Plant Genetics and Genomics: Crops and Models 5

Qifa Zhang
Rod A. Wing
Editors

Genetics and Genomics of Rice

Springer

Plant Genetics and Genomics: Crops and Models

Volume 5

Series Editor
Richard A. Jorgensen

For further volumes:
http://www.springer.com/series/7397

Qifa Zhang • Rod A. Wing
Editors

Genetics and Genomics of Rice

Springer

Editors
Qifa Zhang, Ph.D.
National Key Laboratory of Crop
Genetic Improvement
Huazhong Agricultural University
Wuhan, China, People's Republic

Rod A. Wing, Ph.D.
The Arizona Genomics Institute
School of Plant Sciences
Department of Ecology
and Evolutionary Biology
University of Arizona
Tucson, AZ, USA

ISBN 978-1-4614-7902-4 ISBN 978-1-4614-7903-1 (eBook)
DOI 10.1007/978-1-4614-7903-1
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013944047

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Rice is the most important food crop and a staple food for more than half of the world's population. More than 90 % of the world's rice is grown and consumed in Asia, which is home to 60 % of the Earth's people. Rice accounts for 35–75 % of the calories consumed by more than three billion Asians. World rice production has increased three times from 231 million tons in 1961 to 718 million tons in 2011. This has mainly been achieved through the application of the principles of Mendelian genetics and conventional plant breeding coupled with improved production technologies. These advances in rice production have resulted in the Green Revolution. In 2000, the average per capita food grain availability was 20 % higher than in the 1960s. The resulting food security led to political stability, investment in education, infrastructure development, and industrialization in Asia. Despite these advances in rice production, 800 million people still go to bed hungry every day and most of them are poor rice consumers. It is estimated that we will need to produce 25 % more rice by 2030. Moreover, this increased demand will have to be met utilizing less land, water, labor, and chemicals. Furthermore, rice production and sustainability are continuously threatened by several biotic (diseases, insects) and abiotic (drought, submergence, salinity) stresses. These stresses are becoming increasingly important particularly in the context of global climate change. Thus, to overcome these constraints and ensure continued food security we need to develop genetically superior rice varieties with higher yield potential, possessing multiple resistance to biotic and abiotic stresses and with more palatable and nutritious grain quality. Thanks to the advances in rice genetics and genomics, we have new tools for developing rice varieties which will help us meet the challenge of feeding future rice consumers.

Rockefeller Foundation's International Program on Rice biotechnology (1985–2000) and the International Rice Genome Sequencing Project (IRGS, 2005) contributed much to the advances in rice molecular biology. Rice has become a model plant for genetic and genomic research in higher plants. I am delighted to see that two recognized authorities in this field have undertaken to prepare this authoritative review of the present status of rice genetics and genomics. The first six chapters review the further advances in genomics since publication of the rice genome sequence in 2005. Determining the functions of rice genes is now one of the major thrust areas in rice research. Four chapters are devoted to tools and resources for the functional analysis of rice

genes. Many useful genes have been identified for rice improvement. These have been catalogued in seven chapters. A thorough understanding of rice development and biological processes is crucial for future advances in rice research. Four chapters explore the present understanding of rice biology, while the last three chapters discuss their applications in rice improvement.

The write up of Perspective on synthesis and prospects is thought provoking and lays the road map for future advances in rice genetics and genomics.

This magnum opus should serve as a standard reference for rice researchers for many years to come. I would like to congratulate Drs. Qifa Zhang and Rod Wing for their labor of love in preparing this volume.

Davis, CA Gurdev S. Khush

Preface

Rice is the staple food for a large segment of the world population and global demand for rice production will continue to grow as we add more than two billion human inhabitants to the world population by 2050. Also increasing are the constraints for crop production posed by resource shortages and environmental degradation. As a response to these challenges, the international scientific community has made tremendous progress in rice functional and evolutionary genomics and biotechnology research over the last decade. This includes, but is not limited to, genomic resources such as a gold standard reference genome sequence, the generation of hundreds of thousands of mutant lines, collections of full length cDNAs, and databases for global expression profiles and natural variation. Hundreds of rice genes have now been cloned and molecularly characterized which have led to an enhanced understanding of agronomic traits and the underpinning of important biological processes. This book is devoted to a comprehensive coverage of the advances in such research.

