR* and *A* genes co-segregated and replacement substitutions were detected in the different alleles previously reported. The pattern of anthocyanin pigmentation was compared among the three T65*A* NILs, A5, W025, and W1618, which carried the functional *DFR* allele from the related taxa. The three NILs grown in a greenhouse showed different patterns of colaration in leaf sheath at the seedling stage as well as in apiculus and stigma at the flowering stage, although the recurrent parent of T65 showed no coloration in these tissues. When they were grown under low temperature, pigmentation in apiculus was intensified in different manners, suggesting that the allelic difference affects the temperature sensitivity in gene expression.

Purple stigma3 (Ps3) is responsible for pigmentation in stigma together with *C* and *A*. Since T65 carried *Ps3*, all three NILs showed pigmentation in stigma. Further, *Inhibitor for purple stigma* (*I-Ps*) is known to suppress coloration in stigma (Kinoshita 1995). When each NIL was crossed with Shinriki (japonica type) carrying *I-Ps3*, the stigma was not pigmented only in the F_1 between T65*A*(A5) and Shinriki, although the apiculus was pigmented in all the F_1 hybrids. This indicates that *I-Ps3* interacts differently with the *A* allele of W1618 (*Oryza longistaminata*) and W025 (*O. glaberrima*) from that of A5, and suppression of coloration takes place through an interaction with the *A* allele.

Molecular population analysis of the *A* gene suggested that allelic diversification causing phenotypic change might have resulted from replacement substitutions in the coding region as found in the *C* gene. The 47 *A* sequences of cultivated and wild rice showed that indica, japonica, and wild rice each carried distinct haplotypes, with few shared haplotypes among them. A phylogenetic network for haplotypes of the *A* locus was constructed according to Saitou and Yamamoto (1997). The result revealed networked evolution, indicative of recombination and/or gene conversion (Fig. 4). This implies that bifurcating evolutionary relations are not necessarily expected even among the haplotypes in Asian rice, suggesting a different history from that of *C*.

Diversity in genes governing different agronomic traits

The process of artificial selection is crucial in crops for understanding the importance of genetic diversity that is a major source of varietal improvement (Harlan 1975). The domestication of crops from their wild ancestors leads to a loss of genetic diversity both through bottlenecks and intense selection for agronomic traits (Tanksley and McCouch 1997). The level of genetic diversity was compared in six genes (Table 4). The results indicated that silent site nucleotide diversity (π) was lower in cultivated forms than in the wild ancestor, as expected. However, when compared with π of the wild ancestor, the loss of diversity was much more pronounced in the japonica type

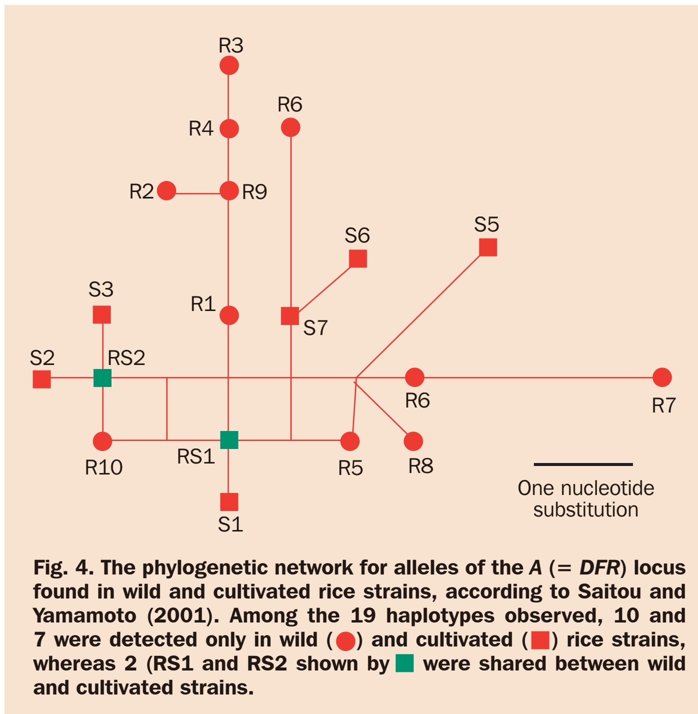


Fig. 4. The phylogenetic network for alleles of the *A* (= *DFR*) locus found in wild and cultivated rice strains, according to Saitou and Yamamoto (2001). Among the 19 haplotypes observed, 10 and 7 were detected only in wild (●) and cultivated (■) rice strains, whereas 2 (RS1 and RS2 shown by ■ were shared between wild and cultivated strains.

(0.0–14.3 %) than in the indica type (31.0–83.1%). The ratio (R/S) of the number of replacement (or nonsynonymous) substitutions per replacement site to the number of synonymous substitutions per synonymous site was calculated. An R/S higher than 1.0 was detected only in the *C* locus of the japonica type (7.75), giving an indication of positive selection. Another test of neutrality, Tajima's D, was also computed. For *RFT1* in the japonica type, Tajima's test gave a significant result (D = –2.3771, $P<0.01$), indicating that less frequent sites are in excess.

These results lead us to consider that each of the genes has a unique history, which might reflect different patterns of natural and/or artificial selection. It seems that some farmers have mined the DGWG allele from the wild population and have maintained it in different landraces, whereas other farmers have selected newly arisen mutations from landraces in the case of *C*, *wx*, and *A* genes, consciously or unconsciously. Therefore, landraces are also expected to preserve agronomically valuable genes, even if a reduction in nucleotide diversity was caused by the population bottleneck due to domestication.

Table 4. A summary of sequence statistics for the six loci examined in Asian rice.[a]

Gene		Cultivated		Wild		% diversity[b]	
		I	J	P	A	I/W	J/W
A (1,669 bp)							
	N[c]	11	21	8	7		
	S[d]	15	16	21	18		
	π[e]	0.0032	0.0002	0.0040	0.0041	76.1	4.7
	R/S[f]	0.19	7.75	0.38	0.14		
	D[g]	0.1626	0.0396	–0.2711	0.2069		
sd1 (1,824 bp)							
	N	18	17	14	13		
	S	15	1	36	32		
	p	0.0040	0.0000	0.0111	0.0102	36.7	0.0
	R/S	0.00	–	0.00	0.00		
	D	–0.0582	–1.1639	–0.1939	–0.3082		
Wx (872 bp)							
	N	12	18	6	6		
	S	23	11	34	16		
	π	0.0045	0.0012	0.0098	0.0085	53.6	14.3
	D	0.7790	–1.1735	–0.2947	–0.1241		
RFT1 (1,536 bp)							
	N	13	17	9	6		
	S	38	26	56	59		
	π	0.0068	0.0016	0.0096	0.0128	31.0	5.8
	R/S	0.42	0.50	0.55	0.53		
	D	1.4153	–2.3771	–0.4808	–0.5070		
C (1,294 bp)							
	N	15	14	5	5		
	S	8	11	16	17		
	π	0.0027	0.0005	0.0087	0.0093	31.0	5.8
	R/S	0.59	0.95	0.28	0.28		
	D	0.9645	0.3285	–0.2673	–0.6845		
alk (4,952 bp)							
	N	10	15	7	5		
	S	61	8	65	57		
	π	0.0059	0.0004	0.0078	0.0066	83.1	5.6
	R/S	0.25	0.56	0.35	0.26		
	D	0.2869	–0.5750	0.9055	–0.4310		

[a]*RFT1* is a rice homolog of *Arabidopsis FT* (AB062675) and *alk* is *soluble starch synthase II* (AP003509). Alignment length examined is shown in parentheses. I and J show indica and japonica (including javanica) types. P and A show perennial and annual forms of wild rice. [b]π divided by total diversity of the wild ancestor. [c]Number of sequences sampled. Total number of segregating sites. [e]Average number of pairwise differences calculated on silent sites. Indels were included as a single mutation. [f]Ratio of replacement substitution to synonymous substitution. Tajima's D statistics based on all sites.

References

Allen DJ, Ort DR. 2001. Impacts of chilling temperatures on photosynthesis in warm-climate plants. Trends Plant Sci. 6:36-42.

Ashikari MA, Sasaki A, Ueguchi-Tanaka M, Itoh H, Nishimura A, Datta S, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M. 2002. Mutation in a gibberellin biosynthetic gene, *GA20 oxidase*, contributed to the rice 'Green Revolution'. Breed. Sci. 52:143-150.

Dung LV, Inukai T, Sano Y. 1998. Dissection of a major QTL for photoperiod sensitivity in rice: its association with a gene expressed in an age-dependent manner. Theor. Appl. Genet. 97:714-720.

Harlan JR. 1975. Crops and man. Madison, Wis. (USA): American Society of Agronomy. 295 p.

Hirano HY, Eiguchi M, Sano Y. 1998. A single base change altered the regulation of the *Waxy* gene at the post-transcriptional level during the domestication of rice. Mol. Biol. Evol. 15:978-987.

Ichitani K, Okumoto Y, Tanisaka, T. 1998. Genetic analysis of the rice cultivar Kasalath with special reference to two photoperiod sensitivity loci, *E1* and *Se-1*. Breed. Sci. 48:51-57.

Itoh Y, Sano Y. 1999. Cryptic differences in the *Se1* alleles as revealed by curvilinear regression analysis under controlled conditions. Rice Genet. Newsl. 16:57-59.

Izawa T, Takahashi Y, Yano M. 2003. Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Curr. Opin. Plant. Biol. 6:113-120.

Jennings PR, Aquino RC. 1968. Studies on competition in rice. 3. The mechanism of competition among phenotypes. Evolution 22:529-542.

Kinoshita T. 1995. Report of the committee on gene symbolization, nomenclature and linkage groups. Rice Genet. Newsl. 12:9-153.

Kratsch HA, Wise RR. 2000. The ultrastructure of chilling stress. Plant Cell Environ. 23:337-350.

Lewontin R. 2000. The triple helix. Cambridge, Mass. (USA): Harvard University Press. 129 p.

Nagamine T, Nakagahra M. 1990. Genetic variation of chilling injury at seedling stage in rice, *Oryza sativa* L. Jpn. J. Breed. 40:449-455.

Nagano H, Onishi K, Ogasawara M, Horiuchi Y, Sano Y. 2005. Genealogy of the "Green Revolution" gene in rice. Genes Genet. Syst. 80:351-356.

Nakai K, Inagaki Y, Nagata H, Miyazaki C, Iida S. 1998. Molecular characterization of the gene for dihydroflavonol 4-reductase of *japonica* rice varieties. Plant Biotech. 15:221-225.

Oka HI. 1988. Origin of cultivated rice. Tokyo (Japan) and Amsterdam (Netherlands): JSSP and Elsevier. 254 p.

Okumoto Y, Ichitani K, Inoue H, Tanisaka T. 1996. Photoperiod insensitivity gene essential to varieties grown in the northern limit region of paddy rice (*Oryza sativa* L.) cultivation. Euphytica 92:63-66.

Olsen KM, Purugganan MD. 2002. Molecular evidence on the origin and evolution of glutinous rice. Genetics 162:941-950.

Purugganan MD, Wessler SR. 1994. Molecular evolution of the plant *R* regulatory gene family. Genetics 138:849-854.

Rieseberg LH, Widmer A, Arntz AM, Burke JM. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Philos. Trans. R Soc. Lond. B Biol. Sci. 358:1141-1147.

Saitoh K, Onishi K, Mikami I, Thidar K, Sano Y. 2004. Allelic diversification at the *C* (*OsC1*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. Genetics 168:997-1007.

Saitou N, Yamamoto F. 1997. Evolution of primate ABO blood group genes and their homologous genes. Mol. Biol. Evol. 14:399-411.

Sano Y. 1984. Differential regulation of *waxy* gene expression in rice endosperms. Theor. Appl. Genet. 68:467-473.

Schlichting CD, Pigliucci M. 1998. Phenotypic evolution: a reaction norm perspective. Sunderland (UK): Sinauer Associates Inc. 340 p.

Takahashi M-E. 1982. Gene analysis and its related problems. J. Fac. Agric. Hokkaido Univ. 61:91-142.

Tanaka A, Kawano K, Yamaguchi J. 1966. Photosynthesis, respiration and plant type of the tropical rice plant. Int. Rice Res. Inst. Tech. Bull. 7:1-46.

Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063-1066.

Thomashow MF. 1990. Molecular genetics of cold acclimation in higher plants. Adv. Genet. 28:99-131.

Tsai KH. 1986. Gene loci and alleles controlling the duration of basic vegetative growth of rice. In: Rice genetics. Manila (Philippines): International Rice Research Institute. p 351-359.

Tsai KH, Oka HI. 1970. Genetic studies of yielding capacity and adaptability in crop plants. 4. Effects of an earliness gene, m^b, in the genetic background of a rice variety, Taichung 65. Bot. Bull. Acad. Sin. 11:16-26.

Vergara, BS, Chang TT. 1985. The flowering response of the rice plant to photoperiod: a review of the literature. 4th ed. Manila (Philippines): International Rice Research Institute. 61 p.

Notes

Authors' address: Plant Breeding Laboratory, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan.

Lessons from applying genomics to wheat and barley improvement

P. Langridge

On the surface, wheat and barley have little to offer the rice genomics research community. They have very large genomes without a physical map, making positional cloning complex, and they are difficult to transform, which hinders the functional analysis of genes and delivery of transgenic technologies. However, shifts in plant genomics research into understanding the basis of diversity and mechanisms involved in creating and maintaining genome complexity have shifted research from a model organism toward more complex species. Wheat and barley are becoming increasingly attractive organisms for many of the new genomics studies. Several key tools have been important for this change, including detailed and well-phenotyped populations, mapping of a large collection of ESTs, and studies of synteny with rice and maize.

Importantly, wheat and barley are widely adapted and there has been extensive monitoring and archiving of genotypes and associated phenotypic data. We also have populations adapted to specific environments and end-uses that have resulted from a long history of selective breeding. These advantages are becoming increasingly significant as analytic tools improve. Early genomics efforts in wheat and barley have delivered useful markers for application in breeding programs and identified key regions of the genome that carry disease-resistance loci, tolerance of abiotic stresses, and components of quality. The expanding resource base for wheat and barley genomics and the new insights being gained into genome organization and behavior of these species offer improvements in our ability to identify new sources of variation and to implement this information in breeding programs.

Keywords: Wheat, barley, synteny, genome, breeding, selection, abiotic stresses

Genomic research has been based on the study of a limited number of model organisms (Davis 2004) that were chosen for their small genome size and experimental tractability. However, the current thrusts of genome science are resulting in a new vision in which the study of diversity and organism complexity is gaining prominence, often at the expense of model organisms.

Developments in crop genomics signal where the significance of "non-crop models" declines as accessibility to genome technologies improves and the social relevance of crop genomics to deliver "public goods" gains prominence. Rice has been ideally positioned because of its major status as a crop and its small genome but it still lags a long way behind *Arabidopsis* as a model for gene discovery and the study of plant developmental processes.

Wheat and barley would appear to have few advantages for genomics studies. The genomes are large and poorly characterized relative to the model species. However, these species do have some potentially valuable characteristics. The level of abiotic tolerance shown by the Triticeae is greater than that of most other crop species and the diversity in the landrace and wild gene pool for wheat and barley tells us that still greater tolerance is achievable.

Wheat, barley, and rye are all closely related. Indeed, stable hybrids can be produced between wheat and barley (Blanco et al 1986) and wheat and rye (Triticale). The high level of genome similarity between wheat, rye, and barley has been recognized for some time. Not only are the sequences of individual genes similar between the species, but gene order on the chromosomes is highly conserved. Indeed, it is possible to replace many wheat chromosomes with related chromosomes from rye or barley. The gene order across the grasses is generally well maintained and clear relationships can be drawn across the grass genomes (Devos 2005). However, at the single-gene level, the order does start to break down. For example, a detailed comparison of a 20-cM region of 3DS of wheat with rice chromosome 1 showed about a 20% loss of synteny. In addition, certain classes of genes, such as the race-specific disease-resistance genes, are of recent evolutionary origin, and these are not usually located at syntenous regions. However, this problem does not seem to apply to many of the key loci controlling abiotic stress tolerance and the rice genome sequence (Yu et al 2002, Goff et al 2002) has proved to be a powerful tool for positional cloning of genes from wheat and barley (Feuillet et al 2003, Yan et al 2003).

Genome structure of the Triticeae

Crop plants such as bread wheat (*Triticum aestivum* L.) were considered good models for cytogenetic investigations and polyploidy research. Wheat has one of the largest and most complex genomes of any species. It is an allopolyploid containing three different ancestral genomes (designated A, B, and D), each of which has seven pairs of homologous chromosomes (2n=6x=42). The homologous chromosomes and genes in the different ancestral genomes are referred to as "homoeologous." Although these genomes are similar in gene content and order, chromosome pairing at meiosis is restricted to homologous chromosomes. This results in disomic inheritance, which greatly simplifies the pattern and interpretation of segregation data. The genome size of wheat (17,000 Mb), which is approximately five times the size of the human genome, was initially viewed as an impediment to genomics research. The start of coordinated genomics initiatives with a global perspective has dramatically changed our knowledge base and led to new opportunities for wheat genomics research.

Two key features of the wheat genome cannot be addressed through studies of the simple model species: polyploidy and the relationship between the large genome size and chromosome behavior, particularly with respect to recombination. To some extent, it is the interrelationship between these features that makes wheat a singularly fascinating organism to study.

Polyploidy is often associated with rapid genetic and epigenetic changes (Osborn et al 2003). Novel patterns of gene expression occur in polypoloids that are not observed in diploid progenitor species (Galitski et al 1999, Osborn et al 2003). The analysis of expression patterns of homoeologous genes in wheat, based around the use of ESTs generated from diverse tissues, has shown that homoeologous genes can be expressed in one but are silent in one or both of the remaining genomes (Mochida et al 2004). Further, the tissue specificity of homoeologous genes was also found to vary. It was particularly surprising to find that 72% of the homoeoloci studied showed expression from only one or two genomes. New work with the wheat Affymetrix array has indicated that the endosperm shows particularly high levels of "genome-specific" gene expression (Schreiber, Baumann, and Langridge, unpublished). The observation that many widely studied plants are ancient polypoids, including rice and *Arabidopsis*, could mean that the altered gene expression seen between homoeologous chromosomes in wheat is a common phenomenon.

Wheat provides an excellent model for studying this effect since synthetic polyploid wheats can be readily generated and wheat lines are available with varying numbers of specific chromosomes, such as the nulli/tetra lines where one chromosome pair is replaced by an additional pair of homoeologous chromosomes, and addition and translocation lines with individual barley, rye, or alien chromosomes or chromosome arms. Relating and understanding altered patterns of gene expression that accompany polyploidy remain an outstanding challenge worthy of pursuit in species such as wheat and may well be related to heterotic effects seen in many species. An attraction in studying these effects in wheat is that it can be readily compared with the diploid crop plant barley. Barley is a normal diploid but it is closely related to wheat. Since both barley and wheat were domesticated at about the same time and both have been actively bred in recent years, a comparison between the genome structure and behavior of the two provides a powerful means of looking at the molecular impact of polyploidization and a tool for examining epistatic and heterotic effects.

Recombinational behavior and chromosome pairing in wheat are also showing some unusual features. It has been known for some time that recombination in wheat chromosomes is focused in the telomeric regions so that the gene position along the chromosomes will affect the exposure of that gene to recombination activity. Genes subject to rapid change, such as many race-specific disease resistance genes, are located in the recombinogenic telometric regions, while more highly conserved genes tend to be positioned closer to the centromere (Dvorak et al 1997, Kunzel et al 2000). The large genome size of wheat may provide a mechanism for developing and maintaining a strong recombination gradient along the chromosomes.

The development of large sets of markers for use in wheat and barley breeding programs has provided an opportunity to look at historical recombination events within

germplasm pools used for wheat improvement. In an early study, we found that large linkage blocks, usually related to alien gene introgressions, are maintained within the wheat genome (Paull et al 1998). The low rates of recombination seen between chromosomes of wheat and its wild relatives were not surprising and have been seen in several studies (for example, Paull et al 1994, Dubcovsky et al 1996). What has been far more surprising has been the outcome of recent linkage disequilibrium studies in wheat and barley that show very high rates of linkage disequilibrium extending over long distances. Work in Australia and in Europe suggests that wheat breeding programs have essentially been shuffling whole chromosomes and there has been little selection for recombinational events (Kruger and Langridge, unpublished; Koebner, personal communication). This result may be a reflection of the nature of wheat selection and breeding in Australia and Europe but it does imply that variation is not being fully exploited in wheat improvement and it does help explain a number of limitations and problems wheat breeders have faced in using specific germplasm pools.

Resources available for genomics research in wheat and barley

Although we do not have access to a full genome sequence of wheat or barley and it is unlikely that these sequences will become available in the foreseeable future, a wide range of other resources can be used for gene and genome characterization. The Triticeae generally show good levels of tolerance of abiotic stresses when compared with rice or maize and it is here that they are likely to be of most value. Wheat, barley, rye, and triticale are grown in highly diverse environments, including some of the most hostile to cropping. The Triticeae provide an opportunity for discovery of genes controlling tolerance of a wide range of stresses, including salinity, water deficit, nutrient deficiency, heat, acid soils, cold, and frost. These genes then provide a target for rice improvement either through transgenic technologies or as a tool for screening for novel alleles in wild rice germplasm.

The key resources available from wheat and barley are the extensively mapped populations and the widely phenotyped germplasm collections. An example is shown in Table 1, which lists differential germplasm available in Australia for studying a range of abiotic stress tolerances. The following section is modified from a recent review on the genomics of abiotic stress tolerance in cereals (Langridge et al 2006).

Mutant populations

Mutation breeding has been extensively used in the cereals, with several important varieties resulting for the selection of mutant phenotypes (Ahloowalia and Maluszynski 2001). However, systematic development of mutant populations as a genomics resource has only commenced recently. A large barley mutant population constructed for mutation screening based around the detection of single-base mismatches was described recently (Caldwell et al 2004) and a similar population was used to identify a series of mutations in the granule-bound starch synthase I, or GBSSI, of wheat (Slade et al 2005). Wheat appears to be particularly suitable for mutant screening since it is able to carry a very high mutation load, presumably because of polyploidy.

Table 1. Wheat, barley, and some rye germplasm is available for analysis of the genetic control of abiotic stress tolerance. Shown are a selection of lines and germplasm accessions that have already been used to construct mapping populations or are in the process of being used in Australia. Many further populations are being developed in other parts of the world.

Stress	Wheat			Barley		
	Adapted		Nonadapted	Adapted		Nonadapted
Boron tolerance	Halberd Bt-Schomburgk Frame Yitpi Spear Dagger Trident Krichauff	×	Cranbrook Schomburgk Janz Sunco Molineux	Sahara	×	Clipper
Mn efficiency	RAC892 Trident		Most Australian lines	Amaji Nijo Haruno Nijo	× ×	WI2585 Galleon
Cu efficiency	Rye translocation and recombinant lines		Most Australian lines			
Zn efficiency	RAC892 Trident		Most Australian lines	Screening under way		
Phosphorus-use efficiency	Brookton		Krichauff	Sahara	×	Clipper
Nitrogen-use efficiency	CIMMYT lines		Frame and Yitpi, poor response to high N	Most current lines		WI3102, poor response to N
Al tolerance	Dollarbird Diamondbird Karlgarin Maringa BR35 Embrapa-16 Rye translocation lines		South Australian lines	Brindabella WB229 Dayton	× × ×	Harrington Yambla Harlan Hybrid
Drought tolerance	CIMMYT *Aegilops tauschii* selections Introgressed *Ae tauschii* lines		Recurrent parents	2 ICARDA populations *H. spontaneum*	 ×	 Barque

Table 1 continued

Table 1 continued.

Stress	Wheat			Barley		
	Adapted		Nonadapted	Adapted		Nonadapted
Proven performers in low-rainfall, low-yielding environments	Excalibur: tillering habit		European and North American lines	Galleon	×	Haruno Nijo
	RAC875: waxy leaves, different plant structure			Chebec	×	Harrington
				Sloop	×	Alexis
				Sloop	×	Halcyon
	Landraces and wild relatives			Landraces and *H. spontaneum* accessions		
Cold tolerance				Diverse list based on geographic origin		Australian lines
Salinity tolerance	Halberd	×	Cranbrook			
	Machete		European germplasm	YU6472: Chinese landrace		Australian lines
	Most South Australian lines are quite good					

Slade et al (2005) identified 246 alleles after screening for each homoeolog in only 1,920 mutated individuals. Several projects are currently under way around the world to develop mutated populations of both wheat and barley and diploid progenitors of wheat, and many of these will be available as public resources.

Work has also been under way to develop transposon-tagged populations of barley using the maize *Ac/Ds* system (Cooper et al 2004). This will also provide an important resource for functional analysis of cloned genes.

Mapping and map-based cloning

The importance of wheat and barley as crop species and the long history of systematic breeding have resulted in extensive information on the genetic control of a wide range of traits and also in the identification of broad diversity for stress and disease tolerance. The advent of molecular marker techniques and links to breeding programs has resulted in a large expansion of mapping studies. Importantly, many of these studies have involved field evaluation of abiotic stress tolerance. The information shown in Table 2 covers only recent studies but indicates both the diversity and complexity of abiotic stress mapping programs.

Molecular markers have also proved important in surveys of cultivated, landrace, and wild relatives of wheat and barley (Table 3). From these studies, extensive variation has been identified and this is being used directly in several wheat and barley breeding programs but it also allows extension of the mapping work. The diversity found in

Table 2. Recent publications describing the mapping of QTLs controlling tolerance of a range of abiotic stresses in wheat and barley. In many cases, additional QTLs were found but were either of low significance or could not be clearly located to chromosomes.

Stress	Chromosome	Reference
Wheat		
Boron tolerance	7B, 7D	Jefferies et al (2000)
Sprouting resistance	5AL, 6A, 3B, 7B	Zanetti et al (2000)
Preharvest sprouting	6B, 7D	Roy et al (1999)
Cold tolerance	5A	Vagujfalvi et al (2000)
Vernalization and cold	5A, 5D	Snape et al (2001)
Flooding tolerance	15 QTLs	St Burgos et al (2001)
Al tolerance	4DL	Riede and Anderson (1996), Luo and Dvorak (1996)
Na^+/K^+ discrimination	4D	Allen et al (1995)
Water stress tolerance (paraquat tolerance)	5A, 5B, 5D	Altinkut and Gozukirmizi (2003)
Drought, low temperature, salinity	Groups 5 & 7	Cattivelli et al (2002)
Nutrient, drought, and salt stress	17 QTL clusters	Quarrie et al (2005)
Flag-leaf senescence	2B, 2D	Verma et al (2004)
Agropyron, drought	3E, 5E, 7E	Farshadfar et al (2004)
Na exclusion	2AL	Lindsay et al (2004)
Barley		
Boron tolerance	2H, 3H, 4H, 6H	Jefferies et al (1999)
Manganese efficiency	4HS	Pallotta et al (2000)
Frost at anthesis	2H, 5H	Reinheimer et al (2004)
Drought	23 QTLs, all except 6H	Peighambari et al (2005)
Drought	Over 70 QTLs	Diab et al (2004), Teulat et al (2001, 2002, 2003)
Seedling desiccation at germination	7 QTLs	Zhang et al (2005)

abiotic stress tolerance will provide an important resource for validation of candidate genes and will also provide a mechanism for rapid delivery of genomics outcomes to pragmatic breeding programs. The concept here is to use candidate genes identified from genomics studies to assess stress tolerance mechanisms used in the landrace or wild lines that show elevated tolerance. Where variation is identified in candidates, either in the structural gene or in expression levels, the locus can be transferred and tested in nonadapted or cultivated germplasm for assessment. In this way, the variation is used to not only validate candidates for stress tolerance but also to provide a tool for allele discovery. Desirable alleles can be transferred by conventional breeding and selection.

