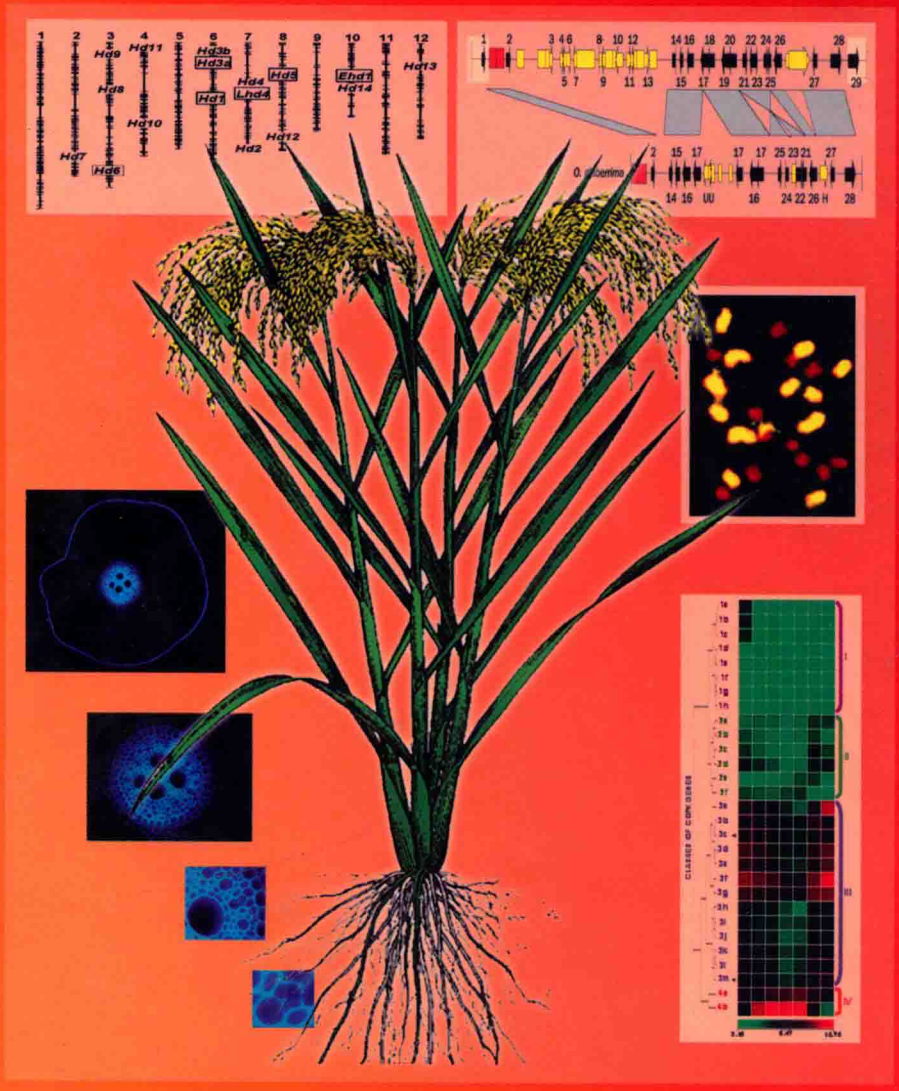
Rice Genetics V



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

IRRI

World Scientific

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

World Scientific

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.

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-13 978-981-270-772-7
ISBN-10 981-270-772-7
ISBN-13 978-971-22-0213-1 (pbk)
ISBN-10 971-22-0213-5 (pbk)

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

Printed in Singapore by World Scientific Printers (S) Pte Ltd

RICE GENETICS V

Proceedings of the Fifth International Rice Genetics Symposium

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.

Contents

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.

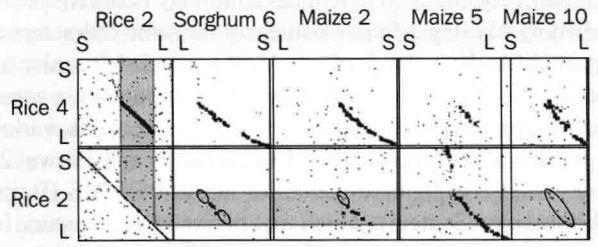


Fig. 1. Defining relationships with duplications. L = long arm, S = short arm; syntenic regions (red dots); part of chromosome 4 duplicate of rice chromosome 2 (gray); regions related to rice chromosome 2 (black circles).

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