The chapters are organized with the following considerations in mind: (1) rice is a model for genomic research of cereals for which we intended to present the features of the rice genome and the tools available and required for genomic studies; (2) rice is a crop that urgently needs genetic improvement for which we provide the current state of our molecular understanding of traits that are vital for varietal development; and (3) the model system of rice is different from Arabidopsis, and thus we must highlight and illustrate the advances in our understanding of the unique and important biological processes of this important cereal. We are very pleased that our goal has been achieved, thanks to the tireless efforts of the contributors.

This book is for the series on Plant Genetics and Genomics by Springer Publishing Co. Credit for initiating this effort goes to Richard Jorgensen, the Series Editor, and Amna Ahmed, Publishing Editor. This book enjoys the advantage that each chapter is presented by an authority on the subject with the latest developments. We sincerely thank all the authors for their dedicated efforts, and their time and talent in writing the chapters. We are particularly indebt to Gurdev Khush for his willingness to write the Foreword. We also thank Daniel Dominguez for his hard work in the communication and progress tracking.

Wuhan, China — Qifa Zhang
Tucson, AZ, USA — Rod A. Wing

Contents

Contributors

Gynheung An, Ph.D. Crop Biotech Institute, Kyung Hee University, Yongin, Korea

Baltazar A. Antonio, Ph.D. Genome Resource Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Darshan S. Brar, Ph.D. Punjab Agricultural University (PAU), Ludhiana, Punjab, India

International Rice Research Institute (IRRI), Laguna, Philippines

Xiaofeng Cao, Ph.D. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Hao Chen, Ph.D. National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Haodong Chen, Ph.D. College of Life Sciences, Peking University, Beijing, China

Rongzhi Chen College of Life Sciences, Wuhan University, Wuhan, Hubei, China

Luca Comai, Ph.D. UC Davis Genome Center, Department of Plant Biology, College of Biological Sciences, University of California, Davis, CA, USA

Xing-Wang Deng, Ph.D. Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, USA

Ludovico Dreni, Ph.D. Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy

Gaëtan Droc, M.Sc. UMR AGAP, CIRAD, Montpellier, Cedex, France

Bo Du, Ph.D. College of Life Sciences, Wuhan University, Wuhan, Hubei, China

Dongying Gao, Ph.D. Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA

Emmanuel Guiderdoni, Ph.D. UMR AGAP, CIRAD, Montpellier, Cedex, France

Bin Han, Ph.D. National Center for Gene Research & Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Chang-Deok Han, Ph.D. Division of Applied Life Sciences, Department of Biochemistry, PMBBRC, Gyeongsang National University, Jinju, Korea

Zhongmin Han, Ph.D. National Key Lab of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Guangcun He, Ph.D. College of Life Sciences, Wuhan University, Wuhan, Hubei, China

Hang He, Ph.D. College of Life Sciences, Peking University, Beijing, China

Yuqing He, Ph.D. National Center of Molecular Breeding, Huazhong Agricultural University, Wuhan, Hubei, China

Hirohiko Hirochika, Ph.D. National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Jing Han Hong Department of Biological Sciences, National University of Singapore, Singapore

Yue-ie C. Hsing, Ph.D. Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

Xuehui Huang, Ph.D. National Center for Gene Research & Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Jianping Hu, Ph.D. Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, MI, USA

Yongfeng Hu, Ph.D. College of Bioengineering, Jingchu University of Technology, Jingmen, Hubei, China

Mayuko Ikeda, Ph.D. Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, Japan

Takeshi Itoh, Ph.D. Bioinformatics Research Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Aiko Iwata, Ph.D. Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA

Scott A. Jackson, Ph.D. Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA

Ning Jiang, Ph.D. Department of Horticulture, Michigan State University, East Lansing, MI, USA

Martin M. Kater, Ph.D. Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy

Yoshihiro Kawahara, Ph.D. Bioinformatics Research Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Hidemi Kitano, Ph.D. Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, Japan

Hei Leung, Ph.D. Genetic Diversity and Gene Discovery Program, International Rice Research Institute, Los Banos, The Philippines

Hongjing Li, Ph.D. National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan, Hubei, China

Jiayang Li, Ph.D. State Key Laboratory of Plant Genomics and Plant Gene Research Center (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Lei Li, Ph.D. Department of Biology, University of Virginia, Charlottesville, VA, USA

Zhi-Kang Li, Ph.D. Institute of Crop Sciences and Shenzhen Institute of Breeding and Innovation, Chinese Academy of Agricultural Sciences, Beijing, China

Xinming Lian, Ph.D. National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan, Hubei, China