Table 3. Recent studies using molecular markers to survey germplasm for salt and drought tolerance.

Crop	Accessions screened	Screening technique	Reference
Wheat			
Salt tolerance in landraces	400	Salt tolerance	Liu et al (1998)
Salt tolerance in wild relatives	14	Salt tolerance	Wang et al (2003)
Drought in wild relatives		Response to osmotic shock	Tyankova et al (2004)
Drought in landraces	28	Drought-tolerant lines from	Moghaddam et al (2005)
Drought in wild emmer	110	C isotope discrimination	Peleg et al (2005)
Barley			
Drought in cultivated species	80	Water stress index, yield	Rizza et al (2004)
Drought in cultivated, landrace,	600	Range of traits under drought-	Forster et al (2004)
Drought in wild barley	275	Microclimate based on position	Nevo et al (2005)

Positional cloning does require high-resolution linkage maps and this has involved large segregating populations. Association mapping may provide an alternative but this is not sufficiently developed in wheat and barley for application at present (Powell and Langridge 2004). The availability of the rice genome sequence and the generally strong synteny between rice and wheat and barley means that rice can be used to generate potential markers close to the target locus. Since most known abiotic stress loci have been identified as QTLs, high-resolution mapping is complicated by difficulties in clear definition of phenotypes. Breaking the trait down to well-defined components or eliminating confounding loci from the populations may help deal with this difficulty.

The extensive mapping of particular traits, such as drought tolerance, in multiple cereals and using different populations may provide a means for locating common loci. For example, components of drought tolerance have been mapped in almost all major cereals by several groups around the world. In wheat, we are consistently seeing QTLs on group 5 and 7 chromosomes appearing in the analyses. However, the merging of populations has not yet been attempted. Similarly, one can look for syntenous loci controlling salt tolerance in wheat, barley, rice, and maize and use the comparison to generate a large "virtual" mapping population. Computational tools for these types of analysis are in the process of development and this approach may be feasible in the not too distant future.

Transcript profiling

Estimates of gene number in the cereals are similar to those of other complex organisms; for example, in barley, gene number estimates range from around 30,000 to 50,000 (Zhang et al 2004). The situation is complicated by the three genomes of wheat. Unusual patterns of gene expression can occur in polyploids. Analysis of gene expression in wheat using EST databases showed that homoeologous genes can be expressed in one genome but be silent in one or both of the remaining genomes (Mochida et al 2004). The tissue specificity of expression from homoeolgous genes may also change. Consequently, a gene in one genome may be expressed in roots, whereas the homoeologs are expressed in leaf tissue.

There are now well over 1 million cereal EST sequences in the public databases, with wheat and barley dominating. The large number of ESTs and the diversity of cDNA libraries that have been used to generate the sequences have made "electronic Northerns" a useful method for assessing gene expression and this provides a good first measure of transcript abundance. Several microarray and macroarray platforms have been generated for the cereals. For wheat and barley, there are several proprietal arrays such as the 10,000 cDNA array reported by Leader (2005). More recently, Affymetrix arrays have been developed for both wheat and barley (Close et al 2004). In a reference experiment using a series of well-defined developmental stages, the 22K Barley 1 Gene Chip identified 18,481 transcripts showing expression above background (Druka et al, personal communication). A similar reference experiment has been conducted using the Affymetrix Wheat Gene Chip. The wheat chip carries 61,127 probe sets representing 55,052 transcripts. However, fewer than 30,000 are

from high-quality sequence data and it is not yet clear how the remaining probe sets will be used.

Currently, there are no published reports on the use of the barley or wheat chips for studying altered gene expression in response to abiotic stress or disease but several such experiments are now under way and we should see reports on the outcomes appearing in the literature within the next year or two.

Proteomics and metabolomics

Again, there have been no reports of the application of proteomics to the study of biotic or abiotic stress tolerance in wheat and barley, although proteome studies have been conducted of wheat leaf (Bahrman et al 2004), grain (Skylas et al 2005), and lemma (Woo et al 2003). These results have indicated the feasibility of differentiating and identifying large numbers of proteins from defined tissues of wheat. The leaf proteomic study resolved 541 proteins, of which 55 were sequenced (Bahrman et al 2004).

The importance of metabolite changes during plant responses to various stresses suggests that detailed metabolite profiling may provide valuable insights into stress response mechanisms. Metabolomics is a relatively new area of research and there are no published reports on its application to stress tolerance in cereals. We are currently using this technique to assay levels of well over 200 metabolites in stressed wheat and barley plants and the results are providing a greatly expanded view on the processes underlying stress responses (Roessner and Bacic, personal communication).

Path to genomics delivery

There are two paths for delivery of genomics outcomes: first, through transformation, and second, through the use of molecular markers. There has been no commercial release of genetically modified (GM) wheat or barley although GM field trials with both species have been conducted in several countries. Conversely, the use of molecular markers to track loci and genome regions in wheat and barley is now routinely applied in many breeding programs. The location of major loci is known for many disease-resistance genes, tolerances of abiotic stresses, and quality traits. The shift from RFLP to SSR markers has also been important in facilitating the tracking of genes. For markers to be effective, they must be closely linked to the target locus and be able to detect polymorphisms in material likely to be used in a breeding program. Generally, about ten different SSRs are required within 10 cM of a target locus to give a reasonable chance of having markers that can be implemented. The prime applications of markers in wheat and barley breeding have been in backcross breeding, where loci are tracked in defect elimination, for the introgression of recessive traits and in the selection of lines with a genome makeup close to that of the recurrent parent. In progeny breeding, markers have proved valuable in building crucial parents and in enriching F_1s from complex crosses. Markers have also improved strategies for gene deployment and enhanced understanding of the genetic control of complex traits such as components of quality.

The recent developments that have occurred in molecular markers for wheat and barley have major implications for the future of the technology. Three key components are particularly significant. First, we now have markers closely linked to many traits of importance in breeding programs. Indeed, we have markers for more loci than we can screen in a conventional breeding program. Second, we have tools that allow marker scanning of the whole genome. SSR markers form the base for this analysis and highly multiplexed SSR screens have been developed for both wheat and barley. This has improved our capacity for whole-genome screens. Third, through association mapping projects, we have or are in the process of developing whole-genome fingerprints for many key lines and varieties of importance in wheat and barley breeding. We can see these developments particularly in barley, for which markers for many of the major disease-resistance clusters, for key components of malting quality, and for several loci conditioning tolerance of abiotic stresses are available. Indeed, now more than 20 different traits and over 30 loci are being screened in Australian barley breeding programs. Molecular markers are being used to screen for 13 loci controlling disease resistance (such as *Rrs1* for scald, *Ha1* and *Ha 2* for cereal cyst nematode resistance, and *Mlo* for powdery mildew resistance), seven abiotic stress-tolerance loci (including boron tolerance, nutrient efficiencies, and frost tolerance), nine quality-related loci (such as *Bamy1*, diastatic power, and malt extract), and three physiological traits (such as the *denso* dwarf and hull-less loci). For wheat, the situation is not as advanced but there are still many tagged loci that are or will be used in the near future. The various groups of markers currently in use in the Australian wheat and barley improvement programs are summarized in Table 4. From the table, it will be clear that many loci can now be successfully tracked. However, the actual number used in a particular program and for a given breeding strategy will be much smaller.

An example of the complexity of marker screening in a wheat breeding program is illustrated by the diagram in Figure 1. The wheat breeding program used to construct the map in Figure 1 is large and has adopted molecular markers to support backcrossing and related applications. Although the map in the figure has been condensed to a single genome, there are several important features apparent that influence the ways in which markers are deployed. First, the target loci are distributed across the genome, with important loci present on each chromosome group. Some loci are clustered. Of these, some can be moved as a group, such as the cluster on the short arm of 2A, while others are in repulsion, such as the root lesion nematode resistance on 7A and the yellow flower color locus. This type of information becomes important in designing a breeding strategy. For example, the 7A loci need to be specifically targeted to identify a line where a recombination event has occurred so that the desirable alleles are combined at each locus, while the 2A region will require only flanking markers to ensure that the desirable combination remains intact. In both cases, molecular markers offer the most efficient method for identifying the desired recombination event (in the 7A example) that separates the desirable from the undesirable alleles or for ensuring that no recombination has occurred (in the 2A example) and that the desirable combination is maintained.

Table 4. Numbers of loci currently being tracked with molecular markers in the Australian wheat and barley improvement programs. Totals show the total number of loci being tracked and also the number of traits; for example, three QTLs for boron tolerance are being tracked in barley.

Crop	Number of loci
Barley	
Disease	13
Abiotic stress	7
Quality	9
Boron tolerance	3
Total	32 (24 traits)
Wheat	
Disease	18
Abiotic stress	6
Quality	10
Total	34 (27 traits)

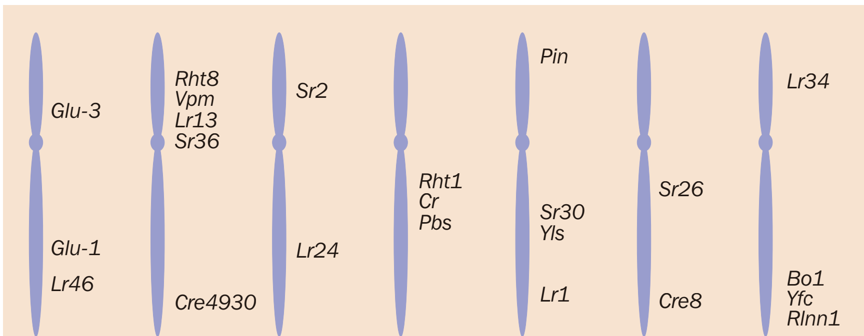


Fig. 1. Loci currently being tracked with molecular markers in an Australian wheat breeding program. The wheat genome view has been simplified by representing each chromosome group rather than the individual A, B, and D genomes. Markers have been placed on the map as if there were only a single genome (from Langridge 2005).

Molecular tools have several important implications for the future of marker-assisted selection (MAS) and breeding strategies in general. Existing strategies for MAS began with a view of markers as providing a rapid and cheap alternative to bioassays and they have largely been used in this role. While highly successful, this strategy does not fully use the technology. The key limitation to an expansion of the scale and complexity of marker use is the size of the populations that would be required if one were to try to select for a large number of loci simultaneously. A further important feature of recent advances has been related to how we best take advantage of the genome information that has been generated for both wheat and barley. We know, for example, that chromosome 2H in barley and group 7 chromosomes of wheat carry clusters of genes (see Fig. 1), often in repulsion, that we would like to break up. Again, conventional use of markers has not been very effective in using such genome regions. Conversely, we know that there are some chromosomes where there is little allelic variation between lines and it is a waste of effort to try to break these up in a breeding program.

The key challenge of new work is to investigate strategies for whole-genome breeding: that is, to see how we can use genome information in the form of graphical genotypes and known locations of key loci and marker tags for both desirable and undesirable alleles, to design optimal breeding strategies that integrate as much of the available information as possible.

Relevance to rice

The key question arising out of the above discussion is, What value do the genomics programs in wheat and barley have for practical rice improvement? In many respects, rice has proved to be a powerful model that has helped researchers in other cereals to discover genes and it is difficult to find examples where information generated in wheat and barley, or maize, has found application for rice. However, this may really be just a reflection of the current state of cereals research. Many of the genes that have been targeted in wheat and barley have little relevance to rice. This applies in particular to genes controlling resistance to disease (Feuillet et al 2003) and traits such as vernalization (Yan et al 2003). The benefits to rice from gene discovery work in wheat and barley are more likely to come from studies of abiotic stress tolerance. This is an area of intense research by many groups and should provide important insights into abiotic stress tolerance in all cereals, at least all C_3 cereals.

A further important lesson that rice researchers and breeders may be able to learn from the wheat and barley community is the development and application of molecular markers. As with rice, most major breeding programs in wheat and barley have been in the public domain and the development of networks of research and breeders has been the key delivery mechanism for technological advances. The early establishment of national marker development and implementation programs in Australia was crucial to technology delivery.

References

Ahloowalia BS, Maluszynski M. 2001. Induced mutations: a new paradigm in plant breeding. Euphytica 118:167-173.

Allen GJ, Jones RGW, Leigh RA. 1995. Sodium-transport measured in plasma-membrane vesicles isolated from wheat genotypes with differing K^+/Na^+ discrimination traits. Plant Cell Environ. 18:105-115.

Altinkut A, Gozukirmizi N. 2003. Search for microsatellite markers associated with water-stress tolerance in wheat through bulked segregant analysis. Mol. Biotechnol. 23:97-106.

Bahrman N, Negroni L, Jaminon O, Le Gouis J. 2004. Wheat leaf proteome analysis using sequence data of proteins separated by two-dimensional electrophoresis. Proteomics 4:2672-2684.

Blanco A, Fracchiolla GV, Greco B. 1986. Intergenic wheat × barley hybrid. J. Heredity 77:98-100.

Caldwell DG, McCallum N, Shaw PP, Muehlbauer GJ, Marshall DF, Waugh R. 2004. A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). Plant J. 40:143-150.

Cattivelli L, Baldi PP, Crosatti C, Di Fonzo N, Faccioli PP, Grossi M, Mastrangelo AM, Pecchioni N, Stanca AM. 2002. Chromosome regions and stress-related sequences involved in resistance to abiotic stress in Triticeae. Plant Mol. Biol. 48:649-665.

Close TJ, Wanamaker SI, Caldo RA, Turner SM, Ashlock DA, Dickerson JA, Wing RA, Muehlbauer GJ, Kleinhofs A, Wise R. 2004. A new resource for cereal genomics: 22K barley GeneChip comes of age. Plant Physiol. 134:960-968.

Cooper LD, Marquez-Cedillo L, Singh J, Sturbaum AK, Zhang S, Edwards V, Johnson K, Kleinhofs A, Rangel S, Carollo V, et al. 2004. Mapping *Ds* insertions in barley using a sequence-based approach. Mol. Genet. Genom. 272:181-193.

Diab AA, Teulat-Merah B, This D, Ozturk NZ, Benscher D, Sorrells ME. 2004. Identification of drought-inducible genes and differentially expressed sequence tags in barley. Theor. Appl. Genet. 109:1417-1425.

Davis RH. 2004. The age of model organisms. Nat. Rev. Genet. 5:69-76.

Devos K. 2005. Updating the 'crop cycle.' Curr. Opin. Plant Biol. 8:155-162.

Dubcovsky J, Luo MC, Zhong GY, Bransteitter R, Desai A, Kilian A, Kleinhofs A, Dvorak J. 1996. Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics 143:983-999.

Dvorak J, Luo M-C, Yang Z-L. 1997. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing *Aegilops* species. Genetics 148:423-434.

Farshadfar E, Mohammadi R, Farshadfar M, Sutka J. 2004. Locating QTLs controlling field and laboratory predictors of drought tolerance in *Agropyron* using multiple selection index. Cereal Res. Comm. 32:17-24.

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B. 2003. Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc. Natl. Acad. Sci. USA 100:15253-15258.

Forster BPP, Ellis RPP, Moir J, Talame VOL, Sanguineti MC, Tuberosa R, This D, Teulat-Merah B, Ahmed I, Mariy S, et al. 2004. Genotype and phenotype associations with drought tolerance in barley tested in North Africa. Ann. Appl. Biol. 144:157-168.

Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. 1999. Ploidy regulation of gene expression. Science 285:251-254.

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296:92-100.

Jefferies SPP, Barr AR, Karakousis A, Kretschmer JM, Manning S, Chalmers KJ, Nelson JC, Islam A, Langridge PP. 1999. Mapping of chromosome regions conferring boron toxicity tolerance in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 98:1293-1303.

Jefferies SPP, Pallotta MA, Paull JG, Karakousis A, Kretschmer JM, Manning S, Islam A, Langridge P, Chalmers KJ. 2000. Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). Theor. Appl. Genet. 101:767-777.

Kunzel G, Korzun L, Meister A. 2000. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154:397-412.

Langridge P, Paltridge N, Fincher GB. 2006. Functional genomics of stress in cereals. Brief. Functional Genomics Proteomics 4:343-354.

Langridge P. 2005. Molecular breeding of wheat and barley. In: Tuberosa R, Phillips RL, Gale M, editors. In the wake of the double helix: from the Green Revolution to the Gene Revolution. Bologna (Italy): Avenue Media. p 279-286.

Leader DJ. 2005. Transcriptional analysis and functional genomics in wheat. J. Cereal Sci. 41:149-163.

Lindsay MP, Lagudah ES, Hare RA, Munns R. 2004. A locus for sodium exclusion (*Nax1*), a trait for salt tolerance, mapped in durum wheat. Functl. Plant Biol. 31:1105-1114.

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell 10:1391-1406.

Luo M C, Dvorak J. 1996. Molecular mapping of an aluminum tolerance locus on chromosome 4D of Chinese spring wheat. Euphytica 91:31-35.

Mochida K, Yamazaki Y, Ogihara Y. 2004. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol. Gen. Genom. 270:371-377.

Moghaddam AE, Trethowan RM, William HM, Rezai A, Arzani A, Mirlohi AF. 2005. Assessment of genetic diversity in bread wheat genotypes for tolerance to drought using AFLPs and agronomic traits. Euphytica 141:147-156.

Nevo E, Beharav A, Meyer RC, Hackett CA, Forster B PP, Russell JR, Powell W. 2005. Genomic microsatellite adaptive divergence of wild barley by microclimatic stress in Evolution Canyon, Israel. Biol. J. Linnean Soc. 84:205-224.

Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19:141-147.

Pallotta MA, Graham RD, Langridge PP, Sparrow DHB, Barker SJ. 2000. RFLP mapping of manganese efficiency in barley. Theor. Appl. Genet. 101:1100-1108.

Paull JG, Chalmers KJ, Karakousis A, Kretschmer JM, Manning S, Langridge P. 1998. Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor. Appl. Genet. 96:435-446.

Paull JG, Pallotta MA, Langridge P, The TT. 1994. RFLP markers associated with *Sr22* and recombination between chromosome 7A of bread wheat and the diploid species *Triticum boeoticum*. Theor. Appl. Genet. 89:1039-1045.

Peleg Z, Fahima T, Abbo S, Krugman T, Nevo E, Yakir D, Saranga Y. 2005. Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. Plant Cell Environ. 28:176-191.

Peighambari SA, Samadi BY, Nabipour A, Charmet G, Sarrafi A. 2005. QTL analysis for agronomic traits in a barley doubled haploid population grown in Iran. Plant Sci. 169:1008-1013

Powell W, Langridge P. 2004. Unfashionable species flourish in the 21st century. Genome Biol. 5:233

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusic D, Waterman E, et al. 2005. A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor. Appl. Genet. 110:865-880.

Reinheimer JL, Barr AR, Eglinton JK. 2004. QTL mapping of chromosomal regions conferring reproductive frost tolerance in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 109:1267-1274.

Riede CR, Anderson JA. 1996. Linkage of RFLP markers to an aluminum tolerance gene in wheat. Crop Sci. 36:905-909.

Rizza F, Badeck FW, Cattivelli L, Lidestri O, Di Fonzo N, Stanca AM. 2004. Use of a water stress index to identify barley genotypes adapted to rainfed and irrigated conditions. Crop Sci. 44:2127-2137.

Roy JK, Prasad M, Varshney RK, Balyan HS, Blake TK, Dhaliwal HS, Singh H, Edwards KJ, Gupta PP K. 1999. Identification of a microsatellite on chromosomes 6B and a STS on 7D of bread wheat showing an association with preharvest sprouting tolerance. Theor. Appl. Genet. 99:336-340.

Skylas D, Van Dyk D, Wrigley CW. 2005. Proteomics of wheat grain. J. Cereal Sci. 41:165-179.

Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D. 2005. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotech. 23:75-81.

Snape JW, Sarma R, Quarrie SA, Fish L, Galiba G, Sutka J. 2001. Mapping genes for flowering time and frost tolerance in cereals using precise genetic stocks. Euphytica 120:309-315.

St Burgos M, Messmer MM, Stamp PP, Schmid JE. 2001. Flooding tolerance of spelt (*Triticum spelta* L.) compared to wheat (*Triticum aestivum* L.): a physiological and genetic approach. Euphytica 122:287-295.

Teulat B, Borries C, This D. 2001. New QTLs identified for plant water status, water-soluble carbohydrate and osmotic adjustment in a barley population grown in a growth-chamber under two water regimes. Theor. Appl. Genet. 103:161-170.

Teulat B, Merah O, Sirault X, Borries C, Waugh R, This D. 2002. QTLs for grain carbon isotope discrimination in field-grown barley. Theor. Appl. Genet. 106:118-126.

Teulat B, Zoumarou-Wallis N, Rotter B, Ben Salem M, Bahri H, This D. 2003. QTL for relative water content in field-grown barley and their stability across Mediterranean environments. Theor. Appl.Genet. 108:181-188.

Tyankova N, Zagorska N, Dimitrov B. 2004. Study of drought response in wheat cultivars, stabilized wheat-wheatgrass lines and intergeneric wheat amphidiploids cultivated in vitro. Cereal Res. Comm. 32:99-105.

Vagujfalvi A, Crosatti C, Galiba G, Dubcovsky J, Cattivelli L. 2000. Two loci on wheat chromosome 5A regulate the differential cold-dependent expression of the *cor14b* gene in frost-tolerant and frost-sensitive genotypes. Mol. Gen. Genet. 263:194-200.

Verma VOL, Foulkes MJ, Worland AJ, Sylvester-Bradley R, Caligari PP DS, Snape JW. 2004. Mapping quantitative trait loci for flag leaf senescence as a yield determinant in winter wheat under optimal and drought-stressed environments. Euphytica 135:255-263.

Wang RRC, Li XM, Hu ZM, Zhang JY, Larson SR, Zhang XY, Grieve CM, Shannon MC. 2003. Development of salinity-tolerant wheat recombinant lines from a wheat disomic addition line carrying a *Thinopyrum junceum* chromosome. Intl. J. Plant Sci.164:25-33.

Woo SH, Kimura M, Higa-Nishiyama A, Dohmae N, Hamamoto H, Jong SK, Yamaguchi I. 2003. Proteome analysis of wheat lemma. Biosci. Biotech. Biochem. 67:2486-2491.

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. 2003. Positional cloning of the wheat vernalization gene *VRN1*. Proc. Natl Acad. Sci. USA 100:6263-6268.

Yu J, Hu S, Wang J, Wang G, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. spp. indica). Science 296:79-92.

Zanetti S, Winzeler M, Keller M, Keller B, Messmer M. 2000. Genetic analysis of pre-harvest sprouting resistance in a wheat × spelt cross. Crop Sci. 40:1406-1417.

Zhang F, Chen G, Huang Q, Orion O, Krugman T, Fahima T, Korol A, Nevo E, Gutterman Y. 2005. Genetic basis of barley caryopsis dormancy and seedling desiccation tolerance at the germination stage. Theor. Appl. Genet. 110:445-453.

Zhang HN, Sreenivasulu N, Weschke W, Stein N, Rudd S, Radchuk V, Potokina E, Scholz U, Schweizer P, Zierold U, et al. 2004. Large-scale analysis of the barley transcriptome based on expressed sequence tags. Plant J. 40:276-290.

Notes

Author's address: Australian Centre for Plant Functional Genomics, University of Adelaide, Glen Osmond, SA 5064, Australia.

The major chromosome pairing locus (*Ph1*) in hexaploid wheat: a perspective

G. Moore

Western civilization owes much of its foundation to pasta and bread wheat. These species are polyploidy, possessing multiple diploid sets of chromosomes. Pasta and bread wheats exist only because the *Ph1* locus stabilizes the pairing of these multiple related chromosomes at meiosis. It provides a high level of fertility and seed set. This article reviews current knowledge of the biological effect of this important locus. It provides insights into how one might induce pairing between related chromosomes for breeding.

Keywords: *Ph1* locus, telomeres, centromeres, meiosis, chromosome pairing

More than 70% of flowering plants are polyploids, possessing two or more sets of related chromosomes. This occurs as a result of the doubling of chromosomes following sexual hybridization within the same species (autopolyploidy) or between closely related species containing related but not completely homologous (homoeologous) genomes (allopolyploidy). Thus, some of the world's most important crops are allopolyploids, including wheat, canola, oats, cotton, tobacco, cultivated sugarcane, and cooking banana; a few are autopolyploid (e.g., potato). Others, such as maize, are cryptic allopolyploids. Their genomes have undergone extensive arrangements following polyploidization, which will have now partially concealed their polyploid origin. For allopolyploids to produce viable gametes and be fertile, they must behave as diploids during meiosis, with only identical chromosomes (homologous) pairing instead of homoeologs.

Breeders use interspecific crosses to introduce desirable genes from wild or related species. These hybrids are produced by sexual hybridization between a polyploid and a wild relative and contain a haploid set of polyploid and wild-relative chromosomes. In some cases, there is a low level of pairing and recombination between the wheat and wild-relative chromosomes, which is a problem for breeding. Understanding the barriers during meiosis that prevent pairing and recombination

of the chromosomes from the two parental species is therefore important for crop improvement strategies.

Meiosis: the basic process

During meiosis, the number of chromosomes needs to be halved before sexual reproduction. This is important because it ensures that the chromosome number does not double with each generation (reviewed by Zickler and Kleckner 1999). After each of the two homologs is replicated to form two sister chromatids, they will remain linked together (four chromatids in total). Each chromosome consisting of two chromatids must recognize and pair with its homolog from among all the chromosomes present in the nucleus. The homologs must then become intimately aligned along their entire lengths and a proteinaceous structure known as the synaptonemal complex (SC) must be assembled between them, a process called synapsis. In this way, meiotic recombination (the exchange of DNA strands between the homologs) is completed, resulting in the formation of chiasmata, physical links that hold the chromosomes together after disassembly of the SC. After the resolution of the physical links, the homologs separate during anaphase I. Thus, each homolog (consisting of two chromatids) will move to opposite poles of the spindle. Following a second division in which the two chromatids of each homolog separate, each gamete will carry only a single copy of each (one of the four chromatids) chromosome. Many components of meiotic recombination and synapsis machinery are known, especially from studies of yeast; however, little is understood about how homologs first recognize each other and how this is controlled. An unlikely model system, namely, hexaploid wheat, is starting to provide insights into this process.

Hexaploid wheat: its structure

In hexaploid (bread) wheat (an allopolyploid), chromosome 1A has a gene order similar to 1B and 1D (karyotype reviewed by Gill et al 1991; RFLP map in Gale et al 1997). Tetraploid (pasta) wheat possesses just the A and B homoeologous genomes, and lacks the D genome. Although there is a similar gene order on the homoeologs, the expression pattern between them is different (Mochida et al 2004). Moreover, dosage compensation means that, when a region of a chromosome is deleted, the transcription from homoeologous regions will compensate for the loss. Thus, the level of transcription is maintained. A major difference between the homoeologous chromosomes is in the repetitive DNA between the genes. Hexaploid and tetraploid wheat both behave as diploids at meiosis, with regular pairing at metaphase I. Thus, at meiosis, wheat chromosome 1A pairs with 1A, and not with either 1B or 1D.

Pairing control in hexaploid wheat

The diploid chromosome pairing behavior that restricts pairing to homologs rather than homoeologs in wheat is under genetic control. A single locus (termed pairing

homoeologous 1, *Ph1*) on the long arm of chromosome 5B has a major controlling effect (Riley and Chapman 1958, Sears and Okamoto 1958). Other loci that affect pairing in wheat have also been identified (e.g., on chromosomes 5AL, 5DL, 5AS, 3DS, 3AL, 3BL, and 3DL; reviewed by Feldman 1993). However, they do not compensate for the absence of the *Ph1* locus. Moreover, to date, little research has been undertaken on these loci compared with that on *Ph1*. In the presence of this locus (i.e., two copies of *Ph1*), chromosome pairing in tetraploid and hexaploid wheat during meiosis is largely restricted to homologous chromosomes, but, in deletion mutants lacking *Ph1*, increasing aberrant pairing can be observed at metaphase I with each generation of the mutant. These *Ph1* mutants accumulate rearrangements with each propagation (Sanchez-Moran et al 2001). To date, no chromosome substituted for 5B in hexaploid wheat from any wild species has been shown to compensate fully for the *Ph1* locus. However, the 5G chromosome from wheat relative *Triticum timopheevi*, when substituted for 5B, does provide a similar effect (Ozkan et al 2001). Other species probably carry loci equivalent to *Ph1* on 5B, but their effect may not be so dramatic. In fact, B chromosomes, which are essentially lumps of heterochromatin, can compensate for the *Ph1* locus in wheat hybrids (Dover and Riley 1972). Thus, the nature of the effect of *Ph1* on chromosome pairing in hexaploid and tetraploid wheat is different from the hybrid situation. Clearly, there are no homologs in the hybrid situation. So, in hexaploid and tetraploid wheat, *Ph1* needs to promote homologous pairing, whereas, in hybrids, *Ph1* needs to reduce homoeologous pairing.

Importance for breeding

Is *Ph1* important for plant breeding? Interspecific hybrids made between wheat lacking *Ph1* and its wild relatives often exhibit extensive pairing at metaphase I between the homoeologs but less in the presence of a single copy (Kimber et al 1981). This observation is of practical importance for plant breeding as there is the potential to generate recombinant chromosomes composed of part wheat and part nonwheat chromatin, allowing the introgression of exotic genetic material from wild relatives. Sears (1972) transferred disease resistance loci from wild relatives of wheat by recombining the chromosomes with those of wheat in lines deficient in the *Ph1* locus. More recently, the International Center for Maize and Wheat Improvement (CIMMYT) exploited *Ph1* to introgress chromosome segments from closely related species so that now 25% of all new varieties released in their program will contain material derived from related species.

Unfortunately, an important complication is that there is still a hierarchy in preferred pairing partners even in the absence of *Ph1*. Most pairing observed at metaphase I (80%) in hexaploid wheat interspecific hybrids lacking the *Ph1* locus occurs between wheat chromosomes from A and D genomes. However, chromosomes of A and B genomes of wheat can pair at metaphase I in interspecific hybrids between tetraploid wheat (carrying just the A and B genomes) and a diploid wild relative (Blanco et al 1988, Jauhar et al 1991).

The *Ph1* locus

The *Ph1* locus is specific to chromosome 5B and to a region defined by 70-Mb deletions in hexaploid wheat and tetraploid wheat (Sears 1977, Gill et al 1993). Chromosome 5B carrying the *Ph1* locus has meiotic drive compared with a 5B chromosome lacking the *Ph1* locus (Sears 1977). It is not found in either the A or D genomes, nor in the diploid A, D, or S genome progenitors. A similar type of *Ph1* activity is also found on chromosome 5G of *T. timopheevi* (AAGG) (Ozkan et al 2001). EMS treatment has failed to generate *Ph1* mutants, but large deletions generated by either X-ray or fast-neutron treatments have yielded mutants (Wall et al 1971, Sears 1977, Roberts et al 1999). The failure of treatments that yield point mutations to produce *Ph1* mutants suggests that the locus is complex in nature and arose following polyploidization of wheat. Thus, several related genes probably contribute to the phenotype, in which knocking out one or more would not produce an altered phenotype. The ability of *Ph1* to hinder chromosome pairing between homoeologs is dosage-dependent. In tetraploid and hexaploid wheat carrying chromosomes from two related rye species—*Secale montanum* (Rm) and *S. cereale* (Rc)—the homoeologs rye chromosomes pair freely at metaphase I with one *Ph1* copy, but few pairs are formed by the rye chromosomes in the presence of two *Ph1* copies (Riley and Miller 1970, Miller and Riley 1972). A single copy of *Ph1* is sufficient to prevent wheat-rye chromosome pairing at metaphase I in tetraploid and hexaploid wheat-rye hybrids (but pairing does occur in its absence). In contrast, six doses of *Ph1* prevent not only homoeologous pairing but also homologous pairing as the six doses result in unpaired homologous chromosomes at metaphase I (Feldman 1966, Yacobi et al 1982, Holm and Wang 1988). Thus, at zero doses of *Ph1*, homoeologous chromosomes regularly pair; with a single dose, some homoeologs can still pair; with two doses, homoeologous chromosomes can't pair; and, at six doses, some homologous chromosomes are unable to pair.

Does *Ph1* affect pairing?

The 42 chromosomes of hexaploid wheat pair as 21 bivalents as scored at metaphase I in the presence of *Ph1*. In the absence of *Ph1*, close to 21 bivalents are observed in most meiocytes at metaphase I (Roberts et al 1999). Thus, in hexaploid wheat, zero to two doses of *Ph1* do not affect substantially the overall level of pairing. If this is not affecting pairing, does it affect how chromosomes recognize each other?

Does *Ph1* affect homolog recognition?

So, what sites do chromosomes first use to recognize each other at the start of meiosis and does *Ph1* affect them? Studies of diploids—yeast (*Sacchizosaccharomyces pombe* and *Saccharomyces cerevisiae*), mammals, and rye—and polyploids—maize (a cryptic polyploid) and wheat—show that telomeres of chromosomes aggregate on the nuclear envelope, forming a telomere cluster or bouquet during meiotic prophase I (Dawe et al 1994, Bass et al 1997, 2000, Chikashige et al 1997, Niwa et al 2000, Trelles-Sticken et

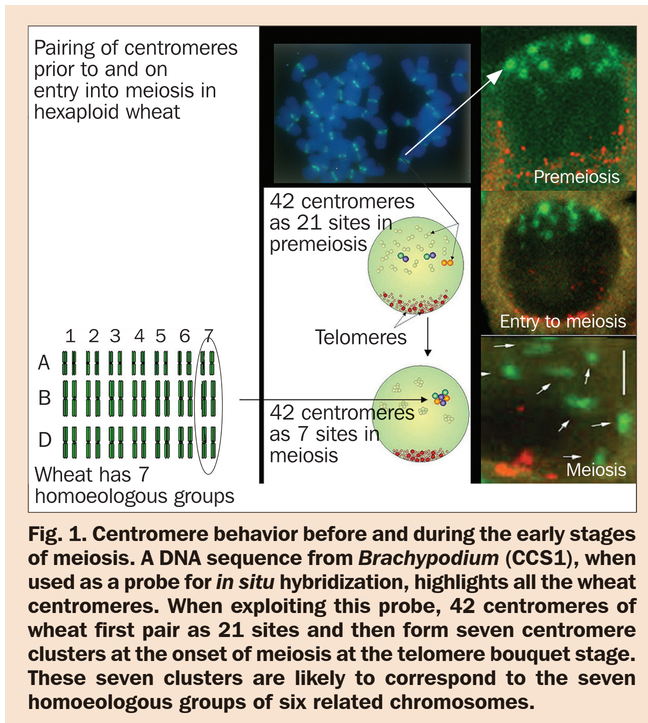


Fig. 1. Centromere behavior before and during the early stages of meiosis. A DNA sequence from *Brachypodium* (CCS1), when used as a probe for *in situ* hybridization, highlights all the wheat centromeres. When exploiting this probe, 42 centromeres of wheat first pair as 21 sites and then form seven centromere clusters at the onset of meiosis at the telomere bouquet stage. These seven clusters are likely to correspond to the seven homoeologous groups of six related chromosomes.

al 1999). These authors propose that this structure facilitates in some way the sorting of chromosomes into homologous pairs. Synaptonemal complex formation initiates near the telomeres in many plants (including maize and rye) (Gillies 1975, 1985). A deletion of the telomere region of one of the homologs is generally sufficient to reduce or eliminate subsequent pairing between these chromosomes at metaphase I, presumably because of the synaptonemal complex initiation (Curtis et al 1991, Lukaszewski 1997). Thus, homologous chromosomes are tethered at a region close to the telomeres during early prophase I and then are intimately aligned along their length as prophase I pairing proceeds. Recent data suggest that the telomere regions of homologs can pair with each other whether *Ph1* is present or absent. Thus, in hexaploid wheat, the telomere regions of homologs pair independently of *Ph1*. There are perturbations in the synteny in the telomere regions of homoeologous chromosomes (Gale et al 1997, Akhunov et al 2003). These perturbations may have led to the telomere regions of the homoeologs becoming distinct.

Does *Ph1* affect centromeres?

Are other chromosome sites being used as initial recognition sites? Centromeres in meiosis are involved in segregating the homologs to opposite poles at metaphase I, but are they involved in the process of homolog recognition? (Fig. 1). In early floral tissue,

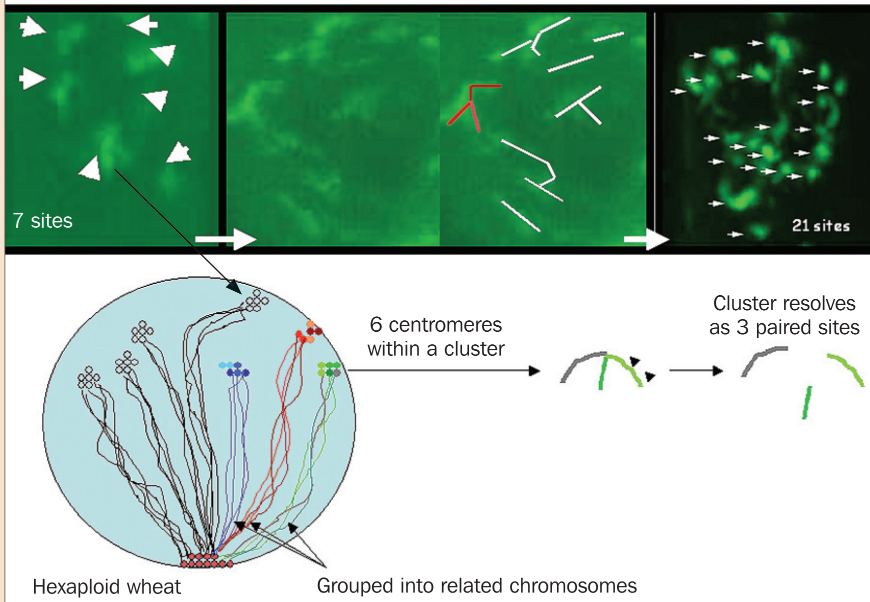


Fig. 2. Resolution of the seven centromere clusters to 21 paired sites in early meiosis. During early meiosis, the centromeres form seven clusters at the telomere bouquet stage. As meiosis progresses, these clusters resolve into tripartite structures, which further resolve back to 21 paired sites.

the centromeres associate as pairs in both hexaploid and tetraploid wheat (Aragon et al 1997, Martinez-Perez et al 2000, 2001). The centromeres do not pair premeiotically in the progenitor diploids (Martinez-Perez et al 2000). A third of these associations are homologous (Aragon et al 1997). At early meiosis, these pairs of centromeres form into seven groups corresponding to the seven homoeologous groups (Fig. 2). After centromere assortment within these groups, the centromeres resolve as homologous pairs (Martinez-Perez et al 2001, 2003). Thus, the centromeres do act as initial sites for homolog recognition. Although the presence of *Ph1* does affect centromere behavior, the centromeres of homologs still eventually pair whether *Ph1* is present or absent (Mikhailova et al 1998). Thus, again in hexaploid wheat, the centromere regions of homologs can still pair independently of *Ph1*. So, both telomeres and centromeres in hexaploid wheat do not require *Ph1* to pair correctly.

Effect of *Ph1* on interstitial sites

So, what is the problem in chromosome pairing following polyploidization? Is it the behavior of the rest of the chromosomes? In hexaploid wheat, most chromosomes intimately pair (synapse) with only one partner during meiotic prophase I. However, chromosomes occasionally start intimately pairing/synapsing with more than one partner (some 12 chromosomes out of 42 can be observed initially in multiple associations in both the presence and absence of *Ph1* (Holm 1986, 1988, Martinez et al 2001a,b). Thus, the specificity at which chromosomes initially synapse from the telomeres is unaffected by the presence or absence of two copies of *Ph1* but is increased in the presence of four copies, which reduce the number of chromosomes that initially engage in multiple associations from some 12 to less than 4. In the presence of *Ph1* (two copies), the incorrectly paired sites are resolved later during prophase I. By contrast, in the absence of *Ph1*, many incorrectly paired sites remain unresolved in later meiotic prophase I involving some 8–12 chromosomes in both hexaploid and tetraploid wheat.

The synapsis data suggest that something is happening to the rest of the chromosomes in the presence and absence of *Ph1* so that the pairing process is subtly altered. What is happening? This is difficult to visualize because the homoeologs possess a similar gene order. How can a pair of homologs be marked to follow the pairing process of interstitial sites on homologs in the presence and absence of *Ph1*? A wheat line carrying a rye segment covering 15% of the distal chromosome arm substituted for the equivalent region of the 1D pair of wheat chromosomes provides a solution. By visualizing the rye segments using genomic *in situ* hybridization, the pairing of the homologs bearing these segments can be visualized. Studies exploiting maize and *Caenorhabditis elegans* have indicated that the onset of pairing at meiosis is associated with conformational changes in the chromosomes (Dawe et al 1994, MacQueen and Villeneuve 2001). Recently, by visualizing rye segments, conformational change has also been observed in hexaploid wheat chromosomes that are pairing at meiosis and, interestingly, the behavior of the conformational change is affected by the *Ph1* locus (Fig. 3). In the absence of *Ph1*, the conformational changes occur asynchronously such that one homolog will undergo a change while the other homolog does not. Moreover, this asynchronous conformation change happens before bouquet formation (Fig. 3). Thus, following polyploidization, homologs can be in different conformations. In hexaploid wheat, *Ph1* ensures that comparable chromosome segments on the homologs undergo the conformational change simultaneously and that, therefore, their condensation state closely mirrors each other. Thus, in hexaploid wheat, *Ph1* is synchronizing the conformation changes so that the homologs are in the same condensation state. If homologs are in different conformational states, this will raise the prospect of them interacting with a homoeolog. This is borne out by the observations that the homologous sites are paired with each other 100% of the time in the presence of *Ph1*, but in its absence can be paired with each other as low as 33% of the time.

In the presence of *Ph1*, each of the two homologous sites undergoes identical conformation change at early meiosis and then pairs with the other. In the absence of

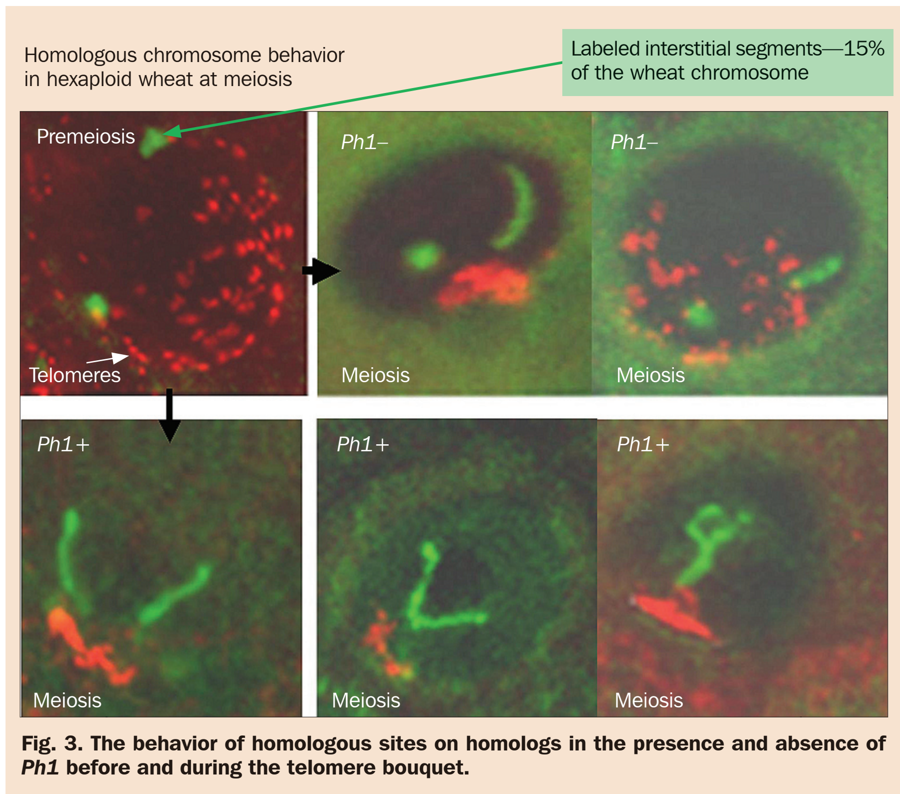


Fig. 3. The behavior of homologous sites on homologs in the presence and absence of *Ph1* before and during the telomere bouquet.

Ph1, in contrast, the homologous sites undergo asynchronous conformational changes, so that they are not in the same conformation when chromosomes are pairing at the telomere bouquet stage.

Effect of *Ph1* in wheat hybrids

In wheat hybrids, which have no homologs, all the related chromosomes are in different conformational states. In this situation, *Ph1* still prevents homoeologs from pairing. As stated above, B chromosomes (which are essentially lumps of heterochromatin) can compensate for the absence of the *Ph1* locus. This suggests that heterochromatin is being affected. In hexaploid wheat-rye hybrids in the absence of *Ph1*, the rye heterochromatin knobs undergo conformational changes and all pair with each other, whereas, in the presence of *Ph1*, the rye heterochromatin knob regions do not undergo conformational changes. Moreover, *Ph1* also affects heterochromatin behavior at the centromeres in these hybrids. In hexaploid wheat-rye hybrids, their 28 centromeres form 14 centromere sites before meiosis and 7 groups at meiosis. However, the 14 centromere sites in the presence of *Ph1* are composed of 7 wheat-rye pairs and 7

wheat-wheat pairs, whereas, in the absence of *Ph1*, there are 7 sites composed of 21 wheat centromeres and 7 single-rye centromeres (Prieto et al 2004b). In the presence of *Ph1*, these sites resolve as unpaired, whereas, in the absence of *Ph1*, they resolve as 14 paired sites (Martinez-Perez et al 2001).

Effect of *Ph1* on recombination

Studies by Dvorak and colleagues found that, in the absence of *Ph1*, recombination occurs between a pair of wheat chromosomes composed of combinations of homoeologous and homologous segments, but, in the presence of *Ph1*, recombination is restricted to homologous segments (Dubcovsky et al 1995, Luo et al 1996, 2000). The fact that homologous segments will be in the same conformation in the presence of *Ph1* means that the homoeologous segments embedded in the homologs will be in different conformations. In the absence of *Ph1*, homologous as well as homoeologous segments will all be in different conformations.

Ph1 induces multiple paired sites

The effect on the telomeric heterochromatin knobs could help to explain the basis for homoeologous pairing in the absence of *Ph1*. In the presence of *Ph1*, regions that are highly homologous (such as heterochromatin) do not undergo conformational change and can therefore be excluded from the pairing process. In contrast, in the absence of *Ph1*, chromatin changes do occur in these highly homologous regions and these regions can engage in multiple associations among homologous, homoeologous, and nonhomologous chromosomes. Moreover, the whole chromosome may be slightly more condensed in the presence of *Ph1* when pairing than in its absence, which might further exclude highly homologous repeats from the pairing process. It has also been observed in the absence of *Ph1* that the elongation of chromatin associated with pairing can occur in meiocytes that have not fully formed the telomere bouquet (Prieto et al 2004a, 2005). In contrast, the conformational changes associated with pairing are observed only at telomere bouquet formation in the presence of *Ph1* (Prieto et al 2004a, 2005). An earlier initiation of pairing in the absence of *Ph1* implies that the chromosomes will pair in slightly different overall condensation states than in the presence of *Ph1*. Thus, in the absence of *Ph1*, the chromosomes will be less condensed than in the presence of *Ph1*. This is consistent with the proposal by Maestra et al (2002) that, although there is no apparent difference in the overall structure of chromosomes in the presence and absence of *Ph1* before meiosis, the chromosomes may be less condensed when pairing during early meiosis in the absence of *Ph1* than in its presence. The more "open chromatin" (less condensed) of the entire chromosome at the time of pairing combined with the ability of highly homologous heterochromatin to extensively elongate as the chromosomes pair may explain the basis of a marked increase in homoeologous and nonhomologous interactions in the absence of *Ph1*. However, it is difficult to provide clear-cut data for this proposal at the telomere bouquet as visualizing the behavior of whole-chromosome additions

is difficult to interpret at this stage, whereas reducing the complexity by visualizing single-arm additions (telosomes) has additional complications. The two telomeres of the telosome join the telomere bouquet, bringing the centromere of the telosome into the bouquet. Thus, the telosome is looped back at this stage and in some cases can stretch right around the nucleus (Martinez-Perez et al 1999, Maestra et al 2002, Carlton and Cande 2002).

The nature of the *Ph1* locus

What is the *Ph1* locus? No allelic variation has been reported for it. The only variation reported has been associated with dosage. Increasing *Ph1* dosage reduces both homologous and homoeologous pairing, whereas, at a single dose, homoeologous rye chromosomes will pair and recombine, which they cannot do at two doses. Thus, *Ph1* can suppress all pairing but different doses determine whether homologous or homoeologous pairing is affected. B chromosomes can compensate for the absence of *Ph1*, which, assuming that B chromosomes are just heterochromatin and lacking genes, means that the mechanisms that maintain and regulate heterochromatin also regulate pairing and recognition as *Ph1* does. EMS mutagenesis has failed to generate any *Ph1* mutants, whereas X-ray and fast-neutron treatments have generated mutants. This suggests that *Ph1* is not simply a "gene" that can be knocked out by point mutation, but is more complex. This would also explain the lack of variation in the *Ph1* phenotype in hexaploid wheat as allelic variation in any one member would not have much effect.

What is next?

What can we expect in the future? The characterization of the region containing the *Ph1* locus should be completed and revealed in the coming months. From this, it should become clear how the *Ph1* locus evolved on polyploidization. However, more importantly, this information can be used to assess whether it is possible to generate allelic variation at the locus. This variation can then be assessed for any difference in phenotype. The factors that suppress the *Ph1* effect are found in *Aegilops speltoides* and *Ae. mutica*. These factors should be elucidated in the future from programs being developed by Dvorak and colleagues for *Ae. speltoides* and by ours for *Ae. mutica*. Thus, it should become possible to promote more easily pairing between homoeologs either by suppressing the *Ph1* effect or by exploiting allelic variation in the *Ph1* locus.

References

Akhunov ED, Akhunova AR, Linkiewicz AM, Dubcovsky J, Hummel D, Lazo G, Chao S, Anderson O, David J, Qi L, et al. 2003. Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. Proc. Natl. Acad. Sci. USA 100:10836-10841.

Aragon-Alcaide L, Reader S, Bevan A, Shaw P, Miller T, Moore G. 1997. Association of homologous chromosomes during floral development. Curr. Biol. 7:905-908.
Bass HW, Marshall WF, Sedat JW, Agard DA, Cande WZ. 1997. Telomeres cluster de novo before the initiation of synapsis: a three-dimensional spatial analysis of telomere positions before and during meiotic prophase. J. Cell Biol. 137:5-18.
Bass HW, Bordoll SJ, Rines HW, Phillips RL, Sedat JW, Agard DA, Cande WZ. 2000. Evidence for the coincident initiation of homolog pairing and synapsis during telomere-clustering (bouquet) stage of meiotic prophase. J. Cell Sci. 113:1033-1042.
Blanco A, Perrone V, Simeone R. 1988. Chromosome pairing variation in *Triticum turgidum* × *Dasypyrum villosum* (L.) candargy hybrids and genome affinities. Proceedings of the 7th International Wheat Genetics Symposium. p 63-68.
Carlton PM, Cande WZ. 2002. Telomeres act autonomously in maize to organise the meiotic bouquet from a semipolarised chromosome orientation. J. Cell Biol. 157:231-242.
Chikashige Y, Ding DQ, Imai Y, Yamamoto M, Haraguchi I, Hiraoka Y. 1997. Meiotic nuclear reorganisation: switching the position of centromeres and telomeres in the fission yeast *Schizosaccharomyces pombe*. EMBO J. 16:193-202.
Curtis CA, Lukaszewski AJ, Chrzastek M. 1991. Metaphase I pairing of deficient chromosomes and genetic mapping of deficiency breakpoints in common wheat. Genome 34:553-560.
Dawe RK, Sedat JW, Agard DA, Cande WZ. 1994. Meiotic chromosome pairing in maize is associated with a novel chromatin organisation. Cell 76:901-912.
Dover GA, Riley R. 1972. Prevention of pairing of homoeologous meiotic chromosomes of wheat by the activity of supernumerary chromosomes of *Aegilops*. Nature 240:159-161.
Dubcovsky J, Luo MC, Dvorak J. 1995. Differentiation between homoeologous chromosome 1A of wheat and 1Am of *Triticum monococcum* and its recogntion by the wheat *Ph1* locus. Proc. Nat. Acad. Sci. USA 92:6645-6649.
Feldman M. 1993. Mode of action of *Ph1* in wheat. Crop Sci. 33:894-897.
Feldman M. 1966. The effect of chromosomes 5A, 5B and 5D on chromosomal pairing in *Triticum aestivum*. Proc. Natl. Acad. Sci. USA 55:1447-1453.
Gale MD, Atkinson CN, Chinoy RL, Harcourt J, Jia J, Li QY, Devos KM. 1997. Genetic maps of hexaploid wheat. Proceedings of the 8th International Wheat Genetics Symposium. p 29-40.
Gill BS, Friebe B, Endo TR. 1991. Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). Genome 34:830-839.
Gill KS, Gill BS, Endo TR, Mukai Y. 1993. Fine physical mapping of *Ph1*, a chromosome pairing regulator gene in polyploid wheat. Genetics 134:1231-1235.
Gillies CB. 1975. An ultrastructural analysis of chromosome pairing in maize. CR Trav. Lab. Carlsberg 40:135-161.
Gillies CB. 1985. An electron microscopic study of synaptonemal complex formation at zygotene in rye. Chromosoma 92:165-175.
Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G. 2006. Molecular characterization of *Ph1* as a major chromosome pairing locus in polypoid wheat. Nature 439:749-752.
Holm PB. 1986. Chromosome pairing and chiasma formation in allohexaploid wheat, *Triticum aestivum* analysed by spreading of meiotic nuclei. Carlsberg Res. Commun. 51:239-294.