Yongjun Lin, Ph.D. National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Xiuying Liu, Ph.D. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Mathias Lorieux, Ph.D. UMR DIADE, Institut de Recherche pour le Développement, Montpellier, France

Takashi Matsumoto, Ph.D. Plant Genome Research Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Makoto Matsuoka, Ph.D. Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, Japan

Nobuko Ohmido, Ph.D. Graduate School of Human Development and Environment, Kobe University, Kobe, Hyogo, Japan

Yidan Ouyang, Ph.D. College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China

Olivier Panaud, Ph.D. Laboratoire Génome et Développement des Plantes, UMR CNRS/IRD/UPVD, Université de Perpignan Via Domitia, Perpignan Cedex, France

Andy Pereira, Ph.D. Department of Crop, Soil and Environmental Sciences, University of Arkansas, Fayetteville, AR, USA

Srinivasan Ramachandran, Ph.D. Rice Functional Genomics, Temasek Life Sciences Laboratory, National University of Singapore, Singapore

Hiroaki Sakai, Ph.D. Bioinformatics Research Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Paul L. Sanchez, Ph.D. School of Plant Sciences, Arizona Genomics Institute (AGI), University of Arizona, Tucson, AZ, USA

Takuji Sasaki, Ph.D. NODAI Research Institute, Tokyo University of Agriculture, Setagaya-ku, Tokyo, Japan

Ko Shimamoto, Ph.D. Laboratory of Plant Molecular Genetics, Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, Japan

Xianwei Song, Ph.D. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Venkatesan Sundaresan, Ph.D. College of Biological Sciences, University of California, Davis, CA, USA

Tsuyoshi Tanaka, Ph.D. Bioinformatics Research Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Hiroyuki Tsuji, Ph.D. Laboratory of Plant Molecular Genetics, Graduate School of Biologocal Sciences, Nara Institute of Science and Technology, Ikoma, Nara, Japan

Narayana Upadhyaya, B.Sc., M.Sc., Ph.D. CSIRO Plant Industry, Canberra, ACT, Australia

Shiping Wang, Ph.D. National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan, Hubei, China

Xiu-jie Wang, Ph.D. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Rod A. Wing, Ph.D. The Arizona Genomics Institute, School of Plant Sciences & Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

Bi Wu, Ph.D. National Key Lab of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Changyin Wu, Ph.D. National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Ping Wu, Ph.D. College of Life Science, Zhejiang University, Hangzhou, Zhejiang, China

Lizhong Xiong, Ph.D. National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan, China

Yongzhong Xing, Ph.D. National Key Lab of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan, Hubei, China

Guohua Xu, Ph.D. College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing, Jiangsu, China

Jian Xu, Ph.D. Department of Biological Sciences, National University of Singapore, Singapore

Zheng Yuan, Ph.D. School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Huihui Yu, Ph.D. Genomic Breeding, Life Science and Technology Center, China National Seed Group Co. Ltd., Wuhan, Hubei, China

Yanchun Yu, Ph.D. School of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, Zhejiang, China

Dabing Zhang, Ph.D. School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Qifa Zhang, Ph.D. National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

Yu Zhao, Ph.D. National Key Laboratory for Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Tian-Qing Zheng, Ph.D. Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China

Dao-Xiu Zhou, Ph.D. Institut de Biologie des Plantes, Université Paris, Orsay, France

Fasong Zhou, Ph.D. Genomic Breeding, Life Science and Technology Center, China National Seed Group Co., Ltd., Wuhan, Hubei, China

Dan Zhu National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

A Reference Rice Genome Sequence in the 10K Genome Era

1

Takeshi Itoh, Baltazar A. Antonio, Yoshihiro Kawahara, Tsuyoshi Tanaka, Hiroaki Sakai, Takashi Matsumoto, and Takuji Sasaki

1 Introduction

The completion of a high quality map-based genome sequence of *Oryza sativa* ssp. *japonica* cv. Nipponbare is one of the most important achievements in science with implications ranging from basic biology to the applied aspects of crop improvement in agriculture. The availability of the complete blueprint of all the genes of rice in the public domain provides an impetus that drives studies on structural, functional, and applied rice genomics. The rice genome sequence has also become a reference genome that is now being used in understanding the genome structure and function of major cereal crops including maize, wheat, barley, and sorghum. As rice has a very rich collection of germplasm resources, the Nipponbare genome sequence has proven to be an indispensable tool in identifying the genetic variability within and among the different species of the genus *Oryza*, an important step in looking for important agronomic traits in distantly related and wild rice species that can be incorporated in modern cultivars.