Holm PB. 1988. Chromosome pairing and synaptonemal complex formation in hexaploid wheat nullisomic for chromosome 5B. Carlsberg Res. Commun. 53:91-110.
Holm PB, Wang X. 1988. The effect of chromosome 5B on synapsis and chiasma formation in wheat, *Triticum aestivum*, cv. Chinese Spring. Carlsberg Res. Commun. 53:191-208.
Jauhar PP, Rieraplizarazu O, Dewey WG, Gill BS, Crane CF, Bennett JH. 1991. Chromosome pairing relationships among the A, B and D genomes of bread wheat. Theor. Appl. Genet. 82:441-449.
Lukaszewski A. 1997. The development and meiotic behaviour of asymmetrical isochromosomes in wheat. Genetics 145:1155-1160.
Luo MC, Dubcovsky J, Dvorak J. 1996. Recognition of homoeology by wheat *Ph1* locus. Genetics 143:1195-1203.
Luo MC, Yang ZL, Kota RS, Dvorak J. 2000. Recombination of chromosomes 3A(m) and 5A(m) of *Triticum monococcum* with homoeologous chromosomes 3A and 5A of wheat: the distribution of recombination across chromosomes. Genetics 154:1301-1308.
Kimber G, Alsonso LS, Sallee OJ. 1981. The analysis of meiosis in hybrids: aneuploid hybrids. Can. J. Genet. Cytol. 23:209-219.
MacQueen AJ, Villeneuve AM. 2001. Nuclear reorganisation and homologue chromosome pairing during meiotic prophase require *C. elegans* Chk-2. Genes Dev. 15:1674-1687.
Maestra B, Hans de Jong J, Shepherd K, Naranjo T. 2002. Chromosome arrangement and behaviour of the rye homologous telosomes at the onset of meiosis in disomic wheat 5RL addition lines with and without the *Ph1* locus. Chromosome Res. 10:655-667.
Martinez M, Cunado N, Carcelen N, Romero C. 2001a. The synaptic behaviour of *Triticum turgidum* with variable doses of the *Ph1* locus. Theor. Appl. Genet. 102:7.
Martinez M, Naranjo T, Cuadrado C, Romero C. 2001b. The *Ph1* and *Ph2* loci play different roles in the synaptic behaviour of hexaploid wheat *Triticum aestivum*. Theor. Appl. Genet. 103:398-405.
Martinez-Perez E, Shaw P, Reader S, Aragon-Alcaide L, Miller T, Moore G. 1999. Homologous chromosome pairing in wheat. J. Cell Sci. 112:1761-1769.
Martinez-Perez E, Shaw P, Moore G. 2000. Polyploidy induces centromere association. J. Cell Biol. 148:233-238.
Martinez-Perez E, Shaw P, Moore G. 2001. The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association. Nature 411:204-208.
Martinez-Perez E, Shaw P, Aragon-Alcaide L, Moore G. 2003. Chromosomes form into seven groups in hexaploid and tetraploid wheat as prelude to meiosis. Plant J. 36:21-29.
Mikhailova E, Naranjo T, Shepherd K, Wennekes-van Eden J, Heyting C, de Jong J, et al. 1998. The effect of wheat *Ph1* locus on chromatin organisation and meiotic chromosome pairing analysed by genome painting. Chromosoma 197:339-350.
Miller TE, Riley R. 1972. Meiotic chromosome pairing in wheat. Genet. Iber. 24:241-250.
Mochida K, Yamazaki Y, Ogihara Y. 2004. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol. Gen. Genet. 270:371-377.
Niwa O, Shimangel M, Miki F. 2000. Telomere-led bouquet formation facilitates homologous chromosome pairing and restricts ectopic interaction in fission yeast meiosis. EMBO J. 19:3831-3840.
Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidy-induced rapid genome evolution in wheat (*Aegilops-Triticu*) group. Plant Cell 13:1735-1747.

Prieto, P, Santos, AP, Moore G, Shaw P. 2004a. Chromosomes associate premeiotically and in xylem vessel cells via their telomeres and centromeres in diploid rice (*Oryza sativa*). Chromosoma 112:300-307.
Prieto P, Shaw P, Moore G. 2004b. Homologue recognition during meiosis is associated with a change in chromatin conformation. Nature Cell. Biol. 6:906-908.
Prieto P, Moore G, Reader S. 2005. Control of conformation changes associated with homologue recognition during meiosis. Theor. Appl. Genet. 111:505-510.
Riley R, Chapman V. 1958. Genetic control of the cytological diploid behaviour of hexaploid wheat. Nature 182: 712-715.
Riley R, Miller T. 1970. Meiotic chromosome pairing in Triticale. Nature 227:82-83.
Roberts MA, Reader SM, Dalgliesh C, Miller TE, Foote TN, Fish LJ, Snape JW, Moore G. 1999. Induction and characterisation of *Ph1* wheat mutants. Genetics 153:1909-1918.
Sanchez-Moran E, Benavente E, Orellana J. 2001. Analysis of karyotypic stability of homoeologous-pairing (ph) mutants in allopolyploid wheats. Chromosoma 110:371-377.
Sears ER, Okamoto M. 1958. Intergenomic chromosome relationships in hexaploid wheat. Proceedings of the 10th International Congress of Genetics. Montreal 2:258-259.
Sears E. 1972. *Agropyron*-wheat transfers through induced homoeologous pairing. Can. J. Genet. Cytol. 14:736.
Sears ER. 1977. An induced mutant with homoeologous pairing in common wheat. Can. J. Genet. Cytol. 19:585-593.
Trelles-Sticken E, Loidl J, Scherthan H. 1999. Bouquet formation in budding yeast: initiation of recombination is not required for meiotic telomere clustering. J. Cell Sci. 112:651-658.
Wall AM, Riley R, Chapman V. 1971. The position of a locus on chromosome 5B of *Triticum aestivum* affecting homoeologous meiotic pairing. Genet. Res. 18:329-339.
Yacobi Y, Mello-Sampayo T, Feldman M. 1982. Genetic induction of bivalent interlocking in common wheat. Chromosoma 87:165-175.
Zickler D, Kleckner N. 1999. Meiotic chromosomes: integrating structure and function. Annu. Rev. Genet. 33:603-754.

Notes

Author's address: Crop Genetics, John Innes Centre, Colney, Norwich. E-mail: graham.moore@bbsrc.ac.uk.

Functional genomics and rice improvement

Functional genomics for gene discovery in abiotic stress response and tolerance

K. Shinozaki and K. Yamaguchi-Shinozaki

Plants respond to abiotic stresses, such as drought, high salinity, and cold, to acquire stress tolerance. Molecular and genomic studies have shown that a number of genes with various functions are induced by abiotic stresses, and that various transcription factors are involved in the regulation of stress-inducible genes in *Arabidopsis* and rice. These gene products function not only in stress tolerance but also in stress response. In this review, recent progress in the analysis of complex cascades of gene expression in drought and cold stress responses is summarized. Various genes involved in stress tolerance are also discussed for their application to molecular breeding of drought, salinity, and/or cold stress tolerance.

Keywords: Drought, cold stress, abscisic acid, gene expression, stress tolerance, signal transduction, microarray analysis

Plants respond and adapt to abiotic environmental stresses, including drought, cold, heat, and high salinity, to survive in severe stress conditions. These stresses induce various physiological and biochemical responses in plants. These responses include stomata closure, repression of cell growth and photosynthesis, and accumulation of various osmolytes and proteins involved in cellular protection from these stresses. Moreover, a variety of genes are induced or repressed by these abiotic stresses at the transcriptional level (Shinozaki et al 2003, Yamaguchi-Shinozaki and Shinozaki 1994, 2005, Zhu 2002, Bartels and Sunkar 2005). Their gene products really function in cellular stress tolerance and response. Many stress-inducible genes have been used to improve the stress tolerance of plants by gene transfer (Zhang et al 2004). It is important to analyze the functions of stress-inducible genes not only to further understand the molecular mechanisms of stress tolerance and response of higher plants but also to improve the stress tolerance of crops by gene manipulation.

Dehydration and high salinity trigger the production of abscisic acid (ABA), which, in turn, not only causes stomata closure but also induces various genes. Many genes are induced by exogenous ABA treatment, whereas some genes are not induced by ABA. There are both ABA-independent regulatory systems and ABA-dependent

ones in stress-responsive gene expression (Yamaguchi-Shinozaki and Shinozaki 2005). *Cis*-acting elements and transcription factors that function in ABA-independent and ABA-responsive gene expression under drought stress have been precisely analyzed at the molecular level (Shinozaki et al 2003, Yamaguchi-Shinozaki and Shinozaki 2005). On the other hand, the ABA-independent regulatory system is important in cold-inducible gene expression. This review article summarizes recent progress in transcriptome analysis of stress-responsive gene expression using microarrays in *Arabidopsis* and rice, and the functions of stress-inducible genes in stress responses and tolerance. *Cis*- and *trans*-acting factors involved in ABA-independent and ABA-dependent gene expression systems are also described. Several stress-inducible genes are discussed that are useful for the improvement of abiotic stress tolerance by gene transfer.

Transcriptome analysis of abiotic stress-inducible gene expression in *Arabidopsis* using microarrays

Microarray technology using cDNAs or oligonucleotides is a powerful and useful tool for analyzing the gene expression profiles of plants that are exposed to abiotic stresses, such as drought, high salinity, and cold, or to ABA treatment (Seki et al 2001, 2002a,b, Kreps et al 2002). This technology is also useful for identifying target genes for stress-related transcription factors, and analyzing gene networks in abiotic stress responses. There are two major forms of microarray technology, the cDNA microarray (Seki et al 2001, 2002a,b) and the oligonucleotide-based microarray, mainly Affymetrix GeneChip (Kreps et al 2002). The cDNA microarray has been used to identify genes induced by drought, cold, and high salinity. Recently, 299 drought-inducible genes, 54 cold-inducible genes, 213 high-salinity-stress-inducible genes, and 245 ABA-inducible genes in *Arabidopsis* were identified using a cDNA microarray containing around 7,000 independent *Arabidopsis* full-length cDNAs (Seki et al 2002a,b). Thousands of stress-inducible genes were identified by the oligonucleotide-based microarray using a GeneChip array containing oligonucleotides for about 8,000 independent *Arabidopsis* genes (Kreps et al 2002). The stress-inducible genes obtained by the cDNA microarray were not the same as those obtained by GeneChip analyses. This is mainly due to the difference in sets of genes arrayed in the two systems (only 1,919 genes were arrayed in both systems) and different conditions of plant growth and stress treatments (Maruyama et al 2004).

More than half of the drought-inducible genes are also induced by high salinity and/or ABA treatments, indicating the existence of significant crosstalk among the drought, high-salinity, and ABA responses. In contrast, only 10% of the drought-inducible genes could also be induced by cold stress. The products of the identified genes can be classified into two groups (Shinozaki et al 2003, Fig. 1). The first group includes proteins that probably function in stress tolerance, such as chaperones, LEA proteins, osmotin, antifreeze proteins, mRNA binding proteins, key enzymes for osmolyte biosynthesis, water channel proteins, sugar and proline transporters, detoxification enzymes, and various proteases. The second group contains regulatory proteins,

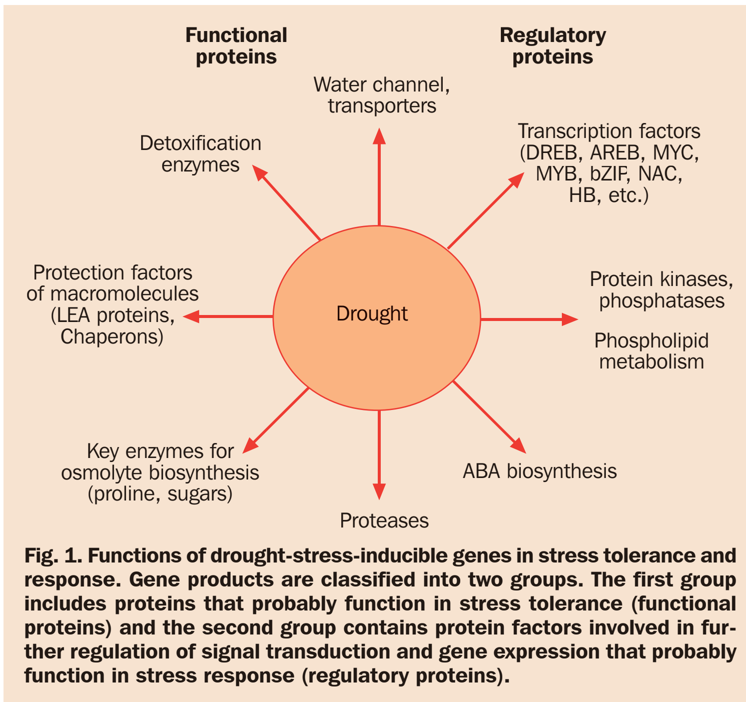


Fig. 1. Functions of drought-stress-inducible genes in stress tolerance and response. Gene products are classified into two groups. The first group includes proteins that probably function in stress tolerance (functional proteins) and the second group contains protein factors involved in further regulation of signal transduction and gene expression that probably function in stress response (regulatory proteins).

that is, protein factors involved in further regulation of signal transduction and gene expression that probably function in stress response; they are various transcription factors, protein kinases, protein phosphatases, enzymes involved in phospholipid metabolism, and other signaling molecules such as calmodulin-binding protein. Many transcription-factor genes were found among the stress-inducible genes, suggesting that various transcriptional regulatory mechanisms function in the drought-, cold-, or high-salinity-stress signal transduction pathways. These transcription factors could regulate various stress-inducible genes cooperatively or separately, and may constitute gene networks in *Arabidopsis*.

Now, whole-genome arrays with more than 20k genes have been produced by companies and used for gene expression profiling. Among them, the Affymetrix 24k ATH1 GeneChip is used worldwide, and the Agilent 22k Arabidopsis 2 oligoarray is also used widely. Recently, the AtGenExpress project on *Arabidopsis* transcriptome using the ATH1 GeneChip collected thousands of transcriptome data of expressed genes in various tissues, growth conditions, stress conditions, and phytohormone treatments (Schmid et al 2005). Moreover, a whole-genome tiling array that covers both strands of whole-*Arabidopsis* genomic DNA with 25 nucleotides of every 30 nulceotides has been developed to analyze all the transcripts, including protein-coding and noncoding genes, miRNAs, antisense RNAs, and other novel RNAs with unknown functions.

Transcriptome analysis of abiotic stress-inducible gene expression in rice

Rice gene expression in response to high-salinity stress was first analyzed using microarray by Kawasaki et al (2001), who used rice cDNA (ESTs) microarray for the analysis. Recently, we analyzed rice stress-inducible genes using a rice cDNA microarray including about 1,700 independent cDNAs (ESTs). These cDNAs were collected from rice plants exposed to drought, cold, or high-salinity stresses (Rabbani et al 2003). Stress-inducible expression of the candidate genes selected by microarray analysis was confirmed by using RNA gel-blot analysis and 73 genes were identified as stress-inducible. Among them, 36, 62, 57, and 43 genes were induced by cold, drought, high salinity, and ABA, respectively. Around 40% of the drought- or high-salinity-inducible genes were induced by cold stress. In contrast, more than 98% of the high-salinity-inducible and 100% of the ABA-inducible genes were also induced by drought stress. This indicated the existence of a substantial common regulatory system or a greater crosstalk between drought and high-salinity stress and between drought and the ABA signaling process than between cold stress and drought stress or between cold and the ABA signaling process (Rabbani et al 2003). These results in rice are consistent with the overlap of gene expression in response to drought and high salinity in *Arabidopsis*.

The products of the identified rice genes are classified into functional proteins and regulatory proteins like those of *Arabidopsis* (Rabbani et al 2003, Fig. 1). Comparative analysis of stress-inducible genes in *Arabidopsis* and rice revealed a considerable level of similarity in stress responses between the two genomes at the molecular level. Among 73 identified stress-inducible genes in rice, 51 have already been reported in *Arabidopsis* with a similar function or gene name. These results indicate that rice has many stress-inducible genes in common with *Arabidopsis*, even though these two plants evolved separately a million years ago. Common stress-inducible genes include LEA proteins, antifreeze proteins, sugar transporter, protease, detoxification enzymes, transcription factors, protein kinases, protein phosphatases, and enzymes involved in phospholipid metabolism. All these genes are up-regulated in response to at least one of the abiotic stresses in rice and reported as stress-inducible genes in *Arabidopsis*. Transcriptome analysis of rice also revealed novel stress-inducible genes, suggesting some differences between *Arabidopsis* and rice in their response to stress (Rabbani et al 2003).

Recently, a larger oligonucleotide microarray has been produced based on full-length cDNA information obtained by the Rice Genome Program of Japan. The rice oligonucleotide array is available from the Agilent Company Ltd. We are now using the Agilent oligoarray for rice transcriptome analysis in abiotic stress responses.

Regulation of gene expression in abiotic stress responses: *cis*-acting elements and transcription factors in ABA-independent gene expression

The promoter of a drought-, high-salinity-, and cold-inducible gene, *RD29A/COR78/LTI78,* has been precisely analyzed to contain two major *cis*-acting elements, ABRE

(ABA responsive element) and DRE (dehydration responsive element)/CRT (C-RepeaT), involved in stress-inducible gene expression (Yamaguchi-Shinozaki and Shinozaki 2005). ABRE and DRE/CRT are *cis*-acting elements that function in ABA-dependent and ABA-independent gene expression in response to abiotic stress, respectively (Fig. 2). Transcription factors belonging to the ERF/AP2 family that bind to DRE/CRT were isolated and termed *CBF/DREB1* and *DREB2* (Yamaguchi-Shinozaki and Shinozaki 2005). The conserved DNA-binding motif of *DREB1A/CBF3* and *DREB2* is A/GCCGAC. The *CBF/DREB1* genes are quickly and transiently induced by cold stress, whose products activate the expression of target stress-inducible genes (Jaglo-Ottosen et al 1998, Liu et al 1998, Kasuga et al 1999). The *DREB2* genes are induced by dehydration stress to express various genes involved in drought-stress tolerance (Liu et al 1998). Overexpression of *CBF/DREB1* in transgenics increased stress tolerance of freezing, drought, and salt, suggesting that the CBF/DREB1 proteins function in the development of cold-stress tolerance without modification (Liu et al 1998). Many *CBF/DREB1* target genes have been identified using both cDNA and GeneChip microarrays (Seki et al 2001, Fowler and Thomashow 2002, Maruyama et al 2004, Vogel et al 2005). Most of the *CBF/DREB1* target genes contain the DRE motif with (A/G)CCGACNT sequence in their promoter regions.

In contrast, overexpression of DREB2 in transgenics does not improve stress tolerance, which suggests the involvement of posttranslational activation of DREB2 protein (Liu et al 1998, Sakuma et al, unpublished data). The DREB2 protein is expressed under normal growth conditions and activated by osmotic stress through posttranslational modification in the early stage of osmotic stress response, whereas dehydration-inducible CBF/DREB1 may function in the next stage of stress response after the induction of the protein. Rice homologs for CBF/DREB1 and DREB2, 10 OsDREB1s and 4 OsDREB2s, respectively, have been identified based on rice genome sequence analyses, which function in stress-inducible gene expression in rice. Overexpression of OsDREB1A in *Arabidopsis* revealed a similar function of the rice genes in stress-responsive gene expression and stress tolerance (Dubouzet et al 2003). Recently, overexpression of OsDREB1 or *Arabidopsis* DREB1 also improved drought and chilling tolerance in rice (Ito et al, unpublished data). This indicates that similar transcription factors function between dicotyledons and monocotyledons.

Several drought-inducible genes do not respond to either cold or ABA treatment, which suggests the existence of another ABA-independent pathway in the dehydration stress response. These genes include *ERD1* that encodes a Clp protease regulatory subunit, ClpD. The *ERD1* gene is not only induced by dehydration but is also up-regulated during natural senescence and dark-induced senescence (Nakashima et al 1997). Promoter analysis of the *ERD1* gene in transgenic plants indicates that the *ERD1* promoter contains *cis*-acting elements involved in both ABA-independent stress-responsive gene expression and senescence-activated gene expression. Analysis of the *ERD1* promoter identified two different novel *cis*-acting elements involved in induction by dehydration stress and dark-induced senescence (Simpson et al 2003). Recently, DNA-binding proteins for the *cis*-elements were identified to be NAC transcription factors (Tran et al 2004).

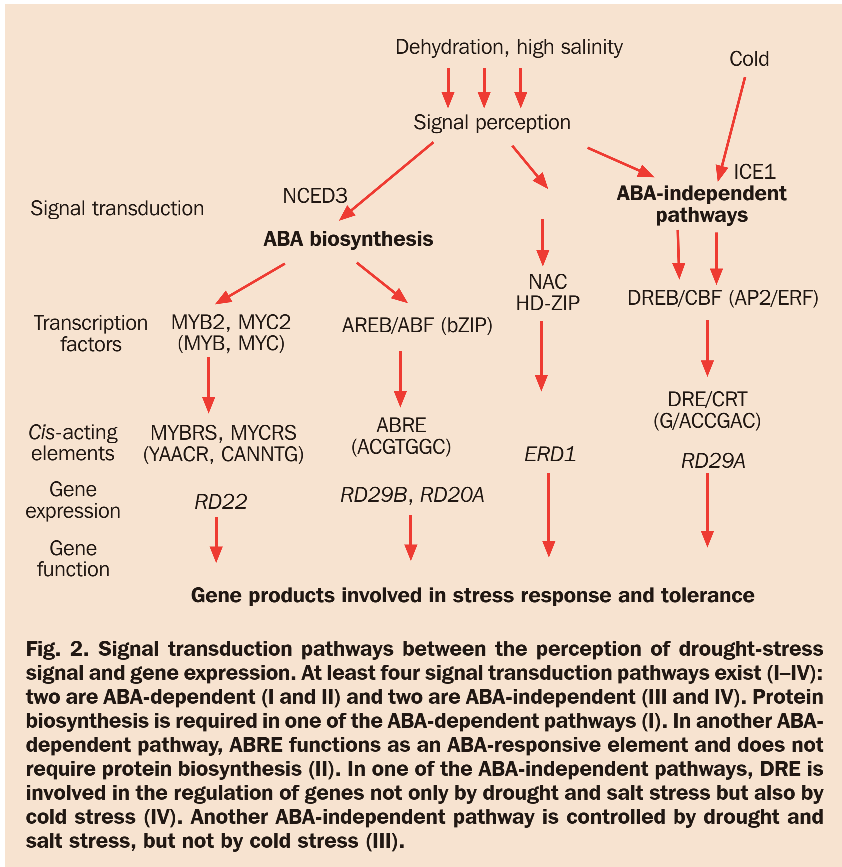


Fig. 2. Signal transduction pathways between the perception of drought-stress signal and gene expression. At least four signal transduction pathways exist (I–IV): two are ABA-dependent (I and II) and two are ABA-independent (III and IV). Protein biosynthesis is required in one of the ABA-dependent pathways (I). In another ABA-dependent pathway, ABRE functions as an ABA-responsive element and does not require protein biosynthesis (II). In one of the ABA-independent pathways, DRE is involved in the regulation of genes not only by drought and salt stress but also by cold stress (IV). Another ABA-independent pathway is controlled by drought and salt stress, but not by cold stress (III).

Cis-acting elements and transcription factors involved in ABA-dependent gene expression under water stress

ABA is synthesized *de novo* in response to mainly drought and high-salinity stress but not to cold stress. Recently, genes involved in ABA biosynthesis and catabolism have been identified based on genetic and genomics analysis (Nambara and Marion-Poll 2005). Several genes involved in ABA biosynthesis are induced by drought and high salinity but not by cold stress. This indicates the important roles of ABA in drought-stress responses. Among the genes involved in ABA biosynthesis, genes for 9-*cis*-epoxycarotenoid dioxygenase, a key enzyme for ABA biosynthesis, are strongly induced by drought stress (Iuchi et al 2001).

ABRE is a major *cis*-acting element in ABA-responsive gene expression (Fig. 2). Two ABRE motifs are important *cis*-acting elements in the ABA-responsive expression of the *Arabidopsis RD29B* gene (Uno et al 2000). Basic leucine zipper (bZIP) transcription factors, AREB/ABF, can bind to ABRE and activate ABA-dependent

gene expression (Uno et al 2000, Choi et al 2000). The AREB/ABF proteins need an ABA-mediated signal for their activation because of their reduced activity in the ABA-deficient *aba2* and ABA-insensitive *abi1* mutants and their enhanced activity in the ABA-hypersensitive *era1* mutant (Uno et al 2000). This is probably due to the ABA-dependent phosphorylation of the AREB/ABF proteins. Overexpression of *ABF3* or *AREB2/ABF4* caused ABA hypersensitivity, reduced the transpiration rate, and enhanced drought tolerance of the transgenics (Kang et al 2002).

Induction of the drought-inducible *RD22* gene is mediated by ABA and requires protein biosynthesis for its ABA-dependent expression. An MYC transcription factor, *RD22BP1 (AtMYC2)*, and an MYB transcription factor, *AtMYB2*, were shown to bind *cis*-elements in the RD22 promoter and cooperatively activate RD22 (Abe et al 1997, 2003, Fig. 2). These MYC and MYB proteins are synthesized after the accumulation of endogenous ABA, which indicates their role in the rather late stage of stress responses. Microarray analysis revealed target genes of MYC/MYB overexpressing in transgenics, such as alcohol dehydrogenase and ABA- or jasmonic acid (JA)-inducible genes (Abe et al 2003). Overexpression of both AtMYC2 and AtMYB2 not only revealed an ABA hypersensitive phenotype but also improved osmotic-stress tolerance of the transgenic plants.

In addition to the above transcription factors that bind to *cis*-acting elements of stress-inducible genes, many genes for various types of transcription factors are induced by drought and/or cold stress. Recently, we identified the drought-inducible *RD26* gene encoding a NAC transcription factor (Fujita et al 2004). The RD26 NAC transcription factor gene is induced by drought, high salinity, ABA, and jasmonic acid (JA) treatment. RD26 protein is localized in the nucleus and has transcriptional activity. RD26 overexpressor was hypersensitive to ABA, and RD26 dominant repressor was insensitive to ABA. ABA- and stress-inducible genes were up-regulated in the RD26 overexpressor and repressed in the RD26 repressor. This indicates an important role of the *RD26* gene in ABA signaling during drought stress. The RD26 is also involved in defense- and senescence-related gene expression in stress response.

Discovery of genes for improvement of abiotic stress tolerance in transgenics

Now, many stress-inducible genes have been identified to function in abiotic stress tolerance using transgenic plants (Bartels and Sunkar 2005). These genes are key enzymes of biosynthesis of osmolytes, such as proline, trehalose, maltose, and so on. Genes for LEA proteins and heat shock proteins are also used to improve drought tolerance in transgenics. We used a gene for galactinol synthase (GolS), a key enzyme involved in raffinose family oligosaccharide biosynthesis, for improvement of drought stress tolerance in transgenics (Taji et al 2002).

Transcription factors have been shown to be quite useful for improving stress tolerance by controlling a number of target genes in transgenic overexpressors (Shinozaki et al 2003). Other regulatory factors, such as protein kinases and enzymes in ABA biosynthesis, are also useful for improving stress tolerance by controlling many stress-related genes. We showed that overexpression of a gene for 9-*cis*-epoxycarot-

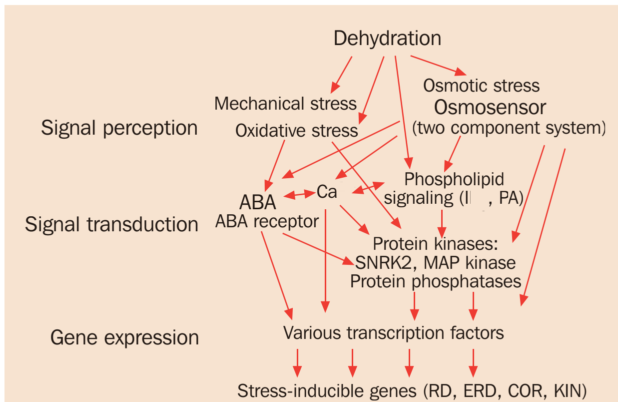


Fig. 3. Second messengers and protein factors involved in the signal perception and signal transduction in drought-stress response. Two-component histidine kinase is thought to function as an osmosensor in plants. Calcium and phospholipids are most probably cellular second messengers of drought-stress signal. The phosphorylation process functions in water-stress and ABA signal-transduction pathways. SnRK2s play important roles in dehydration and ABA signaling. MAP kinases also function in stress signaling. ABA plays important roles in the regulation of both gene expression and physiological responses during water stress.

enoid dioxygenase (NCED), a key enzyme in ABA biosynthesis, improves drought stress tolerance in transgenics (Iuchi et al 2001).