The completion of the genome sequence has also paved the way for more extensive projects on rice genomics. The *Oryza* Map Alignment Project has been initiated with the aim of using the Nipponbare genome sequence as a reference to characterize the wild species of rice and to create a genome-level experimental system for understanding the evolution, domestication, and genome organization of the genus *Oryza* [32]. The International Rice Functional Genomics Consortium (IRFGC) initiative that focused on sequence analysis of diverse rice cultivars of rice for the purpose of identifying SNPs based on 20 *japonica*/*indica* cultivars and landraces has elucidated the genotypic and phenotypic diversity of domesticated rice and allowed for identification of 160,000 SNPs [22]. More recently, the genome-wide association study (GWAS) has been widely adopted in rice as a strategy to characterize many common genetic variants across different accessions, to elucidate how these variants are associated with complex agronomic traits, and to reveal the heterogeneity of genetic architecture among diverse rice cultivars [10, 11, 36].

T. Itoh, Ph.D. • Y. Kawahara, Ph.D. • T. Tanaka, Ph.D. H. Sakai, Ph.D.
Bioinformatics Research Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

B.A. Antonio, Ph.D.
Genome Resource Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

T. Matsumoto, Ph.D.
Plant Genome Research Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

T. Sasaki, Ph.D. (✉)
NODAI Research Institute, Tokyo University of Agriculture, Sakuragaoka 1-1-1, Setagaya-ku, Tokyo 156-8502, Japan
e-mail: t5sasaki@nodai.ac.jp

Q. Zhang and R.A. Wing (eds.), *Genetics and Genomics of Rice*, Plant Genetics and Genomics: Crops and Models 5, DOI 10.1007/978-1-4614-7903-1_1, © Springer Science+Business Media New York 2013

2 Thousand-Genome Era

With the rapid advances in genome sequencing technology, the scenario has changed dramatically since the completion of the rice genome sequence in 2004. Genome researchers around the world are now focusing on sequencing not just a single representative of a species but rather hundreds and even thousands of closely related genomes or representatives of a single species [6, 7]. In plants, the 1,000 Plant Genomes Project is large-scale endeavor to obtain the transcriptome of representative species in the entire plant kingdom including algae, nonflowering plants, and angiosperms (http://www.onekp.com/). For the model dicot plant, the Arabidopsis genome research community has also embarked on a project that involved the whole genome sequencing of 1,001 Arabidopsis strains [31]. These highly ambitious genomics initiatives take advantage of next-generation sequencing technologies that could facilitate sequencing, resequencing, and assembling of entire genomes more accurately and efficiently, at a much lower cost and at a much greater depth.

With a large variation in the genus *Oryza* and the urgent need in agriculture to increase yield and improve the quality of crops, hundreds of cultivars and wild rice species have already been partially sequenced and analyzed by genome-wide association studies [10, 11, 33]. In addition, the genome sequences of more than 10,000 cultivars representing global rice genetic diversity are expected to be released in the near future. In particular, the "Rice 10,000 Genome Project" has been initiated as a joint collaborative effort of the International Rice Research Institute (IRRI), Chinese Academy of Agricultural Science (CAAS), and Beijing Genomics Institute (BGI) to characterize the global diversity of rice (http://bgiamericas.com/scientific-expertise/collaborative-projects/). This project would lead to the genome sequencing of modern rice varieties, traditional rice varieties, and wild relatives selected from the gene bank collections at IRRI and would guarantee more robust rice breeding programs.

The value of sequencing thousands of rice varieties could be enhanced if a high quality reference genome is available. Therefore, it is an urgent issue to provide a reference genome sequence with ultrahigh quality that can be used for many genomic studies such as large-scale SNP detections among cultivars. In this review, the major features of the reference rice genome sequence are described with emphasis on current initiatives to improve the sequence via concerted efforts to generate a unified genome sequence assembly. The annotation efforts of the Rice Annotation Project Database (RAP-DB) are described with new added features that will further enhance the genome sequence and provide more information on each predicted gene of rice.

3 Generating the Map-Based Genome Sequence

The initiative that led to the sequencing of the rice genome started in 1992 with the Rice Genome Research Program (RGP), a national project funded by the Ministry of Agriculture, Forestry and Fisheries of Japan with the aim of characterizing the rice plant, which is considered as the main pillar of Japanese agriculture. During the first phase of the program, it has successfully established a high-density linkage map of rice [9, 15], an extensive catalog of all expressed rice genes [34], and a yeast artificial chromosome (YAC)-based physical map covering the entire genome [16]. Even at the early stages of the project, these molecular tools were proven to be useful in understanding the genomic structure of rice and other cereal grasses, which were eventually found to share extensive similarity in their genetic make-up.