Recently, we showed that ABA-activated SnRK2 protein kinase (OST1/SRK2E) functions in an ABA signal transduction pathway in stomata closure (Yoshida et al 2002, Mustilli et al 2002). SnRK2 is a member of the SNF1-related PKase family, and contains 10 members in *Arabidopsis*. SnRK2s are activated by drought, salinity, and ABA (Yoshida et al 2002). SRK2E/OST1 is involved in stomata closure but not in seed germination. Another SnRK2, SRK2C, is activated by osmotic stress, salt stress, and ABA (Umezawa et al 2004). SnRK2C is strongly expressed in the root chip and root axis, and is involved in root response to drought. SnRK2C is involved in stress-responsive gene expression to improve stress tolerance. SnRK2 protein kinases may be involved in the activation of transcription factors in osmotic-stress-responsive gene expression. SRK2C is a useful tool for improving stress tolerance by controlling many stress-related genes (Fig. 3).

Conclusions and future perspectives

High-quality nucleotide sequences of *Arabidopsis* and rice genomes were determined in 2000 and 2005, respectively. Molecular and genomic analyses have revealed many genes that are induced by abiotic stress and their products function in stress response

and tolerance. Transcriptome analyses based on microarray have provided powerful tools in gene discovery of stress-responsive genes from various crops and trees. Transgenic plants with antisense or RNAi constructs and T-DNA- or transposon-tagged mutants are used to analyze their function based on phenotypes due to loss-of-function of genes. Moreover, transgenic overexpressors are also useful not only for the functional analyses of genes but also for improving stress tolerance by gene transfer. The combination of molecular, genomic, and genetic analyses will elucidate complex regulatory systems in abiotic-stress-responsive gene expression (Fig. 3). *Arabidopsis* stress-related genes are useful both in the improvement of abiotic stress tolerance in transgenic crops and trees and in the discovery of stress-related genes of various crops and trees based on comparative genomics.

References

Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K. 1997. Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9:1859-1868.

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63-78.

Bartels D, Sunkar. 2005. Drought and salt tolerance in plants. Crit. Rev. Plant Sci. 24:23-58.

Choi H, Hong JH, Ha J, Kang JY, Kim SY. 2000. ABFs, a family of ABA-responsive element-binding factors. J. Biol. Chem. 275:1723-1730.

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J. 33:751-763.

Fowler S, Thomashow MF. 2002. *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14:1675-1690.

Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, Tran LSP, Yamaguchi-Shinozaki K, Shinozaki K. 2004. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J. 39:863-876.

Iuchi S, Kobayshi M, Taji T, Naramoto M, Seki M, Kato T, Tabata S, Kakubari Y, Yamaguchi-Shinozaki K, Shinozaki K. 2001. Regulation of drought tolerance by gene manipulation of 9-*cis*-epoxycarotenoid, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. Plant J. 27:325-333.

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF. 1998. *Arabidopsis* CBF1 overexpression induces COR genes and enhances freezing tolerance. Science 280:104-106.

Kang JY, Choi HI, Im MY, Kim SY. 2002. *Arabidopsis* basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell 14:343-357.

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. Nature Biotech. 17:287-291.

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert H. 2001. Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13:889-905.

Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. Plant Physiol. 130:2129-2141.

Liu Q, Sakuma Y, Abe H, Kasuga M, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain, separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell 10:1391-1406.

Maruyama K, Sakuma Y, Kasuga M, Ito Y, Seki M, Goda H, Shimada Y, Yoshida S, Shinozaki K, Yamaguchi-Shinozaki K. 2004. Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. Plant J. 38:982-993.

Mustilli AC, Merlot S, Vavasseur A, Fenzi F, Giraudat J. 2002. *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. Plant Cell 14:3089-3099.

Nakashima K, Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K. 1997. A nuclear gene, *erd1*, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally up-regulated during senescence in *Arabidopsis thaliana*. Plant J. 12:851-861.

Nambara E, Marion-Poll M. 2005. Abscisic acid biosynthesis and catabolism. Annu. Rev. Plant Biol. 56:165-185.

Rabbani, MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. Monitoring expression profiles of rice (*Oryza sativa* L.) genes under cold, drought and high-salinity stresses, and ABA application using both cDNA microarray and RNA gel blot analyses. Plant Physiol. 133:1755-1767.

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. 2002. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem. Biophys. Res. Commun. 290:998-1009.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. Nature Genet. 37:501-506.

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguhci-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K. 2001. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13:61-72.

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, et al. 2002a. Monitoring the expression profiles of ca. 7000 *Arabidopsis* genes under drought, cold, and high-salinity stresses using a full-length cDNA microarray. Plant J. 31:279-292.

Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, et al. 2002b. Monitoring the expression pattern of ca. 7000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. Func. Integ. Genom. 2:282-291.

Shinozaki K, Yamaguchi-Shinozaki K, Seki M. 2003. Regulatory network of gene expression in the drought and cold stress responses. Curr. Opin. Plant Biol. 6:410-417.

Simpson SD, Nakashima K, Narusaka Y, Seki M. Shinozaki K, Yamaguchi-Shinozaki K. 2003. Two different novel *cis*-acting elements of erd1, a clpA homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant J. 33:259-270.

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K. 2002. Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. Plant J. 29:417-426.

Tran LSP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M. Shinozaki K, Yamaguchi-Shinozaki K. 2004. Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive *cis*-element in the *early responsive to dehydration stress 1* promoter. Plant Cell 16:2481-2498.

Umezawa T, Yoshida R, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K. 2004. SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 101:17306-17311.

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. 2000. *Arabidopsis* basic leucine zipper transcriptional transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc. Natl. Acad. Sci. USA 97:11632-11637.

Vogel JT, Zarka DG, Van Buskirk HA, Fowler SG, Thomashow MF. 2005. Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. Plant J. 412:195-211.

Yamaguchi-Shinozaki K, Shinozaki K. 1994. A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell 6:251-264.

Yamaguchi-Shinozaki K, Shinozaki K. 2005. Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci. 10:88-94.

Yoshida R, Hobo T, Ichimura K, Mizoguchi T, Takahashi F, Alonso J, Ecker JR, Shinozaki K. 2002. ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in *Arabidopsis*. Plant Cell Physiol. 43:1473-1483.

Zhang JZ, Creelman RA, Zhu JK. 2004. From laboratory to field: using information from *Arabidopsis* to engineer salt, cold, and drought tolerance in crops. Plant Physiol. 135:615-621.

Zhu JK. 2002. Drought and salt stress signal transduction in plants. Annu. Rev. Plant Biol. 53:247-273.

Notes

Authors' addresses: K. Shinozaki, RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 203-0045, Japan (e-mail: shinozaki@rtc.riken.go.jp, www.psc.riken.go.jp/); K. Yamaguchi-Shinozaki, Japan International Research Center for Agricultural Sciences (JIRCAS), 1-1 Ohwashi, Tsukuba 305-8686, Japan (e-mail: kazukoys@jircas.affrac.go.jp).

Expression and functional analysis of rice genes involved in reproductive development and stress response

A.K. Tyagi, J.P. Khurana, P. Khurana, S. Kapoor, V.P. Singh, A.K. Singh, J.K. Thakur, V. Gupta, S. Anand, S. Vij, M. Jain, S. Ray, P. Agarwal, R. Arora, P. Sharma, S. Mukherjee, A. Nijhawan, J. Giri, and R. Khurana

The rice genome sequenced and annotated by the IRGSP has identified 37,544 protein-coding genes. In an effort to identify genes encoding transcription factors and signal transduction components, more than 7,000 genes belonging to 87 classes have been used to prepare a local database. Detailed analysis of genes for plant hormone response, CDPKs, C_2H_2 zinc-finger, and SET domain proteins unraveled interesting evolutionary aspects in relation to genes and the rice genome. A 51k microarray, SAGE analysis, and real-time polymerase chain reaction revealed differential expression of target genes during reproductive development and stress conditions. Several genes specific to reproductive floral organs and seed development have been identified. A large number of SAGE tags are observed from intergenic regions and antisense strands reflecting the unexplored transcription potential of the rice genome. Analysis of rice gene promoter activities has been undertaken in transgenic tobacco/*Arabidopsis* to demarcate regions conferring anther-/pollen-specific expression. *OSISAP1*, a gene coding for a stress-associated zinc-finger protein, and its promoter have been functionally validated in transgenic tobacco and rice. Genes for proteins interacting with OSISAP1 have also been found to be stress-inducible. Investigations on functional analysis of stress-responsive genes are in progress.

Keywords: Functional genomics, microarray, phylogenetic analysis, reproductive development, rice, SAGE, signal transduction, stress, transgenics, transcription factors

Among food-grain crops, rice occupies an eminent position and it represents an important agricultural activity. For about three billion Asians, rice provides 25–80% of their calories and most of the global production is consumed directly by humans. Although rice shows a wide range of variability as well as adaptability and breeders have regularly improved rice varieties, the rice plant needs to be redesigned to improve production and incorporate better potential to ward off diseases or insects and adapt to unfavorable environments. This can be aided by recent advances in genomics and biotechnology. Progress in these areas is represented by the availability of integrated

genetic and physical maps, the sequence of the entire genome, as well as functional validation of several genes/QTLs, molecular breeding, and transgenics in rice (Tyagi et al 1999, 2004, International Rice Genome Sequencing Project 2005, Sasaki et al 2005, Yu et al 2005). The availability of 18,828 simple sequence repeats (SSRs) and a large number of single nucleotide polymorphisms (SNPs) provides tremendous potential for mining useful alleles. The challenge of precise annotation of genes and understanding their function at the molecular level still lies ahead. This requires development of expression profiles of a set of genes acting in concert during a particular state and functional validation by gene-knockout, overexpression, or proteome analyses. We have initiated a program to analyze genes for transcription factors and signal transduction components in relation to reproductive development and stress response in rice. We expect that such studies will provide a better understanding of the processes responsible for better yield and survival.

Analysis of rice genes for transcription factors and signal transduction components

The map-based finished quality sequence, representing >95% of an estimated 389-Mb rice genome, has been annotated into 37,544 non-transposable-element-related protein coding genes (International Rice Genome Sequencing Project 2005). The number of genes unique to rice and other cereals is estimated to be 2,859, whereas 26,837 (71%) genes have putative homologs in *Arabidopsis* and 22,840 (61%) could be supported by rice expressed sequence tags (ESTs) or full-length cDNAs. It can also be noted that almost 60% of the rice genome is duplicated. To identify genes for this investigation, NCBI, TIGR, and KOME databases for rice were accessed. In all, 4,852 signal transduction–related genes in 42 categories and 2,341 transcription factor genes in 43 categories were identified. A local database of 7,193 genes with information regarding their nucleotide sequence, protein sequence, chromosome location, 5′ and 3′ mRNA boundaries, availability of cDNA, expression profile, and probable orthologs in other systems was prepared. It can be noted that this number represents predicted genes as well as alternatively spliced mRNAs. Out of signal transduction–related genes, 1,679 represent kinases, followed by 646 genes for transporters, 439 genes for phosphatases, 300 genes for ATPases, and 259 genes for pentatricopeptide repeat proteins. For transcription factors, zinc-finger proteins are represented by at least 798 genes, followed by MYB, NAM/NAC, ERF, and bHLH proteins being represented by at least 270, 134, 133, and 107 genes, respectively. Some of these categories have been analyzed in depth from the structural and phylogenetic point of view.

Auxin and cytokinin response genes

The phytohormone auxin plays a vital role in regulating many fundamental processes throughout the life cycle of plants, including cell division, apical dominance, tropic responses, stem elongation, lateral root development, vascular differentiation, and embryogenesis. To understand the molecular mechanism of auxin action, a large num-

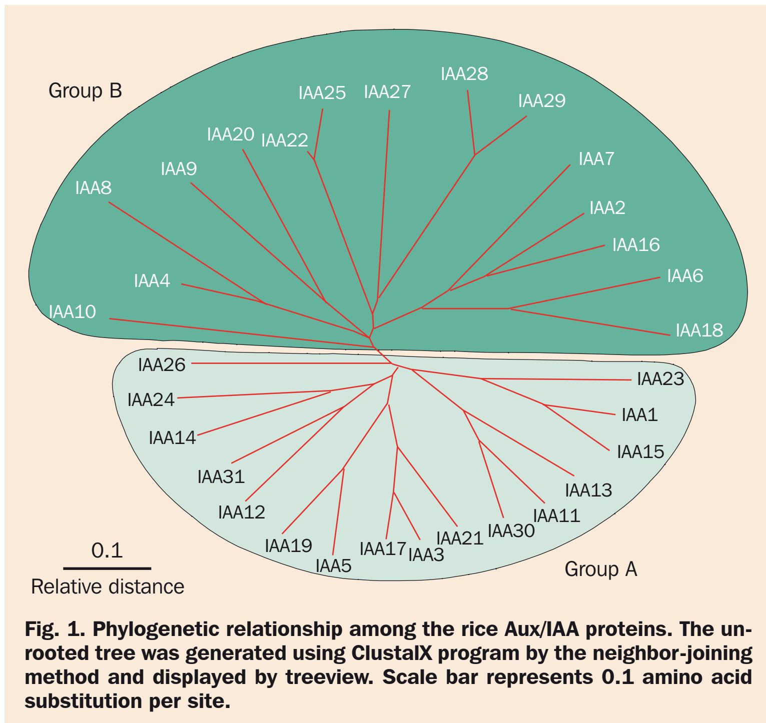


Fig. 1. Phylogenetic relationship among the rice Aux/IAA proteins. The unrooted tree was generated using ClustalX program by the neighbor-joining method and displayed by treeview. Scale bar represents 0.1 amino acid substitution per site.

ber of auxin-responsive genes have been identified and characterized from different plants (Hagen and Guilfoyle 2002, Thakur et al 2001, 2005). These auxin-responsive genes have been broadly grouped into three major classes: auxin/indoleacetic acid (*Aux/IAA*), *GH3*, and small auxin-up RNA (*SAUR*). In *Arabidopsis thaliana*, the auxin-responsive genes *Aux/IAA, GH3*, and *SAUR* are represented by large multigene families containing 19, 28, and 72 members, respectively (Hagen and Guilfoyle 2002, Liscum and Reed 2002). The analysis of the complete genome of rice revealed that auxin-inducible *GH3* and *Aux/IAA* gene families contain 12 and 31 members, designated as *OsGH3-1* to *-12* and *OsIAA1* to *31*, respectively (Jain et al 2006a,b). A preliminary analysis showed that the *SAUR* gene family is also represented by more than 50 members in rice.

The 31 rice *Aux/IAA* genes identified were found to be distributed on 10 of the 12 rice chromosomes. Most of the OsIAA proteins showed the presence of all four characteristic conserved domains. The coding sequence of the majority of the *Aux/IAA* genes (13 among 31) was disrupted by four introns at perfectly conserved positions with respect to their amino acid sequence, suggesting a common ancestral gene with a classical pattern of five exons and four introns. The phylogenetic analysis clustered all the rice Aux/IAA protein sequences into two major groups (A and B) and formed 12 sister pairs (Fig. 1), nine of which were found to be located on the duplicated chro-

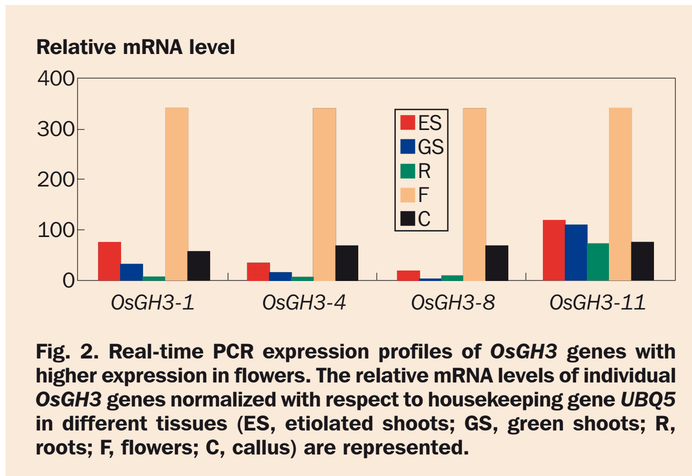


Fig. 2. Real-time PCR expression profiles of *OsGH3* genes with higher expression in flowers. The relative mRNA levels of individual *OsGH3* genes normalized with respect to housekeeping gene *UBQ5* in different tissues (ES, etiolated shoots; GS, green shoots; R, roots; F, flowers; C, callus) are represented.

mosomal blocks as described by Paterson et al (2004). In *Arabidopsis* also, *Aux/IAA* genes formed ten sister pairs and all of them were found to be located on homologous duplicated chromosomal segments (Remington et al 2004). Thus, it is remarkable that the duplication and retention of the sister pairs of *Aux/IAA* genes is associated with chromosomal block duplications in both rice and *Arabidopsis*. The fact that the *Aux/IAA* gene family in both rice and *Arabidopsis* showed a high degree of duplicated gene retention is particularly interesting as both species experienced different events during genome evolution. All 31 members showed a complexity of expression patterns in various organs or in response to exogenous auxin and light stimuli (Jain et al 2006a), indicating their crucial role in plant growth and development.

The 12 identified rice *GH3* genes are distributed on only 5 of the 12 rice chromosomes. The *Arabidopsis* GH3 proteins have been classified into three groups on the basis of their protein structure and specificity to adenylate plant hormones (Staswick et al 2002). In the phylogenetic analysis, all the rice and *Arabidopsis* GH3 proteins clustered distinctly into three groups. Group III, however, included only *Arabidopsis* GH3 proteins with unknown functions. The homologs of genes belonging to this group could not be identified either in the rice sequence database or EST databases of other monocots. We can speculate that group III GH3 proteins were lost in rice after the divergence of monocots and dicots or evolved in dicots after the divergence from monocots and may perform dicot-specific functions. As far as the expression profile is concerned, real-time PCR analysis showed that *OsGH3* genes are differentially expressed in various tissues/organs and some *OsGH3* genes (*OsGH3-1*, *-4*, *-8*, and *-11*) showed significantly higher expression in flowers than in others tissues (Fig. 2), indicating their role in floral development (Jain et al 2006b).

The response regulators represent the elements of a bacterial two-component system. Analysis of the *Arabidopsis* genome revealed the existence of 32 putative response regulator genes classified into type-A and type-B response regulators (Hwang

et al 2002). We have identified ten genes encoding type-A response regulators in rice (designated as *Oryza sativa* response regulators, *OsRR1* to *10*) based on their high sequence similarity within the receiver domain. The type-A response regulators represent the primary cytokinin response genes because the steady-state transcript levels of most of these genes increases rapidly (within 15 min) on exogenous cytokinin application even in the absence of *de novo* protein synthesis. Moreover, the expression of the *OsRR6* gene is enhanced in rice seedlings exposed to salinity, dehydration, and low-temperature stress, indicating its role in cross-talk between abiotic stress and cytokinin signaling. It is important to note that cytokinin levels might be important for rice grain production (Ashikari et al 2005). Our results provide a foundation for future studies on elucidating the precise role of auxin and cytokinin signal transduction genes in rice.

Calcium-dependent protein kinase genes

Calcium is a ubiquitous second messenger of the eukaryotic signal transduction cascade. Calcium-dependent protein kinases (CDPKs) are found throughout the plant kingdom from algae to angiosperms (Harmon et al 2001). CDPKs possess a characteristic structure with our domains: the N-terminal domain, the kinase domain, the autoinhibitory domain, and the calmodulin-like domain (Cheng et al 2002, Hrabak et al 2003). The CDPK genes do not depend on the exogenous calmodulin, but can be directly activated by calcium-binding to the EF-hand calcium-binding sites. An *in silico* analysis of *Arabidopsis* and rice genome sequences had earlier revealed 34 and 29 CDPK genes, respectively (Cheng et al 2002, Hrabak et al 2003, Asano et al 2005). CDPKs are found to be involved in diverse functions such as pollen development, embryogenesis, seed development, germination, tuberization, and various stress responses (Ludwig et al 2004). To identify rice CDPK genes, a genome-wide *in silico* analysis was performed. This exercise revealed 33 potential genes encoding calcium-dependent protein kinases, ranging in size from 51 to 65 kDa. The kinase domain was found to exhibit a high degree of sequence conservation at the active site. The calmodulin-like domain was 50–217 amino acids long and contained calcium-binding EF-hands. Phylogenetic analysis revealed that rice CDPKs are clustered in four distinct groups, as in dicots (*Arabidopsis thaliana*). The highest degree of homology between any two CDPKs of *Arabidopsis* and rice was found to be 80.4%, and the maximum divergence between any two CDPKs was 68.5%. The four groups were found to be evenly distributed on all 12 chromosomes. Rice chromosomes seem to have undergone segmental, tandem, and background duplication during evolution (Yu et al 2005). Chromosomes 11 and 12 are known to have a duplicated region in the distal end of the short arm (The Rice Chromosomes 11 and 12 Sequencing Consortia 2005). Incidentally, a pair of CDPK genes located on the short arm of both chromosomes 11 and 12 may have resulted from the segmental duplication mentioned above. Duplicated CDPK genes on 7.6- and 10.8-cM positions, on chromosomes 11 and 12, respectively, have 97.5% homology in their nucleotide sequences. The order of the adjacent genes too is highly conserved. The other duplicated CDPK genes on 19- and

27.6-cM positions of chromosomes 11 and 12, respectively, showed 83.5% homology. In this case, however, the order of the adjacent genes was found to be shuffled. A comparison of the cDNA sequence with its respective genomic loci has revealed that most CDPK genes contain six to seven introns. The number of introns, however, has been found to vary from none to 20.

C_2H_2 zinc-finger genes

Cys_2/His_2 zinc-finger transcription factors have been reported to play an important role in development and in imparting stress tolerance in the model plant *Arabidopsis* (Sagasser et al 2002, Sakamoto et al 2004). As stated earlier, databases were searched for genes encoding C_2H_2 zinc-finger transcription factors by performing name searches and BLAST searches with the plant-specific conserved domain QALGGH and a consensus sequence of the zinc-finger motif taken from petunia (Takatsuji et al 1994). All the sequences that had a C_2H_2 motif were aligned by ClustalX version 1.83 (Thompson et al 1997). A total of 173 unique sequences were found. When these sequences were analyzed by PROSITE, many of the motifs could not be identified as C_2H_2 zinc-fingers. This could be because of extensive modification during the course of evolution. This observation is similar to that of the yeast zinc-finger analysis in which the PROSITE pattern to search for the C_2H_2 finger had to be extended to identify all the zinc-finger proteins (Bohm et al 1997). In rice, three types of C_2H_2 motifs could be distinguished: QALGGH motif (Q type), modified QALGGH (M type), and without QALGGH (C type). Although the Q-type zinc-finger is exclusively found in plants, rice showed a maximum of C-type zinc-fingers. The motif length was found to vary from 15 to 34 amino acids. These zinc-finger proteins (ZFPs) showed varying numbers of zinc-finger domains per protein, ranging from one to nine. In proteins having one or two zinc-fingers, the Q type was the most common.

Rice shows the presence of both tandemly placed and widely separated zinc-finger domains. The tandem zinc-fingers of rice are maximally present in proteins containing C-type zinc-fingers. None of the Q-type zinc-fingers were found to be tandemly placed. The residues between the second histidine of the previous zinc-finger and the first cysteine of the next zinc-finger form the "spacer" or the "linker" between the two zinc-finger domains (Takatsuji 1998). Like in other species, spacer length in rice was found to be highly variable. Although most fingers have a spacer length of less than 100 residues, some are extremely long, extending into the hundreds of amino acids.

The evolutionary relationship among all 173 rice ZFPs was analyzed by constructing an unrooted phylogenetic tree taking *Synechocystis* sequence as the outgroup. This exercise resulted in four distinct groups. All the proteins of group I have a C-type zinc-finger with the number varying from one to nine per protein. A large number of group II members were found with at least two Q-type zinc-fingers. Groups II and III mainly have Q-type zinc-finger domains, whereas groups I and IV predominantly have C-type domains. Most of the proteins with only one zinc-finger fall in group III. When compared with *Arabidopsis*, yeast, and human proteins, rice ZFPs showed

maximum similarity with the proteins of *Arabidopsis*. Members of groups I and IV showed limited homology also to their yeast counterparts. Therefore, we can infer that these two groups would have evolved relatively early. When the intron analysis of these genes was carried out using SIM4 software, a large number of genes were found to be intronless. However, four genes were found to contain as many as 11 introns. No significant bias was seen for chromosome-specific distribution of genes.

Genes for SET domain proteins

An important domain present within many Polycomb and Trithorax proteins is the SET domain named after three *Drosophila* proteins in which they were originally identified; suppressor of variegation [Su(var)3-9], enhancer of zeste [E(Z)], and trithorax. These proteins play an important role in regulating development by modulating chromatin structure. At least 29 genes encoding SET domain proteins have been reported in *Arabidopsis thaliana* (Baumbusch et al 2001). In rice, 44 SET domain protein genes were identified. The number of SET domain-encoding genes, however, stands at 41 since primary transcripts of three genes undergo alternative splicing to generate two different protein products each. Full-length cDNAs corresponding to at least 32 of these genes are available in the KOME database. Three genes seem to be duplicated. The SET domain proteins were further categorized into four groups, E(Z), TRX, ASH, and Su(var), depending upon the similarity in the SET domain and other domains/regions of putative structural and biochemical significance. A majority of the proteins fall into classes Su(var) and ASH. Only two proteins represent the E(Z) class—OsiEZ1 and OsiCLF. The *OsiEZ1* gene was earlier characterized in our laboratory and found to express preferentially in the reproductive tissue (Thakur et al 2003). Further characterization of *OsiEZ1* and *OsiCLF* is in progress.

Reproductive development

Reproduction, being the most important feature of living organisms to maintain the continuity of the race, has always fascinated the scientific community. In recent years, a significant amount of work has been done in the area of molecular plant reproduction, especially in *Arabidopsis*, *Antirrhinum*, petunia, and maize. One of the major achievements in this area is the conceptualization of the ABC model, which explains how a set of genes can specify floral organ identity. In addition, several other genes involved in the transition of vegetative to flowering shoot apex, flower development, gamete development, fertilization, and embryo/seed development have been identified (Drews and Yadegari 2002, Bommert et al 2005, Itoh et al 2005). Genome-wide assessment of such genes in rice is still required, as is their functional analysis.

Gene expression profile during flower development

We have attempted to use high-throughput techniques such as SAGE and microarray to get a broader idea of the genes involved in flower development. Recently, Bao et al (2005) have used SAGE to study the molecular mechanisms of heterosis by compar-

ing two cultivars of rice. We have employed this technique to identify differentially expressed genes at different stages of rice panicle development because in SAGE small differences in transcript and genes expressing at a very low level can also be studied. For the sake of making the study simpler and making description applicable to different species and mutant phenotypes, panicle development was divided into four categories on the basis of landmark events in the development. RNA isolated from these stages of flower development along with that from mature leaves was used to generate five long-SAGE libraries. In addition, one library was also generated from 3.5–4.0 cm of panicle that represents the male meiosis-specific stage. After sequencing the positive clones, tags were extracted from selected sequences using SAGE 2000, version 4.5 analysis software (www.invitrogen.com/sage). Ambiguous base calls were removed by applying a Phred score cut-off of 20 to all sequences. Genes were assigned to tags by comparing our library with the virtual library constructed from cDNA clones in KOME and the reference library available in TIGR (The Rice Full-Length cDNA Consortium 2003). The tag to gene assignment was confirmed for a few genes by manual BLAST analysis in NCBI. More than 150,000 tags have been sequenced and analyzed, which amounts to approximately 41,000 unique tags. A large number of them correspond to transposons and retrotransposons. About 60% of the unique tags are represented singly, implying that most of the transcripts are expressed at very low levels. Genes for metallothionein and α-amylase were found to be highly expressed in these libraries.

As a preliminary attempt, the expression pattern of 35 MADS box transcription factors was analyzed by combining data of all five libraries. Expectedly, three MADS box genes, which are known to have a role in flower development, *OsMADS1, OsMADS6,* and *OsMADS16,* were found to express in flowers at high levels and were absent in the leaf library. On the other hand, genes such as *RBCS*l and *CAB* that are involved in photosynthesis were found to be highly expressed in the leaf library. One of the important observations during our analysis is that a significant number of tags match with the antisense of annotated transcripts and the intergenic regions, thus opening new vistas to study the unexplained territory of the genome.

Because of the availability of cDNA as well as oligonucleotide arrays, microarray technology has been employed to study the genes involved in reproductive organ development. In rice, cDNA microarrays have been employed to identify and characterize novel anther-specific genes (Endo et al 2004) and genes involved in pollination and fertilization (Lan et al 2004). In our study, we are using the 51k rice Genechip® to study the genes differentially expressed during floral-organ development. Stages similar to SAGE analysis were used for microarrays to help compare data from both sources. Each experiment was performed in triplicate and good correlation was found in biological replicates of each tissue. The initial data analysis for MADS box genes revealed a few panicle-specific genes. After comparing SAGE and microarray data for the MADS box genes, similar expression patterns were revealed in both data sets, except for a few genes (1g and 1h) for which a SAGE expression pattern could not be obtained because of the low level of expression (Fig. 3). Information on genes for

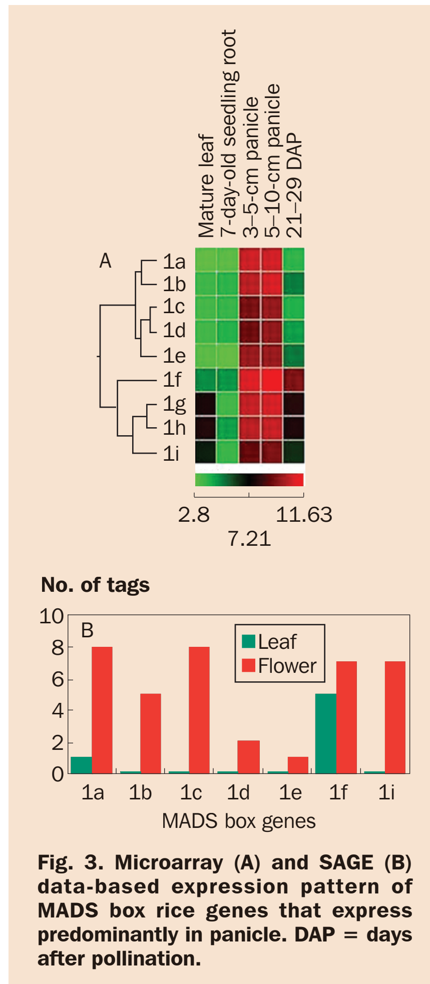


Fig. 3. Microarray (A) and SAGE (B) data-based expression pattern of MADS box rice genes that express predominantly in panicle. DAP = days after pollination.

transcription factors and signal transduction components will be used to select targets for functional analysis.

Analysis of anther-/pollen-specific promoters

Gene expression involves a number of interactive processes (Tyagi 2001). In pollen, regulation of gene expression occurs at the level of transcription as well as post-transcription. The 5′ upstream region of genes contains specific sequences, GTGA, AGAAA, and TCCACCATA, that are shown to be important for the regulation of pollen gene expression (Filichkin et al 2004, Jiang et al 2005). The role of 5′UTR sequences and the evolutionarily conserved genetic programs in pollen led to the hypothesis that the 5′UTRs of pollen-expressed genes share regulatory sequence elements (Hulzink et al 2003). Analysis of promoters has been made possible by the

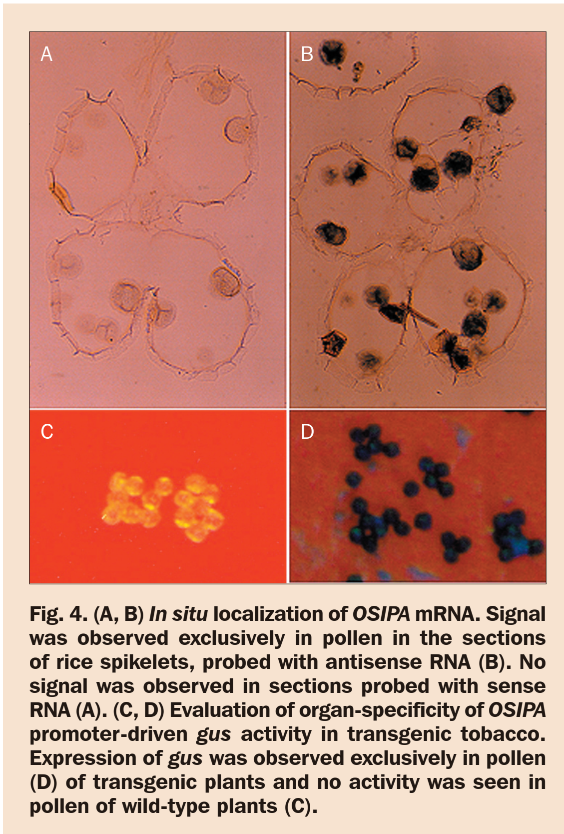


Fig. 4. (A, B) *In situ* localization of *OSIPA* mRNA. Signal was observed exclusively in pollen in the sections of rice spikelets, probed with antisense RNA (B). No signal was observed in sections probed with sense RNA (A). (C, D) Evaluation of organ-specificity of *OSIPA* promoter-driven *gus* activity in transgenic tobacco. Expression of *gus* was observed exclusively in pollen (D) of transgenic plants and no activity was seen in pollen of wild-type plants (C).

development of techniques for generating fertile transgenic plants containing chimeric promoter-reporter gene constructs. Moreover, databases of eukaryotic promoters (EPD) and plant promoter sequences (PlantProm) enhance promoter research and analysis (Praz et al 2002, Shahmuradov et al 2003).

Three anther- and pollen-specific genes (*OSIPA*, *OSIPK*, and *OSIPP2*) were isolated by differential screening of cDNA libraries made from prepollinated (PP) and postfertilization (PF) stage inflorescence of rice. Their promoters were also isolated by screening of the rice genomic DNA library and found to contain anther- and pollen-specific elements. *OSIPA* is expressed in a pollen-specific manner (Fig. 4). To characterize the activity of the *OSIPA* promoter, the region between –1,823 and +64 bp was cloned upstream to *gus*. Transgenic tobacco and *Arabidopsis* lines generated were found to show *gus* expression right from the early stages of pollen development (Fig. 4). As the flower matures, *gus* expression keeps on increasing and the maximum expression level is observed in pollen after the dehiscence of the anthers. Another flower-specific gene, *OSIPK1* (a calcium-dependent protein kinase), expresses specifically in anthers. The activity of the *OSIPK1* promoter was also checked by transforming tobacco and

Arabidopsis plants. The transgenic tobacco lines analyzed revealed that *gus* expression started in the anthers (including pollen) of 11-mm-stage flowers and expression continued till the maturity of the flowers. No activity was seen in anther and pollen where the flower opened and anther dehisced. *OSIPP2* is a pollen-preferential gene from rice. The activity of the *OSIPP2* promoter was studied in both a heterologous and homologous system. In transgenic tobacco plants, reporter gene (*gus*) activity is maximum in pollen grains and the growing tips of the pollen tube. Transgenic rice plants expressed *gus* (driven by the *OSIPP2* promoter) preferentially in pollen grains; however, low-level activity was detected in other tissues. Efforts are on to delineate the pollen-specific element. The functionality of elements and the presence of other novel elements in these promoters can be elucidated by mutational and gain-of-function analysis. A tapetum-specific promoter has been exploited to induce male sterility in tobacco earlier and recently also in cabbage (Lee et al 2003). Similarly, these pollen-specific promoters may have a potential to target genes of interest in pollen and have immense biotechnological application.

Gene expression during seed development

The edible part in rice is the caryopsis or seed. The husk or hull of the rice seed encloses the true fruit or brown rice, which consists mainly of the embryo and the endosperm. In rice, fertilization follows immediately after pollination. The day of floral anthesis is considered as 0 DAP (days after pollination). The zygote divides to form a globular embryo at 3 DAP. The first morphogenic change occurs at 4 DAP with the formation of the shoot apical meristem and radicle along with coleoptile initiation. Prior to this stage, the fertilized zygote is barely visible to the naked eye. At 5 DAP, the first leaf primordium appears. Then, the embryo undergoes a rapid developmental phase and, by 10 DAP, morphological growth of the rice embryo is complete. Thereafter, the seed enters a phase of maturation and dormancy and, beyond the 20-DAP stage, there is no significant morphological development (Itoh et al 2005).

To study the expression profile of CDPK family genes, microarray analysis was preformed using five stages of seed development along with tissue from the root of a 7-day-old seedling and mature leaves. Hierarchical cluster analysis of 29 CDPKs grouped them into two broad classes according to the level of expression. On the basis of differential expression pattern, however, these genes were clustered in four classes (Fig. 5). Class I contained eight CDPK genes, which showed a higher level of expression during early seed development in comparison with roots and leaf RNA as well as mature seeds. In class II, six CDPK genes were found to express in reproductive and vegetative organs at a similar level. The highest number (13) of CDPK genes, belonging to class III, exhibited high levels of expression in vegetative tissues; however, their respective RNA levels showed a gradual decline during seed maturation. It is interesting to note that two class III CDPK genes, which may have resulted from segmental duplication on chromosomes 11 and 12, showed a similar expression pattern. In class IV, two CDPK genes were found to have a high level of expression in all the seed developmental stages of rice. When the expression intensity of the CDPK genes in seed was compared with that in root and leaf, nine genes

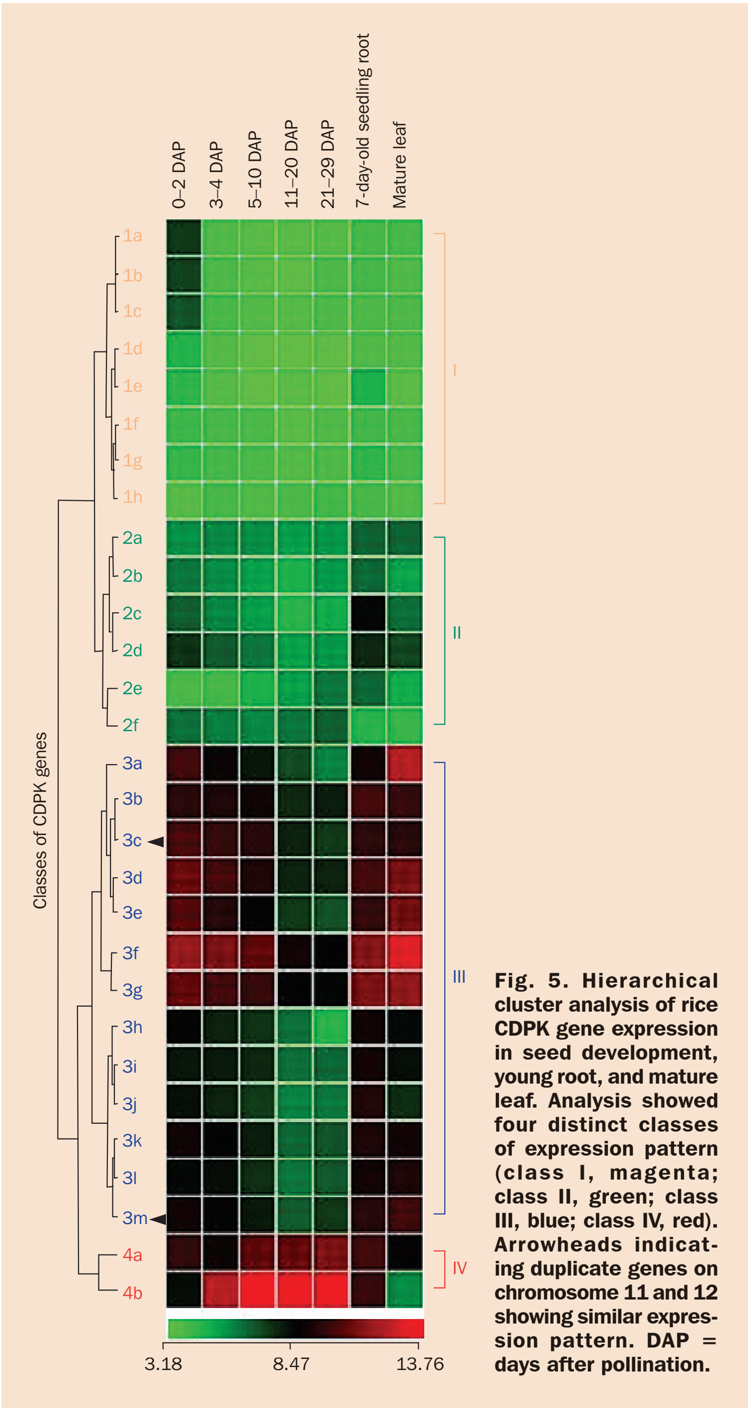


Fig. 5. Hierarchical cluster analysis of rice CDPK gene expression in seed development, young root, and mature leaf. Analysis showed four distinct classes of expression pattern (class I, magenta; class II, green; class III, blue; class IV, red). Arrowheads indicating duplicate genes on chromosome 11 and 12 showing similar expression pattern. DAP = days after pollination.

showed more than a twofold increase in expression in the seed developmental stages. To further extend our study on the expression patterns of the CDPK genes, quantitative polymerase chain reaction (PCR) was performed on three seed developmental stages. This analysis validated the expression of the genes, which showed seed-specific differential expression in microarray analysis. These data show that rice CDPK class I genes might be important during the early stages of seed development, whereas class IV genes, which show sustained higher expression throughout the development of seed, could be involved in seed storage functions. Detailed investigations to understand the roles of individual CDPKs in controlling seed development are under way.

Out of 173 rice C_2H_2 zinc-finger transcription factors, 155 were represented on the microarray being used. Fourteen genes showed significantly higher expression in seeds than the controls. In addition, 57 genes showed temporal expression in seeds during development, although they also showed considerable expression in the controls. On the basis of their expression patterns in the seed samples, we have categorized these genes into four broad groups. Further characterization of these genes would highlight their role in seed development.

Analysis of genes for stress tolerance

Abiotic stresses such as drought, salinity, extremes of temperature, chemical toxicity, and oxidative stress are reported to account for the maximum loss of plant productivity. The plant's response to abiotic stress is marked by morphological, physiological, biochemical, and molecular changes, which have a negative effect on both plant growth and productivity (Wang et al 2003). At the molecular level, the changes are seen in the form of altered gene expression profiles (Cooper at al 2003, Rabbani et al 2003). The aim of studying the genes involved in abiotic stress response at the cellular level is not only to understand the complexity of stress response but also to identify the gene(s) that can be genetically manipulated to obtain stress-tolerant crops.

OSISAP1, a gene coding for a stress-associated zinc-finger protein from rice, was characterized to understand its role in stress-response (Mukhopadhyay et al 2004). It contains two zinc-fingers identified as A20 and AN1 types. In rice and *Arabidopsis*, 10 and 11 genes with A20- and AN1-type zinc-fingers, respectively, have been identified. *OSISAP1* is induced by various kinds of stresses, including cold, salt, dehydration, heavy metals, and injury. The gene was overexpressed in tobacco to evaluate its role in stress tolerance. The transgenic plants showed tolerance of cold, salt, and dehydration stress at the seedling stage. *OSISAP1* has also been overexpressed in *Arabidopsis* using the floral dip method. All the transgenic plants showed a reduction in height and leaf size, with a prolonged life span. To confirm whether such changes are related to the constitutive expression of *OSISAP1*, T1 plants were grown and their transgenic nature was tested by PCR analysis. Interestingly, all PCR-positive plants showed altered phenotype with a prolonged life span. On the other hand, all PCR-negative plants showed normal vegetative growth. Experiments are also under way to express *OSISAP1* under the stress-inducible *rd29A* promoter in *Arabidopsis* (Kasuga et al 1999).

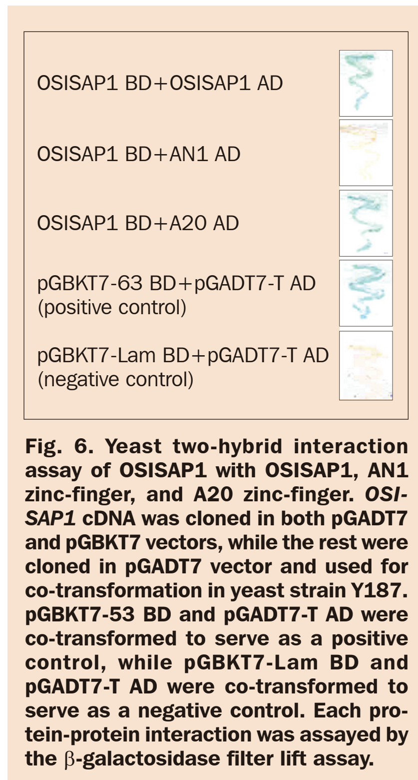


Fig. 6. Yeast two-hybrid interaction assay of OSISAP1 with OSISAP1, AN1 zinc-finger, and A20 zinc-finger. *OSISAP1* cDNA was cloned in both pGADT7 and pGBKT7 vectors, while the rest were cloned in pGADT7 vector and used for co-transformation in yeast strain Y187. pGBKT7-53 BD and pGADT7-T AD were co-transformed to serve as a positive control, while pGBKT7-Lam BD and pGADT7-T AD were co-transformed to serve as a negative control. Each protein-protein interaction was assayed by the β-galactosidase filter lift assay.

Protein-protein interaction studies revealed that OSISAP1 can interact with itself, possibly through A20 zinc-fingers (Fig. 6). This analysis also led to the identification of two other interacting proteins encoded by *OSISAP2* and *OSIRLCK* genes of rice. The response of *OSISAP2* and *OSIRLCK* to abiotic stress was evaluated by real-time PCR analysis and results were compared with those of *OSISAP1*. The mRNA levels of all three genes showed an increase in response to cold, salt, and dehydration stress compared to the unstressed control. To further study the role of *OSISAP2* and *OSIRLCK* in stress response, transgenic *Arabidopsis* plants overexpressing these genes under the control of the constitutive *CaMV35S* promoter have been obtained. We need to define the function of these genes in stress response, individually and in combination.

Conclusions

The expression profile of genes for signal transduction components and transcription factors emerges with a complexity and diversity of a very high order. Further, certain genes associated with reproductive development and stress response have already been identified and the function of certain regulatory sequences as well as genes unraveled.

Our next challenge would be to identify the key regulatory components among such genes and define their function.

References

Asano T, Tanaka N, Yang G, Hayashi N, Komatsu S. 2005. Genome-wide identification of the rice calcium-dependent protein kinase and its closely related kinase gene families: comprehensive analysis of the CDPKs gene family in rice. Plant Cell Physiol. 46:356-366.

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M. 2005. Cytokinin oxidase regulates rice grain production. Science 309:741-745.

Bao JY, Lee S, Chen C, Zhang XQ, Zhang Y, Liu SQ, Clark T, Wang J, Cao ML, Yang HM, Wang SM, Yu J. 2005. Serial analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. Plant Physiol. 138:1216-1231.

Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB. 2001. The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionary conserved classes. Nucl. Acids Res. 29:4319-4333.

Bohm S, Frishman D, Mewes H W. 1997. Variations of the *C2H2* zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. Nucl. Acids Res. 25:2464-2469.

Bommert P, Satoh-Nagasawa N, Jackson D, Hirano HY. 2005. Genetics and evolution of inflorescence and flower development in grasses. Plant Cell Physiol. 46:69-78.

Cheng SH, Willmann MR, Chen HC, Sheen J. 2002. Calcium signaling through protein kinases: the *Arabidopsis* calcium-dependent protein kinase gene family. Plant Physiol. 129:469-485.

Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S, Guimil S, Dunn M, Luginbuhl P, Ellero C, Goff SA, Glazebrook J. 2003. A network of rice genes associated with stress response and seed development. Proc. Natl. Acad. Sci. USA 100:4945-4950.

Drews GN, Yadegari R. 2002. Development and function of the angiosperm female gametophyte. Annu. Rev. Genet. 36:99-124.

Endo M, Tsuchiya T, Saito H, Matsubara H, Hakozaki H, Masuko H, Kamada M, Higashitani A, Takahashi H, Fukuda H, Demura T, Watanabe M. 2004. Identification and molecular characterization of novel anther-specific genes in *Oryza sativa* L. by using cDNA microarray. Genes Genet. Syst. 79:213-226.

Filichkin SA, Leonard JM, Monteros A, Liu P-P, Nonogaki H. 2004. A novel endo-beta-mannanase gene in tomato, *LeMAN5,* is associated with anther and pollen development. Plant Physiol. 134:1080-1087.

Hagen G, Guilfoyle TJ. 2002. Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Mol. Biol. 49:373-385.

Harmon AC, Gribskov M, Gubrium E, Harper JF. 2001. The CDPK superfamily of protein kinases. New Phytol. 151:175-183.

Hrabak EM, Chan CWM, Gribskov M, Harper JF, Choi JH, Halford N, Kudla J, Luan S, Nimmo HG, Sussman MR, et al. 2003. The *Arabidopsis* CDPK-SnRK superfamily of protein kinases. Plant Physiol. 132:666-680.

Hulzink RJ, Weerdesteyn H, Croes AF, Gerats T, van Herpen MM, van Helden J. 2003. *In silico* identification of putative regulatory sequence elements in the 5′-untranslated region of genes that are expressed during male gametogenesis. Plant Physiol. 132:75-83.

Hwang I, Chen HC, Sheen J. 2002. Two-component signal transduction pathways in *Arabidopsis*. Plant Physiol. 129:500-515.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793-800.

Itoh J-I, Nonomura KI, Ikeda K, Yamaki S, Inukai Y, Yamagishi H, Kitano H, Nagato Y. 2005. Rice plant development: from zygote to spikelet. Plant Cell Physiol. 46:23-47.

Jain M, Kaur N, Garg R, Thakur, JK, Tyagi AK, Khurana JP. 2006a. Structure and expression analysis of early auxin-responsive *Aux/IAA* gene family in rice (*Oryza sativa*). Funct. Integr. Genomics 6:47-59.

Jain M, Kaur N, Tyagi AK, Khurana JP. 2006b. The auxin-responsive *GH3* gene family in rice (*Oryza sativa*). Funct. Integr. Genomics 6:36-46.

Jiang SY, Cai M, Ramachandran S. 2005. The *Oryza sativa* no pollen (*Osnop*) gene plays a role in male gametophyte development and most likely encodes a C2-GRAM domain-containing protein. Plant Mol. Biol. 57:835-853.

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. Nature Biotechnol. 17:287-291.

Lan L, Chen W, Lai Y, Suo J, Kong Z, Li C, Lu Y, Zhang Y, Zhao X, Zhang X, et al. 2004. Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (*Oryza sativa* L.) with a 10k cDNA microarray. Plant Mol. Biol. 54:471-487.

Lee YH, Chung KH, Kim HU, Jin YM, Kim HI, Park BS. 2003. Induction of male sterile cabbage using a tapetum-specific promoter from *Brassica campestris* L. ssp. *Pekinensis*. Plant Cell Rep. 22:268-273.

Liscum E, Reed JW. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol. Biol. 49:387-400.

Ludwig AA, Romeis T, Jones JDG. 2004. CDPK-mediated signalling pathways: specificity and cross-talk. J. Exp. Bot. 55:181-188.

Mukhopadhyay A, Vij S, Tyagi A.K. 2004. Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. Proc. Natl. Acad. Sci. USA 101:6309-6314.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc. Natl. Acad. Sci. USA 101:9903-9908.

Praz V, Périer R, Bonnard C, Bucher P. 2002. The eukaryotic promoter database, EPD: new entry types and links to gene expression data. Nucl. Acids Res. 30:322-324.

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Itoh Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol. 133:1755-1767.

Remington DL, Vision TJ, Guilfoyle TJ, Reed JW. 2004. Contrasting modes of diversification in the *Aux/IAA* and *ARF* gene families. Plant Physiol. 135:1738-1752.

Sagasser M, Lu GH, Hahlbrock K, Weisshaar B. 2002. *A. thaliana* TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. Genes Dev. 16:138-149.

Sakamoto H, Maruyama K, Sakuma Y, Meshi T, Iwabuchi M, Shinozaki K, Yamaguchi-Shinozaki K. 2004. *Arabidopsis* Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. Plant Physiol. 136:2734-2746.

Sasaki T, Matsumoto T, Antonio BA, Nagamura Y. 2005. From mapping to sequencing, post-sequencing and beyond. Plant Cell Physiol. 46:3-13.

Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM, Solovyev VV. 2003. PlantProm: a database of plant promoter sequences. Nucl. Acids Res. 31:114-117.

Staswick PE, Tiryaki I, Rowe ML. 2002. Jasmonate response locus JAR1 and several related *Arabidopsis* genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. Plant Cell 14:1405-1415.

Takatsuji H. 1998. Zinc-finger transcription factors in plants. Cell Mol. Life Sci. 54:582-596.

Takatsuji H, Nakamura N, Katsumoto Y. 1994. A new family of zinc finger proteins in petunia: structure, DNA sequence recognition, and floral organ-specific expression. Plant Cell 6:947-958.

Thakur JK, Jain M, Tyagi AK, Khurana JP. 2005. Exogenous auxin enhances the degradation of a light down-regulated and nuclear-localized OsiIAA1, an Aux/IAA protein from rice, via proteasome. Biochim. Biophys. Acta 1730:196-205.

Thakur JK, Malik MR, Bhatt V, Reddy MK, Sopory SK, Tyagi AK, Khurana JP. 2003. A *POLYCOMB* group gene of rice (*Oryza sativa* L. subspecies *indica*), *OsiEZ1*, codes for a nuclear-localized protein expressed preferentially in young seedlings and during reproductive development. Gene 314:1-13.

Thakur JK, Tyagi AK, Khurana JP. 2001. *OsIAA1*, an *Aux/IAA* cDNA from rice, and changes in its expression as influenced by auxin and light. DNA Res. 8:193-203.

The Rice Chromosomes 11 and 12 Sequencing Consortia. 2005. Rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. BMC Biology 3:20.

The Rice Full-Length cDNA Consortium. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379.

Thompson JD, Gibson TJ, Plavniak F, Jeanmongin F, Higgins DG. 1997. The CLUSTAL-X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25:4876-4882.

Tyagi AK. 2001. Plant genes and their expression. Curr. Sci. 80:161-169.

Tyagi AK, Khurana JP, Khurana P, Raghuvanshi S, Gaur A, Kapoor A, Gupta V, Kumar D, Ravi V, Vij S, Khurana P, Sharma S. 2004. Structural and functional analysis of rice genome. J. Genet. 83:79-99.

Tyagi AK, Mohanty A, Bajaj S, Chaudhury A, Maheshwari SC. 1999. Transgenic rice: a valuable monocot system for crop improvement and gene research. Crit. Rev. Biotechnol. 19:41-79.

Wang W, Vinocur B, Altman A. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:1-14.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, et al. 2005. The genomes of *Oryza sativa*: a history of duplications. PLoS Biol. 3:e38.

Notes

Authors' addresses: Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110 021, India; V.P. Singh and A.K. Singh, Division of Genetics, Indian Agricultural Research Institute, New Delhi 110 012, India.

Acknowledgment: Research in our laboratories is financially supported by the Department of Biotechnology and the University Grants Commission, as well as by research fellowships from CSIR/UGC, Government of India.

Designing and constructing novel gene promoters to generate stress-tolerant plants without yield penalty

Tuan-hua David Ho, Chwan-Yang Hong, Ming-Tsair Chan, and Sumay Yu

Although genetic engineering has become an important practice in agricultural biotechnology, how to properly control the expression of transgenes in transgenic plants remains a challenging task. Strong constitutive promoters are routinely used in plant transformation, but sometimes their use leads to undesirable secondary effects and negatively affects the overall performance of transgenic plants. In order to maximize the benefits of transgenes and to avoid unexpected negative impact, tissue-specific stress-/ABA-inducible promoters have been designed and constructed based on knowledge learned from studies of native promoters. Microarray analysis and bioinformatics are also employed in an extensive search for stress-/ABA-inducible and tissue-specific promoters, and information obtained is used to broaden the foundation for constructing synthetic designer promoters. The efforts include (1) optimization of upstream ABA-/stress-responsive *cis*-acting elements such as AB responsive element (ABRE) and coupling element (CE), (2) a search for the most efficient minimal promoter and desirable introns, and (3) the addition of tissue-specific determinants. Various versions of synthetic stress-/ABA-inducible promoters have been constructed, and some of them have been tested in transgenic plants for the expression of beneficial genes in conferring stress tolerance. Although transgenic plants with either a strong constitutive promoter or synthetic stress-inducible promoter acquire an elevated level of stress tolerance, only the latter display normal growth and development without any apparent yield penalty under normal conditions.

Keywords: genetic engineering, gene expression, tissue-specific inducible promoters, yield penality, transgene

Genetic engineering for the improvement of plant performance has become a routine procedure in both basic and applied research. Herbicide-resistant and insect-resistant crop plants have been commercially available for several years. Transgenic plants with enhanced levels of tolerance of biotic and abiotic stresses have been reported by many laboratories (for a review, see Ho and Wu 2004). Strong constitutive promoters, such as the 35S promoter and those from actin and ubiquitin genes, are used to drive

the expression of transgenes in most of these projects. Although this is adequate for transgenes whose function is needed in most of the tissues throughout the development of plants, the use of strong constitutive promoters could occasionally lead to an unnecessary burden for the development and productivity of transgenic plants. For example, transgenic *Arabidopsis* plants overexpressing a transcription factor, DREB1A, show remarkable enhancement in stress tolerance, but also display a dwarf phenotype (Kasuga et al 1999). This is mostly likely caused by the constitutive expression of a beneficial gene whose function is needed only when plants are under stress. Therefore, it is essential to develop tissue-specific inducible promoters that are active in targeted cell types during specific developmental stages and/or under certain environmental conditions. These "molecular gene switches" would be useful tools in future genetic engineering research for properly controlling the expression of transgenes so that their beneficial effects could be realized without any undesirable side effects. Toward this goal, we have launched a program to analyze and improve stress-inducible promoters in cereals so that we can (1) reveal the structural and functional components of these promoters, (2) design and construct better versions of these promoters, and (3) apply these promoters in the production of stress-resistant transgenic plants.

Analysis of stress-/abscisic acid (ABA)-inducible promoters

It has long been recognized that the phytohormone abscisic acid (ABA) is an important secondary messenger for many responses to abiotic stresses (Quatrano et al 1997). As the initial effort in our research program, we have chosen to study stress/ABA promoters in cereals such as rice and barley. Analyzing several stress-/ABA-induced genes encoding LEA (late embryogenesis abundant) proteins, by both loss- and gain-of-function approaches, we have concluded that, in order to be up-regulated by ABA, a promoter needs to have two *cis*-acting sequences (Fig. 1). One of them is an element containing an ACGT core with a proper flanking sequence, also known as an ABA response element (ABRE) (Shen et al 1996, 2004, Shen and Ho 1995). The other sequence is either another copy of the ACGT element or a coupling element (CE) (Shen et al 1996). One type of coupling element, CE3, is an apparent variant form of the ACGT element with a GCGT in the middle (a possible A to G mutation). The other type of coupling element, CE1, is totally different from the ACGT element in terms of sequence similarity. At least the ACGT element and CE3 interact with a special subclass of the basic leucine zipper transcription factor, ABI5, which is also regulated by phosphorylation (Casaretto and Ho 2003). The relationship between the two elements in the stress/ABA response complex is quite intriguing. The distance between the ACGT element and CE3 has to be very close in order to confer a high level of activity (Fig. 2). However, the distance between the ACGT element and CE1 (note that the CE1 sequence is quite different from that of the ACGE element) has to be an integral of 10, that is, 10, 20, 30, etc., to have the highest activity (Fig. 2). It is conceivable that the ACGT element and CE1 have to be on the same side of the DNA double helix (10 base pairs per turn) to have the proper interactions with each other. These lines of information about stress-/ABA-inducible promoters are not just

ACGT boxes and coupling elements in ABA-responsive promoters

Plant	Gene	ACGT box	CE/ACGT box	
Barley	*HVA1*	CCTACGTGGC	AACGCGTGTC	A + C (CE3)
Rice	*Osem*	CGTACGTGGT	GACGCGTGTC	A + C
Rice	*rab16b*	AGTACGTGGC	GCCGCGTGGC	A + C
Maize	*rab28*	GCCACGTGGG	GACGCGCCTC	A + C
Wheat	*Em*	GACACGTGGC	CACACGTGCC	A + A
Maize	*rab17*	GAGACGTGGC	CACACGTCCC	A + A
Barley	*HVA22*	GCCACGTACA	TGCCACCGGC	A + C′ (CE1)
Maize	*rab17*	CCCACGTACA	GGCCACCGAC	A + C′
Arabidopsis	*Rd29*			A + DRE

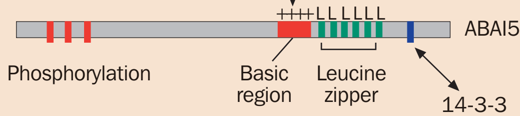


Fig. 1. Two essential *cis*-acting elements in stress-/ABA-responsive promoters. The ACGA element is needed for all stress-/ABA-responsive promoters. The other element is either another copy of the ACGT element or a coupling element (CE). It has been shown that a special class of the basic leucine-zipper transcription factor, ABI5, can specifically interact with the ACGT element as well as CE3.

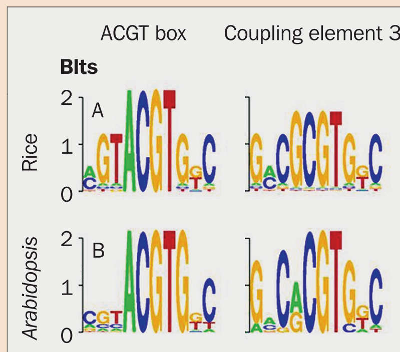


Fig. 2. Bioinformatics analysis of all stress-/ABA-responsive promoters in rice and *Arabidopsis*. From Zhang et al (2005).

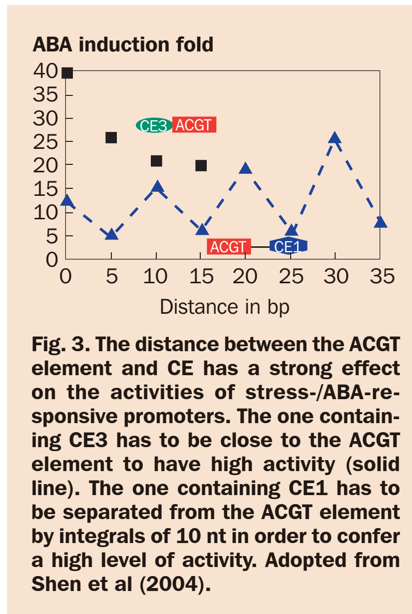
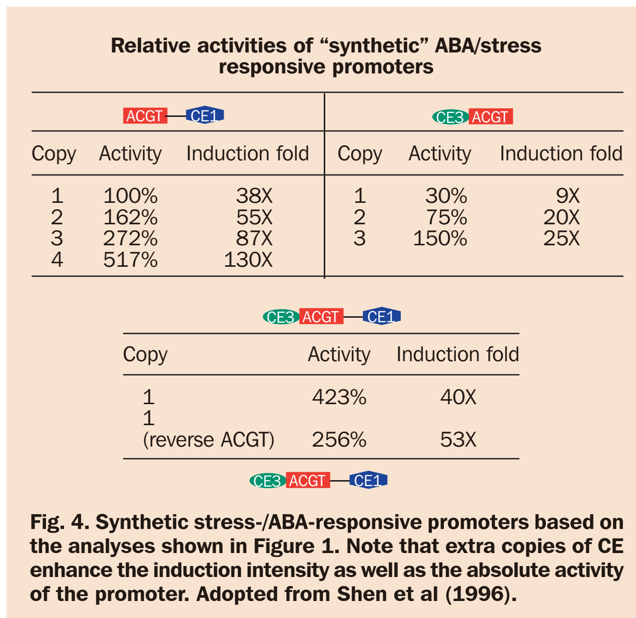


Fig. 3. The distance between the ACGT element and CE has a strong effect on the activities of stress-/ABA-responsive promoters. The one containing CE3 has to be close to the ACGT element to have high activity (solid line). The one containing CE1 has to be separated from the ACGT element by integrals of 10 nt in order to confer a high level of activity. Adopted from Shen et al (2004).

limited to a few genes that have been studied experimentally. Recent bioinformatics analysis suggests that promoters of virtually all ABA up-regulated genes in both rice and *Arabidopsis* follow the same structural "rules" that have been unraveled by our previous analyses (Fig. 3 and Zhang et al 2005). Based on the information obtained, we have been able to design and construct quite a few synthetic promoters that have even higher stress/ABA inducibility than naturally occurring promoters. As shown in Figure 4, by simply adding extra copies of the ACGT element and/or CE, the performance of these promoters can be enhanced by a large margin. Basically, the synthetic promoters can be much more tightly regulated by stress/ABA than the native ones. Moreover, orientations and relative positions of these elements could be further manipulated.

Minimal promoter/intron enhancements and development of tissue-specific stress-/ABA-inducible promoters

The basic components of a promoter can also be enhanced to achieve higher levels of activity. Besides elements involved in stress/ABA induction, minimal promoters and introns are also surprisingly important in regulating the level of expression. We have observed that minimal promoters isolated from some seed protein genes can be even better than the commonly used 35S, actin, and ubiquitin gene minimal promoters. Furthermore, although it is generally recognized that the first intron of a gene, or gene construct, possesses a general enhancer function, an intron could also confer specificity in gene expression. For example, the rice actin intron appears to be a general



Relative activities of "synthetic" ABA/stress responsive promoters

ACGT—CE1			CE3 ACGT		
Copy	Activity	Induction fold	Copy	Activity	Induction fold
1	100%	38X	1	30%	9X
2	162%	55X	2	75%	20X
3	272%	87X	3	150%	25X
4	517%	130X			

CE3 ACGT—CE1

Copy	Activity	Induction fold
1	423%	40X
1 (reverse ACGT)	256%	53X

Fig. 4. Synthetic stress-/ABA-responsive promoters based on the analyses shown in Figure 1. Note that extra copies of CE enhance the induction intensity as well as the absolute activity of the promoter. Adopted from Shen et al (1996).

enhancer, but barley alpha-amylase and LEA introns are important only in specific tissues in the context with certain hormone-regulated promoters (Y. Yamauchi and T.H.D. Ho, unpublished observations). In addition to stress/ABA inducibility provided by the ACGT element and CE and the absolute level of activities determined by the minimal promoter and intron, it would be desirable to incorporate tissue specificity determinants into the synthetic promoters. Although many genes in plants are differentially expressed in various tissues, well-defined promoter elements for tissue specificity are still very rare.

We followed a gene expression profiling approach using microarray analysis as a means to identify genes that are stress-/ABA-induced only in specific tissues/cell types. Rice plants were treated with ABA for different lengths of time and RNA samples were isolated from roots and shoots separately. These RNA samples were analyzed on Agilent rice microarray chips to identify genes that were induced by ABA. By comparing the expression profiles between root and shoot samples, we then identified the subsets of genes that were induced by ABA only in roots or shoots. To our surprise, the profiles of stress-/ABA-induced genes are quite different between these two tissues, with roots having many more genes up-regulated by ABA than shoots. On the other hand, more genes are down-regulated by ABA in shoots than in roots. Several biochemical and physiological processes are differentially regulated by ABA in these two organs. For example, genes involved in fatty acid synthesis are up-regulated by ABA in roots but not in shoots. The degree of differential stress/ABA responses in these genes has been further verified and more precisely quantified by northern (RNA) blot and real-time RT-PCR analyses as shown in Figure 5. Some of these genes are very

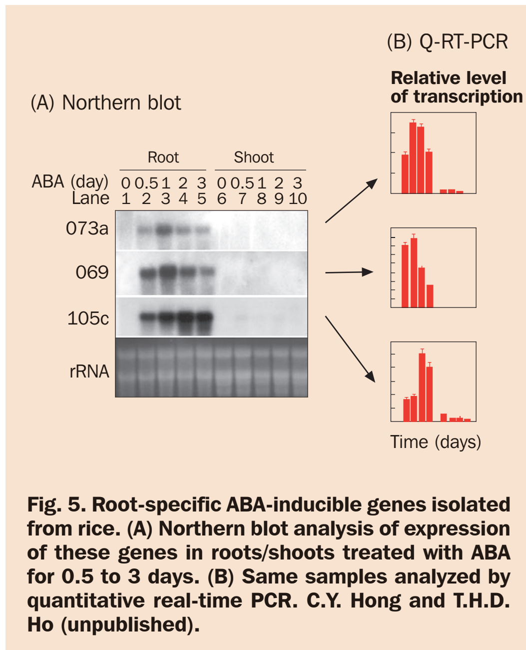


Fig. 5. Root-specific ABA-inducible genes isolated from rice. (A) Northern blot analysis of expression of these genes in roots/shoots treated with ABA for 0.5 to 3 days. (B) Same samples analyzed by quantitative real-time PCR. C.Y. Hong and T.H.D. Ho (unpublished).

tightly regulated by ABA and have a high degree of tissue specificity. For example, gene 069 is not expressed at all in roots and shoots in the absence of ABA. However, it is quickly induced by more than several thousand-fold by ABA in roots, but remains totally inactive in shoots (Fig. 5). If the level of its RNA is strictly regulated at the transcriptional level, as most of the genes are, its promoter must contain strong tissue specificity determinant(s) in addition to stress/ABA elements that we have already identified.

We carried out more analysis with the rice stress-/ABA-inducible genes, and found several of them not expressed in both mature and germinating seeds. As shown in Figure 6, gene 073 is highly induced by ABA in roots, but its level of expression is very low in shoots and in seeds. This additional tissue specificity is highly desirable because seeds are usually consumed by humans/animals; thus, it is an additional food safety benefit not having the transgenes expressed in seeds. Besides, most seeds are intrinsically stress resistant; thus, there is probably no need to provide additional protection against stress in seed tissues.

Applying stress-/ABA-inducible promoters to generate transgenic tomato without an apparent yield penalty

Could the use of stress-/ABA-inducible promoters be beneficial in regulating the expression of a transgene only when and where it is needed? We have tested the use of one of the synthetic stress/ABA promoters in the generation of transgenic tomato

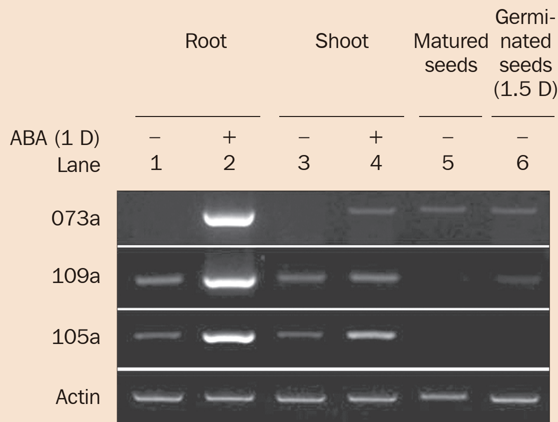


Fig. 6. Tissue-specific ABA-responsive rice genes. These genes are up-regulated by ABA in roots with little or no activity in shoots and seeds. C.Y. Hong and T.H.D. Ho (unpublished).

overexpressing a transcription factor, CBF1. CBF1 is one of the transcription factors regulating many downstream stress response processes/genes in *Arabidopsis* and other plants (Jaglo et al 2001). It has been demonstrated that using the 35S promoter to drive the expression of CBF1 in transgenic tomato leads to resistance to stresses such as drought (Hsieh et al 2002). However, constitutive expression of CBF1 also causes severe dwarfism in transgenic tomato plants. As shown in Figure 7, tomato plants harboring the 35S::CBF1 transgene have a mature height only about half that of the nontransgenic wild type. However, by replacing the 35S promoter with a stress-/ABA-inducible promoter, the CBF1 transgenic tomato plants still have the same level of enhanced stress resistance but also maintain the same height as the wild type. More importantly, the amount of tomato fruit produced in the 35S::CBF1 plants is much lower than that in the wild type, but this yield reduction does not take place in plants with the stress-inducible promoter (Fig. 8). Thus, by simply using a stress-inducible promoter, instead of a strong constitutive promoter, a CBF1 transgenic plant can have the same intended benefit but without the detrimental secondary effects.

Future prospects

We have demonstrated that the use of a stress-/ABA-inducible promoter to drive the expression of a beneficial transcription factor, CBF1, leads to high levels of drought and cold stress resistance in tomato plants without an apparent yield penalty. Furthermore, we have engineered the promoters so that they are (1) more tightly regulated by stress, that is, they have a very low background in the absence of stress, but are highly induced under stress; and (2) specifically induced by stress in roots or shoots

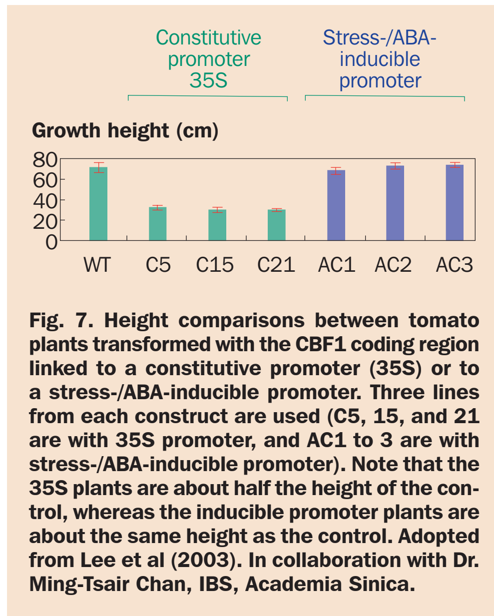


Fig. 7. Height comparisons between tomato plants transformed with the CBF1 coding region linked to a constitutive promoter (35S) or to a stress-/ABA-inducible promoter. Three lines from each construct are used (C5, 15, and 21 are with 35S promoter, and AC1 to 3 are with stress-/ABA-inducible promoter). Note that the 35S plants are about half the height of the control, whereas the inducible promoter plants are about the same height as the control. Adopted from Lee et al (2003). In collaboration with Dr. Ming-Tsair Chan, IBS, Academia Sinica.

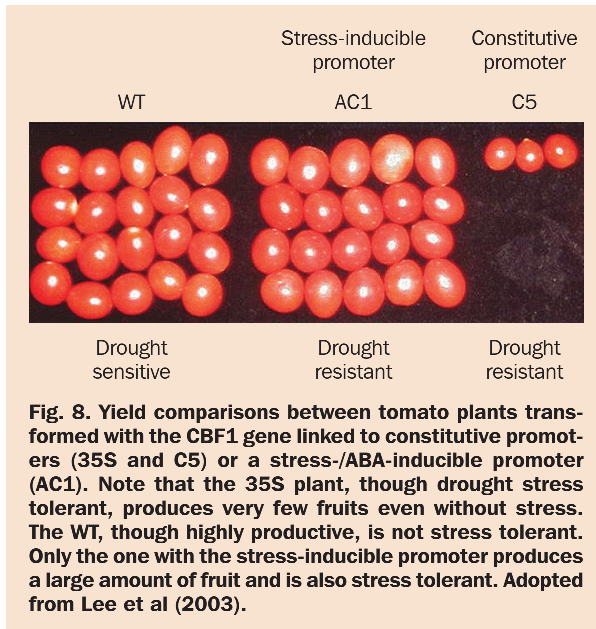


Fig. 8. Yield comparisons between tomato plants transformed with the CBF1 gene linked to constitutive promoters (35S and C5) or a stress-/ABA-inducible promoter (AC1). Note that the 35S plant, though drought stress tolerant, produces very few fruits even without stress. The WT, though highly productive, is not stress tolerant. Only the one with the stress-inducible promoter produces a large amount of fruit and is also stress tolerant. Adopted from Lee et al (2003).

of rice plants but with very low or no activities in seeds. A "tool box" of these types of designer promoters has been generated so that researchers could choose the most appropriate "molecular switch" to properly control the expression of their favored transgene. By effective regulation of transgene expression, the performance of transgenic plants could then be optimized with minimal undesirable side effects, hence achieving the ultimate goal of genetic engineering.

References

Casaretto J, Ho T-HD. 2003. The transcription factors HvABI5 and HvVP1 are required for the ABA induction of gene expression in barley (*Hordeum vulgare* L.). Plant Cell 15:271-284.

Hsieh TH, Lee JT, Charng YY, Chan MT. 2002. Tomato plants ectopically expressing *Arabidopsis* CBF1 show enhanced resistance to water deficit stress. Plant Physiol. 130:618-626.

Ho T-HD, Wu R. 2004. Genetic engineering for enhancing plant productivity and stress tolerance. In: Nguyen H, Blum A, editors. Physiology and biotechnology integration for plant breeding. New York, N.Y. (USA): Marcel Dekker. p 489-502.

Jaglo, KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF. 2001. Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. Plant Physiol. 127:910-917.

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. 1999. Improving plant drought, salt, and freezing tolerance by gene transfer of single stress-inducible transcription factor. Nature Biotechnol. 17:287-291.

Lee J-T, Prasad V, Yang P-T, Wu J-F, Ho T-HD, Charng Y-Y, Chan, M-T. 2003. Expression of *Arabidopsis* CBF1 regulated by an ABA/stress inducible promoter in transgenic tomato confers stress tolerance without affecting yield. Plant Cell Environ. 26:1181-1190.

Quatrano R, Bartel D, Ho THD, Pages M. 1997. New insights into ABA-mediated process. Plant Cell 8:470-475.

Shen Q, Ho THD. 1995. Functional dissection of an ABA-inducible gene reveals two independent ABA responsive complexes each containing a G-box and a novel *cis*-acting element. Plant Cell 7:295-307.

Shen Q, Zhang P, Ho THD. 1996. Modular nature of ABA response complexes: composite promoter units that are necessary and sufficient for ABA-induced gene expression in barley. Plant Cell 7:1107-1119.

Shen JQ, Casaretto J, Zhang P, Ho THD. 2004. Functional definition of ABA response complexes: the promoter units necessary and sufficient for ABA induction of gene expression in barley (*Hordeum vulgare* L.). Plant Mol. Biol. 54:111-124.

Zhang W, Ruan J, Ho THD, You Y, Yu T, Quatrano R. 2005. *Cis*-regulatory element based targeted gene finding: genome-wide identification of ABA- and abiotic stress-responsive genes in *Arabidopsis thaliana*. Bioinformatics 21:3074-3081.

Notes

Authors' addresses: Tuan-hua David Ho and Chwan-Yang Hong, Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan; Ming-Tsair Chan, Institute of BioAgricultural Sciences, Academia Sinica, Taipei, Taiwan; Sumay Yu, Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan.

Acknowledgments: The primary work performed in the authors' laboratories was supported by grants from the National Science Council, Academia Sinica, and National Science and Technology Program on Agricultural Biotechnology in Taiwan.

Rice: an emerging model for plant systems biology

A. von Zychlinski, S. Baginsky, and W. Gruissem

Proteomics has become a powerful technique to investigate cellular processes and network functions. This became possible as a result of major progress in the sensitivity of mass spectrometry instrumentation and data analysis software. As proteomics technologies are now becoming available to the wider scientific community, efforts are under way to identify complete proteomes. This information is used to improve genome annotation and to identify and confirm protein splice variants. Analysis of protein modifications and protein variants uses novel scoring and prediction tools independent of established protein databases. We discuss the proteomics tools and analysis pipelines that can be applied to rice in order to facilitate our understanding of rice genome structure and function.

Keywords: Rice, proteomics, transcriptomics, metabolomics, TILING arrays, high throughput

As one of the most important crop plants worldwide, rice (*Oryza sativa* L.) is the model organism of choice for research to understand mechanisms of function and performance. Complete high-quality genome sequence information on the two rice subspecies, japonica and indica, allows the molecular analysis of genetic differences. The available rice sequence information also facilitated the development of a commercially available GeneChip® with a unique array design to investigate the transcriptional activity of most gene models. Increased complexities of the rice proteome and metabolome continue to pose significant challenges, however, because no single standardized procedure is available for the analysis of the proteome or metabolome (reviewed in Aebersold and Mann 2003, Bino et al 2004). We discuss various technologies that are being developed and efforts that are under way for rice to integrate transcriptomics, proteomics, and metabolomics into a systems-level approach to understanding cellular processes.

Full-genome expression analysis tools

Based on the available DNA sequence, the complete rice genome has been assembled on TILING arrays. TILING arrays are high-density microarrays that cover the entire genome, including regions that are not predicted to contain protein-coding genes. This approach is useful for several reasons. First, TILING arrays will provide information on the coding capacity of intergenic regions and identify transcription units that have not been annotated as gene models. Second, annotated gene models can be confirmed and refined, including the identification of natural antisense transcripts, because expression can be monitored simultaneously for both DNA strands. Third, TILING arrays will be helpful for the discovery and characterization of alternative RNA splicing, which allows a single gene to encode multiple functionally distinct proteins.

Two recent studies for *Arabidopsis* using full-genome TILING arrays on RNA isolated from cell-culture lines have already provided a number of surprising results (Yamada et al 2003, Stolc et al 2005). It can be expected that transcriptome analysis using TILING arrays for rice will produce equally exciting new data. For example, both *Arabidopsis* studies have detected high levels of antisense transcription that in some cases exceeded the transcription activity of the corresponding sense strand of gene models. The discovery of transcription activity in intergenic regions suggests that the transcriptional capacity of the *Arabidopsis* genome exceeds the current estimates based on genome annotation (Yamada et al 2003, Stolc et al 2005). This raises the question of to what extent antisense transcription of gene models regulates the translation rate of the corresponding mRNA *in vivo*. Furthermore, are transcripts from "nonannotated" intergenic regions actually translated into proteins, or are they simply a consequence of transcriptional read-through? Answers to these questions require global and deep analysis of all expressed proteins. Transcriptome and proteome analyses combined will provide the necessary and complementary information about genome structure, activity, and regulation to support or contradict results from TILING array experiments. Furthermore, proteomics approaches will reveal posttranslational protein modifications and provide new insights into subcellular localization of proteins.

Quantitative proteome analysis

Proteomics has a central role in systems biology because it integrates information derived from the other "-omics" technologies, including transcriptomics and metabolomics. Transcriptome analysis is a powerful first step in predicting proteome complexity based on temporal and spatial gene expression patterns, but alone it does not provide insights into subcellular protein localization, pathway compartmentalization, or protein concentrations. A positive correlation between transcript level and protein abundance suggests transcriptional control of protein expression. Although several studies have reported such a positive correlation (Gygi et al 1999, Ideker et al 2001, Griffin et al 2002, Washburn et al 2003, Kleffmann et al 2004, Tian et al 2004, reviewed in Greenbaum et al 2003, Hack 2004), not all mRNAs correlate in their abundance with the accumulation of their corresponding proteins. Additional regula-

tory mechanisms that control mRNA translation rate or protein stability can account for such differences. Recent evidence suggests that antisense-RNA transcription may also contribute significantly to the final concentration of a protein (Yamada et al 2003, Stolc et al 2005). These few examples demonstrate that precise quantitative information at the transcriptome, proteome, and metabolome level is required to understand cellular network functions and regulation.

Data management and mining tools

The Systems Biology Experiment Analysis Management System (SBEAMS, www.sbeams.org) was developed at the Institute of Systems Biology in Seattle as an integrative platform for most if not all "-omics"-type data. SBEAMS is a relational database management system for systems biology experiments—an integrated software framework with a diverse set of tools to store, manage, and query experimental data. It contains a Web interface for querying the database and provides integrated access to all available data. All data analysis tasks are greatly simplified to ensure quality control. SBEAMs is currently being installed at a few other locations worldwide, including the Functional Genomics Center Zurich (FGCZ, www.fgcz.ethz.ch), to support the management of high-throughput proteomics data. Data-mining tools such as Genevestigator® (https://www.genevestigator.ethz.ch), which was developed at ETH Zurich (Zimmermann et al 2004, 2005), currently provide access to large databases of gene expression experiments. The data-mining platform is designed such that quantitative proteome and metabolome data can be easily integrated in the future for systematic exploration of cellular networks. New model organisms such as rice can be easily incorporated into the Genevestigator® software tool box provided that rice researchers agree on a single microarray platform (e.g., Affymetrix, NimbleGen) as the community standard.

PeptideAtlas was developed as an open-source platform and database to which researchers can submit their MS/MS data to make them publicly available for genome annotation. Such a data collection from different sources is a very efficient strategy for building up a "peptide-validated" genome. At the current rate of data collection, proteome coverage increases linearly with the number of submitted MS/MS data (Desiere et al 2005), suggesting that proteome analysis can provide significant new insights into genome structure and expression. High-throughput proteome sequencing projects are currently under way for human, *Drosophila*, *C. elegans*, and *Arabidopsis* (see Center for Model Organisms Proteomics, www.systemsx.ch/about/nodes/cmop.html). The rate of discovering new peptides will slow as more data are deposited into PeptideAtlas. This stage has not been reached, however, and high-throughput MS/MS data collection is at present still the most efficient strategy for increasing proteome coverage. As is the case for *Arabidopsis*, high-throughput proteome analysis in rice will also contribute significantly to the refinement of the rice genome structure and expression. Integration of tools such as SBEAMS, Genevestigator®, and PeptideAtlas will therefore greatly accelerate our understanding of plant genome function and regulation. The available proteomics tools and their utility in rice systems analysis

through the integration of functional genomics approaches are discussed in the next section.

Exploiting proteomics for systems biology: development of high-throughput analysis pipelines

High-throughput proteomics techniques focus on the identification of the complete proteome that represents the expressed genome in specific cells, organs, or organisms (reviewed in Aebersold and Mann 2003, Yates 2004). Although significant advances have been made, we should also mention that high-throughput proteomics at present has two limitations. First, because of the dynamic range limitations of current MS/MS instruments, even sophisticated biochemical separation strategies allow the identification of only relatively abundant proteins in a complex protein mixture (McGregor and Dunn 2003). Techniques are currently being developed to enrich low-abundance proteins. Second, high-throughput proteome analysis is mostly restricted to proteins and peptides available in protein databases (e.g., www.expasy.ch) or to organisms with fully sequenced genomes or extensive EST collections. Low-abundance proteins, proteins translated from unpredicted alternatively spliced mRNAs, or peptides with posttranslational modifications will not be routinely detected and require *de novo* sequencing strategies, whose discussion is beyond the scope of this article.

As introduced above, high-throughput proteomics data can significantly improve genome annotation and gene models by mapping identified and experimentally verified peptides to the genome. Using such high-throughput proteomics data, Desiere and colleagues (2005) have reported the unambiguous detection of several SNPs and confirmation of splice junctions. In contrast to EST and cDNA information, peptide sequences provide unequivocal information and posttranslational modifications can be included directly in the genome annotation. Using appropriate cellular fractionation techniques, organelle proteome sequencing strategies provide further insights into subcellular protein localization and pathway compartmentalization. Our own recent study of highly purified *Arabidopsis* chloroplasts (Kleffmann et al 2004) and rice etioplast (von Zychlinski et al 2005) has shown that computational prediction of protein localization fails to assign all proteins correctly, and therefore protein localization must be verified by additional experiments. Other recent large-scale proteome analyses of *Arabidopsis* chloroplasts (Friso et al 2004), mitochondria (Heazlewood et al 2004), and vacuoles (Carter et al 2004) have confirmed our observations.

To facilitate proteome data analysis, the Trans-Proteomic Pipeline (TPP) was developed at the Institute for Systems Biology (Seattle, Washington; www.proteome-center.org/software.php) as an open-source analysis platform for unified storage and analysis of MS/MS data. In TPP, MS/MS data are first automatically converted into XML file formats, which can be used by various database search engines. Tools such as PeptideProphet and ProteinProphet are then applied to the data for the identification of peptides and proteins. MS/MS spectra that cannot be assigned are subsequently analyzed using various quality-scoring tools to identify high-quality spectra that can be used for modified database searches or *de novo* sequencing approaches. This

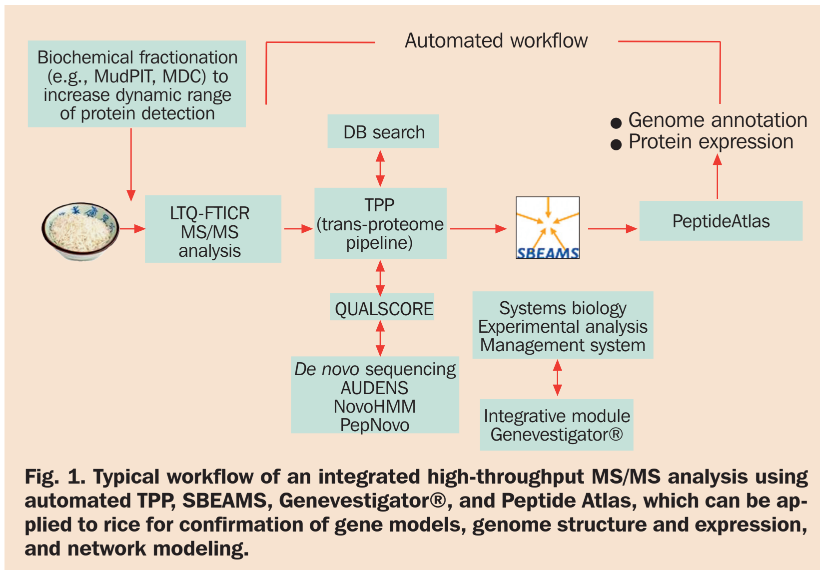


Fig. 1. Typical workflow of an integrated high-throughput MS/MS analysis using automated TPP, SBEAMS, Genevestigator®, and Peptide Atlas, which can be applied to rice for confirmation of gene models, genome structure and expression, and network modeling.

can be accomplished with the help of tools such as QUALSCORE (Nesvizhskii et al 2006) and PepNovo (Frank and Pevzner 2005), which have been implemented as integral components of TPP. Figure 1 illustrates a modified proteomics workflow, in which MS/MS data are first subjected to a standard protein database search that provides protein identifications based on existing and annotated databases. The remaining MS/MS spectra that were not assigned using standard protein databases are then evaluated using database-independent MS/MS spectrum-scoring tools such as QUALSCORE to identify true peptide-derived spectra from low-quality noise fragmentation or contaminants. MS/MS spectra identified by QUALSCORE can then be searched against modified databases, such as a genome database or protein database including different posttranslational modifications. In a parallel approach, MS/MS spectra can also be analyzed directly in a fully database-independent way using *de novo* sequencing tools. Several prediction tools that are now available achieve good results when applied to high-quality MS/MS spectra (Chen et al 2001, Johnson and Taylor 2002, Ma et al 2003, Zhang 2004, Searle et al 2004, Fischer et al 2005, Frank and Pevzner 2005, Grossmann et al 2005).

In the example shown in Figure 1, a complex rice protein mixture of interest is first separated into different fractions before MS/MS analysis in order to increase the dynamic range of protein detection. Commonly used techniques are MudPIT or multidimensional chromatography, either off-line or directly on-line on an MS/MS instrument. The following automated workflow ideally uses a high-resolution, fast-scanning mass spectrometer (e.g., LTQ-FTICR) to generate MS/MS spectra, which are then processed in TPP. The MS/MS spectra of peptides that correspond to entries in protein databases result in direct rice protein identifications. The remaining set

of MS/MS spectra is searched using QUALSCORE to extract high-quality MS/MS spectra that most likely originate from rice peptides. Following this filtering process, modified database searches are performed, which again result in additional rice protein identifications and unassigned MS/MS spectra that are further processed to produce amino acid sequence information using *de novo* sequencing tools. Subsequent rice peptide data analysis can be performed in SBEAMS and peptide data stored in PeptideAtlas. Automated mapping of identified peptides to the rice genome facilitates genome annotation and confirmation of gene models. If combined with data-mining tools such as Genevestigator®, a unified rice functional genomics data-processing approach could greatly facilitate experimental analysis to obtain information from single genes to pathway functions and genetic networks.

Current status of plant proteomics: shifting the focus to rice

Most of the large-scale proteome analyses from rice and other plants reported to date used isolated organelles, membrane fractions, or subcellular structures (reviewed in Peck 2005). These studies include different plastid types (Peltier et al 2002, Froehlich et al 2003, Ferro et al 2003, Huber et al 2004, Peltier et al 2004, Kleffmann et al 2004, Friso et al 2004, Baginsky et al 2004, reviewed in Baginsky and Gruissem 2004, van Wijk 2004, von Zychlinski et al 2005), mitochondria (Brugiere et al 2004, Lister et al 2004, Heazlewood et al 2004, Millar et al 2005), peroxisomes (Fukao et al 2002), vacuoles (Shimaoka et al 2004, Carter et al 2004), the plasma membrane (Ephritikhine et al 2004, Alexandersson et al 2004, Sazuka et al 2004, Marmagne et al 2004, Borner et al 2005), the cell wall (Chivasa et al 2002, Borderies et al 2003, Boudart et al 2005), and cytosolic ribosomes (Chang et al 2005). Table 1 summarizes the reports that focus on the rice proteome. To date, only two large-scale proteome analyses have been reported from whole rice plants or cultured rice cells. Koller et al (2002) identified 2,528 unique proteins using a combined MudPIT and 2D-gel electrophoresis approach. Komatsu (2005) built a rice proteome database from 2D-gel electrophoresis maps. To date, this database features 23 protein maps from different rice tissues, resulting in 5,092 highly redundant protein entries.

Fractionation of cell organelles and subcellular compartments before proteome analysis provides useful insights into protein localization, protein sorting into different cell compartments, and compartmentalization of biochemical pathways (Taylor et al 2003). All organelle proteomics studies reported to date identified proteins that were not predicted to localize to the respective organelle when computational prediction tools such as TargetP were used as a benchmark. This is not surprising, however, because the original training set of proteins used for TargetP showed only 85% accuracy, suggesting that intracellular protein trafficking is more complex and unpredictable than anticipated and that unexpected pathways for protein import might exist.

Unequivocal identification and assignment of proteins to specific subcellular locations therefore require stringent purification protocols. The rapid development of high-resolution and high-accuracy mass spectrometers in the future will lead to the detection of even minor contaminations of proteins from other cellular organelles,

Table 1. Summary of reported rice proteome studies.

Tissue	Method	Number of identified proteins	Reference
Leaf, root, and seed	2D-PAGE MS/MS	556	Koller et al (2002)
	MudPIT	2,363	
Mitochondria	2D-PAGE MS/MS	136	Heazlewood et al (2003)
	Blue-native PAGE MS/MS		
	LC MS/MS		
Nuclear proteins, rice seedlings	2D-PAGE MS/MS	257	Khan and Komatsu (2004)
	2D-PAGE Edman sequencing		
Rice seedlings	2D-PAGE MS/MS	5,092 redundant protein entries	Komatsu (2005) (Rice proteome database)
Cultured suspension cells	2D-PAGE Edman sequencing		
Leaf sheath	2D-PAGE MS/MS2	44	Shen et al (2002)
	D-PAGE Edman sequencing	58	
Plasma membranes	2D-PAGE MS/MS		Tanaka et al (2004)
Vacuolar membranes	2D-PAGE Edman sequencing	43	
Golgi membranes		46	
Mitochondria		146	
Chloroplasts		89	
Etioplasts	LC MS/MS	240	von Zychlinski et al (2005)

which could result in misinterpretation of the proteome data (Warnock et al 2004). As a consequence, it is important to develop criteria that assess the purity of subcellular proteome fractions. In addition, experiments are necessary to validate the subcellular targeting of identified proteins. Proteomics data provide an excellent starting point for the design of such experiments. Using rice etioplast as an example, we developed an extensive protein profiling approach to validate the purity of the isolated organelle and to facilitate high-confidence assignment of proteins to the rice organelle (von Zychlinski et al 2005).

Finally, complete rice genome annotation and confirmation of gene models require that rice proteome analysis reaches a sufficient depth to identify and validate low-abundance proteins. At present, the reported rice proteome studies (Table 1) do not provide sufficiently deep information because they relied on established techniques and available protein databases. Novel experimental and computational tools as described in this chapter, including modified analysis pipelines, can now be included

in the rice proteome workflow to increase proteome coverage and to facilitate our systems understanding of rice function and performance.

References

Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. Nature 422:198-207.

Alexandersson E, Saalbach G, Larsson C, Kjellbom P. 2004. *Arabidopsis* plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. Plant Cell Physiol. 45:1543-1556.

Baginsky S, Gruissem W. 2004. Chloroplast proteomics: potentials and challenges. J. Exp. Bot. 55:1213-1220.

Baginsky S, Siddique A, Gruissem W. 2004. Proteome analysis of tobacco bright yellow-2 (BY-2) cell culture plastids as a model for undifferentiated heterotrophic plastids. J. Proteome Res. 3:1128-1137.

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, et al. 2004. Potential of metabolomics as a functional genomics tool. Trends Plant Sci. 9:418-425.

Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerre-Tugaye MT, Boudet A, Pont-Lezica R. 2003. Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: a critical analysis. Electrophoresis 24:3421-3432.

Borner GH, Sherrier DJ, Weimar T, Michaelson LV, Hawkins ND, Macaskill A, Napier JA, Beale MH, Lilley KS, Dupree P. 2005. Analysis of detergent-resistant membranes in *Arabidopsis*: evidence for plasma membrane lipid rafts. Plant Physiol. 137:104-116.

Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerre-Tugaye MT, Pont-Lezica R. 2005. Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: identification by mass spectrometry and bioinformatics. Proteomics 5:212-221.

Brugiere S, Kowalski S, Ferro M, Seigneurin-Berny D, Miras S, Salvi D, Ravanel S, D'Herin P, Garin J, Bourguignon J, Joyard J, Rolland N. 2004. The hydrophobic proteome of mitochondrial membranes from *Arabidopsis* cell suspensions. Phytochemistry 65:1693-1707.

Carter C, Pan S, Zouhar J, Avila EL, Girke T, Raikhel NV. 2004. The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. Plant Cell 16:3285-3303.

Chang IF, Szick-Miranda K, Pan S, Bailey-Serres J. 2005. Proteomic characterization of evolutionarily conserved and variable proteins of *Arabidopsis* cytosolic ribosomes. Plant Physiol. 137:848-862.

Chen T, Kao MY, Tepel M, Rush J, Church GM. 2001. A dynamic programming approach to *de novo* peptide sequencing via tandem mass spectrometry. J. Comput. Biol. 8:325-337.

Chivasa S, Ndimba BK, Simon WJ, Robertson D, Yu XL, Knox JP, Bolwell P, Slabas AR. 2002. Proteomic analysis of the *Arabidopsis thaliana* cell wall. Electrophoresis 23:1754-1765.

Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, Aderem A, Boyle R, Brunner E, Donohoe S, et al. 2005. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. Genome Biol. 6:R9.

Ephritikhine G, Ferro M, Rolland N. 2004. Plant membrane proteomics. Plant Physiol. Biochem. 42:943-962.

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N. 2003. Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol. Cell Proteomics 2:325-345.

Fischer B, Roth V, Roos F, Grossmann J, Baginsky S, Widmayer P, Gruissem W, Buhmann JM. 2005. NovoHMM: a hidden Markov model for de novo peptide sequencing. Anal. Chem. 77:7265-7273.

Frank A, Pevzner P. 2005. PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal. Chem. 77:964-973.

Friso G, Giacomelli L, Ytterberg AJ, Peltier JB, Rudella A, Sun Q, Wijk KJ. 2004. In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. Plant Cell 16:478-499.

Froehlich JE, Wilkerson CG, Ray WK, McAndrew RS, Osteryoung KW, Gage DA, Phinney BS. 2003. Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J. Proteome Res. 2:413-425.

Fukao Y, Hayashi M, Nishimura M. 2002. Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of *Arabidopsis thaliana*. Plant Cell Physiol. 43:689-696.

Greenbaum D, Colangelo C, Williams K, Gerstein M. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. 4:117.

Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. 2002. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. Mol. Cell Proteomics 1:323-333.

Grossmann J, Roos FF, Cieliebak M, Liptak Z, Mathis LK, Muller M, Gruissem W, Baginsky S. 2005. AUDENS: a tool for automated peptide de novo sequencing. J. Proteome Res. 4:1768-1774.

Gygi SP, Rochon Y, Franza BR, Aebersold R. 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell Biol. 19:1720-1730.

Hack CJ. 2004. Integrated transcriptome and proteome data: the challenges ahead. Brief Funct. Genomic Proteomic 3:212-219.

Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. 2004. Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. Plant Cell 16:241-256.

Huber CG, Walcher W, Timperio AM, Troiani S, Porceddu A, Zolla L. 2004. Multidimensional proteomic analysis of photosynthetic membrane proteins by liquid extraction-ultracentrifugation-liquid chromatography-mass spectrometry. Proteomics 4:3909-3920.

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292:929-934.

Johnson RS, Taylor JA. 2002. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. Mol. Biotechnol. 22:301-315.

Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, Baginsky S. 2004. The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. Curr. Biol. 14:354-362.

Koller A, Washburn MP, Lange BM, Andon NL, Deciu C, Haynes PA, Hays L, Schieltz D, Ulaszek R, Wei J, Wolters D, Yates JR, 3rd. 2002. Proteomic survey of metabolic pathways in rice. Proc. Natl. Acad. Sci. USA 99:11969-11974.

Komatsu S. 2005. Rice proteome database: a step toward functional analysis of the rice genome. Plant Mol. Biol. 59:179-190.

Lister R, Chew O, Lee MN, Heazlewood JL, Clifton R, Parker KL, Millar AH, Whelan J. 2004. A transcriptomic and proteomic characterization of the *Arabidopsis* mitochondrial protein import apparatus and its response to mitochondrial dysfunction. Plant Physiol. 134:777-789.

Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G. 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun. Mass Spectrom. 17:2337-2342.

Marmagne A, Rouet MA, Ferro M, Rolland N, Alcon C, Joyard J, Garin J, Barbier-Brygoo H, Ephritikhine G. 2004. Identification of new intrinsic proteins in *Arabidopsis* plasma membrane proteome. Mol. Cell Proteomics 3:675-691.

McGregor E, Dunn MJ. 2003. Proteomics of heart disease. Hum. Mol. Genet. 12 Spec. No. 2:R135-144.

Millar AH, Heazlewood JL, Kristensen BK, Braun HP, Moller IM. 2005. The plant mitochondrial proteome. Trends Plant Sci. 10:36-43.

Nesvizhskii AI, Roos FF, Grossmann J, Vogelzang M, Eddes JS, Gruissem W, Baginsky S, Aebersold R. 2006. Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. Mol. Cell Proteomics. 5:652-670.

Peck SC. 2005. Update on proteomics in *Arabidopsis*: where do we go from here? Plant Physiol. 138:591-599.

Peltier JB, Ytterberg AJ, Sun Q, van Wijk KJ. 2004. New functions of the thylakoid membrane proteome of *Arabidopsis thaliana* revealed by a simple, fast and versatile fractionation strategy. J. Biol. Chem. 279:49367-49383.

Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Soderberg L, Roepstorff P, von Heijne G, van Wijk KJ. 2002. Central functions of the lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. Plant Cell 14:211-236.

Sazuka T, Keta S, Shiratake K, Yamaki S, Shibata D. 2004. A proteomic approach to identification of transmembrane proteins and membrane-anchored proteins of *Arabidopsis thaliana* by peptide sequencing. DNA Res. 11:101-113.

Searle BC, Dasari S, Turner M, Reddy AP, Choi D, Wilmarth PA, McCormack AL, David LL, Nagalla SR. 2004. High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results. Anal. Chem. 76:2220-2230.

Shimaoka T, Ohnishi M, Sazuka T, Mitsuhashi N, Hara-Nishimura I, Shimazaki K, Maeshima M, Yokota A, Tomizawa K, Mimura T. 2004. Isolation of intact vacuoles and proteomic analysis of tonoplast from suspension-cultured cells of *Arabidopsis thaliana*. Plant Cell Physiol. 45:672-683.

Stolc V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC, Hegeman A, Nelson C, Rancour D, Bednarek S, et al. 2005. Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays. Proc. Natl. Acad. Sci. USA 102:4453-4458.

Taylor SW, Fahy E, Ghosh SS. 2003. Global organellar proteomics. Trends Biotechnol. 21:82-88.

Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, et al. 2004. Integrated genomic and proteomic analyses of gene expression in mammalian cells. Mol. Cell Proteomics 3:960-969.
van Wijk KJ. 2004. Plastid proteomics. Plant Physiol. Biochem. 42:963-977.
von Zychlinski A, Kleffmann T, Krishnamurthy N, Sjolander K, Baginsky S, Gruissem W. 2005. Proteome analysis of the rice etioplast: metabolic and regulatory networks and novel protein functions. Mol. Cell Proteomics 4:1072-1084.
Warnock DE, Fahy E, Taylor SW. 2004. Identification of protein associations in organelles, using mass spectrometry-based proteomics. Mass Spectrom. Rev. 23:259-280.
Washburn MP, Koller A, Oshiro G, Ulaszek RR, Plouffe D, Deciu C, Winzeler E, Yates JR, 3rd. 2003. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 100:3107-3112.
Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, et al. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842-846.
Yates JR, 3rd. 2004. Mass spectral analysis in proteomics. Annu. Rev. Biophys. Biomol. Struct. 33:297-316.
Zhang Z. 2004. De novo peptide sequencing based on a divide-and-conquer algorithm and peptide tandem spectrum simulation. Anal. Chem. 76:6374-6383.
Zimmermann P, Hennig L, Gruissem W. 2005. Gene-expression analysis and network discovery using Genevestigator. Trends Plant Sci. 10:407-409.
Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. 2004. GENEVESTIGATOR: *Arabidopsis* microarray database and analysis toolbox. Plant Physiol. 136:2621-2632.

Notes

Authors' address: Institute of Plant Sciences, ETH Zurich, Universitätstrasse 2, 8092 Zurich, Switzerland, e-mail: wgruissem@ethz.ch.

AN OVERVIEW: FIVE INTERNATIONAL RICE GENETICS SYMPOSIA (1985-2005)

G.S. Khush

In his overview of the First International Rice Genetics Symposium (IRGS) held in May 1985, Sir Ralph Riley, world-renowned geneticist, began by saying, "Rice would not be the organism of choice for basic research. So, all who undertake its study do so with the objective of helping those who grow and eat rice."

I can safely say today that rice is now the model plant for genetic research on crop plants, and those who work on rice do so not only to help grow and eat it but also to advance the frontiers of genetics and molecular biology. In 1985, rice was a poor cousin to maize, wheat, barley, and tomato for genetic knowledge. There was no acceptable system of numbering rice chromosomes, no system of gene nomenclature existed, and different gene symbols were assigned to the same genes. There was no coordination in linkage mapping. There was no mechanism for maintaining mutant gene stocks. However, several developments during the last 20 years have made rice the organism of choice for research on crop plants, and it has become a reference genome.

First was the establishment of the Rice Genetics Cooperative (RGC) during the first International Rice Genetics Symposium (IRGS) for promoting cooperation and coordination among rice geneticists. Several committees were set up under the RGC to coordinate activities. The committee on rice gene symbolization and nomenclature prepared rules for gene symbolization and set up a mechanism to coordinate the assignment of symbols to newly identified rice genes. It also helped to establish a unified system of numbering rice chromosomes and linkage maps. The RGC also decided to publish a Rice Genetics Newsletter annually and this has proved to be an excellent medium for communication among rice scientists. Two centers, one at IRRI and the other at Kyushu University, were established to maintain and distribute seeds of rice mutants.

The second development was the establishment of the International Program on Rice Biotechnology by the Rockefeller Foundation in 1985. During its 15-year tenure, several laboratories in the United States, Europe, and Asia supported by this Rockefeller program made major advances in rice molecular and cellular biology. Considerable investment was made in human resource development, particularly in developing countries.

The third development was the start of the Rice Genome Research Program (RGRP) in Japan under the leadership of Dr. Takuji Sasaki in 1991. Densely populated molecular genetic maps of rice were prepared and numerous ESTs and BAC libraries were generated. This eventually led to the establishment of the International Rice Genome Sequencing Project (IRGSP) in 1999. Sequencing of japonica and indica genomes through international collaboration has been a major landmark in higher plant genetics and has opened new avenues for determining the functions of each of the 29,757 rice genes. The establishment of the International Network on Rice Functional Genomics by IRRI is a logical and timely development. Rice genes identified through these international efforts will be extremely useful for developing more productive and more nutritious varieties for feeding five billion rice consumers in 2030.

All these advances in rice genetics have been reflected in the scientific content of five successive IRGS. Two hundred rice scientists attended the First IRGS, held in 1985, and 90% of the papers were on classical or Mendelian genetics and a few on isozymes and tissue culture. The highlight of the symposium, as mentioned earlier, was the establishment of the RGC.

Three hundred scientists from 24 countries attended the Second IRGS (1990). Approximately 50% of the papers were on classical genetics and the remaining covered molecular genetics and biotechnology. A unified system of numbering rice chromosomes and linkage groups was established. The first molecular genetic map of rice and protocols for transformation of indica and japonica rice were reported. This set the stage for an upsurge of research on molecular and cellular biology of rice.

Five hundred scientists from 31 countries attended the Third IRGS (1995) and the vast majority of the papers were on molecular genetics. A dense molecular genetic map of more than 2,000 markers was reported. Orientation of classical and molecular genetic maps was another advance.

Five hundred twenty scientists from 32 countries attended the Fourth IRGS (2000). A majority of the papers covered molecular genetics, gene tagging, and transformation.

The fifth IRGS (2005) had 710 registered participants from 38 countries and the program consisted of 26 plenary papers, 54 oral presentations, 4 workshops, and 400 poster presentations. Excellent papers were presented in different sessions on breeding, mapping of genes and QTLs, identification and cloning of candidate genes for biotic and abiotic stresses, gene expression, and genomic databases. There were excellent reports on mutant induction for functional genomics. A meeting of the RGC was held on 21 November and new officers were selected. It was decided that the Rice Genetics Newsletter will be published online twice a year instead of having an annual printed version. Submission of manuscripts is now possible through the Oryzabase and Gramene Web sites. Considerable confusion has occurred in gene symbolization, particularly in naming annotated gene sequences. The RGC decided that Gramene and Oryzabase and other interested parties should formulate a mechanism for online submission and approval of new gene symbols.

I am extremely happy to see the participation of numerous young scientists in this symposium. I am confident that, during the Sixth IRGS in 2010, many of these

young scientists will be leaders in their respective fields of research. I hope there will be stronger communication between molecular biologists and plant breeders for using the exciting developments in molecular biology for developing more productive rice varieties for sustainable food security.