In 1998, the RGP embarked on the second phase of genome analysis, which was aimed at sequencing the entire rice genome. Even with a relatively small genome size estimated at 430 Mb, the task of sequencing the entire genome was undeniably enormous at a time when the high-throughput DNA sequencing technology was still at relatively early stages of development. The genome sequencing initiative in Japan served as a stimulus for the USA as well as other Asian and European countries to establish similar programs on rice genome analysis.

Eventually an international collaboration that has evolved into the International Rice Genome Sequencing Project (IRGSP) was organized with the aim of sharing resources and technology to accelerate the completion of sequencing the rice genome [26]. The IRGSP, formally established in 1998, pooled the resources of sequencing groups in ten nations to obtain a complete finished quality sequence of a single inbred cultivar, Nipponbare, and adopted a hierarchical clone-by-clone method using bacterial artificial chromosome (BAC) clones and P1 artificial chromosome (PAC) clones. This strategy employed a high-density genetic map, expressed sequence tags (ESTs), YAC-based physical map, and PAC/BAC-based physical map, BAC-end sequences, and draft sequences contributed by two commercial companies, namely, Monsanto [5] and Syngenta [8].

The map-based sequence of the Nipponbare genome consisted of 3,401 BAC/PAC clones which were sequenced to approximately tenfold sequence coverage, assembled, ordered, and finished to a sequence quality of less than one error per 10,000 bases [12]. A majority of physical gaps in the BAC/PAC tiling path were bridged using a variety of substrates including PCR fragments, 10 kb plasmids, and 40 kb fosmid clones. In total, the finished quality sequence covered 95 % of the 389 Mb genome including virtually all of the euchromatin regions and the two centromeres. A total of 37,544 non-transposable element-related protein coding genes were identified, 71 % of which had a putative homolog in Arabidopsis. Fourteen percent of the 37,544 predicted genes were found to appear in tandem duplications.

The publication on the rice genome sequence was based on the Build 3.0 assembly, which was also used as the template for manual curation of annotation in conjunction with the First Rice Annotation Project Meeting (RAP1) which was held in 2004. A year later, another update led to the construction of Build 4.0 genome assembly. Subsequently, the nucleotide sequences of seven new clones mapped on the euchromatin–telomere junctions were added, several clones in the centromere region of chromosome 5 were improved, and one gap on chromosome 11 was closed which led to the release of the Build 5.0 genome assembly at the end of 2008. The most recent version of the PAC/BAC-based physical map of rice that served as the template for genome sequencing is shown in Fig. 1.1. Although significant progress has been achieved since the initial publication of the entire genome, particularly in closing the gaps and characterizing the centromere and telomere-ends, some regions still remain to be completely sequenced.

4 Highly Accurate Genome Sequence

In order to further improve the quality of the map-based genome sequence, we resequenced the Nipponbare genome using the Illumina Genome Analyzer and obtained a total of 70 million 36-bp single-end reads and 60 million 51-bp single-end reads. These sequences were combined with more than 269 million 76-bp paired-end reads obtained independently by the Cold Spring Harbor Laboratory (W.R. McCombie, personal communication). After removing the low-quality sites and adapter sequences, the reads were mapped to the Nipponbare reference genome sequence by the Burrows–Wheeler Alignment (BWA) tool that could efficiently align short sequencing reads against a reference sequence [17]. The uniquely mapped reads could cover 90.6 % of the genome with an average depth of 43.6, which should be large enough to thoroughly validate the sequencing errors. We searched for sites covered by at least ten reads and found that 80 % or more were different from the reference genome sequence so that the total number of the nucleotide sites that could be used for further analysis were 321 Mb. As a result, a total of 3,447 sites were detected as SNP-type errors in the reference genome. In addition, a total of 1,439 small insertion/deletion-type errors were also found. Hence, a sequencing error rate of 1.5×10^{-5} per site was obtained. In addition to the Illumina reads, we also used 2,706,353 reads (1.0 Gb) generated by GS FLX Titanium. To detect large gaps, these reads were aligned to the reference genome by megablast [35]. After manual inspection of the results, merely five erroneous gaps remained and were corrected.

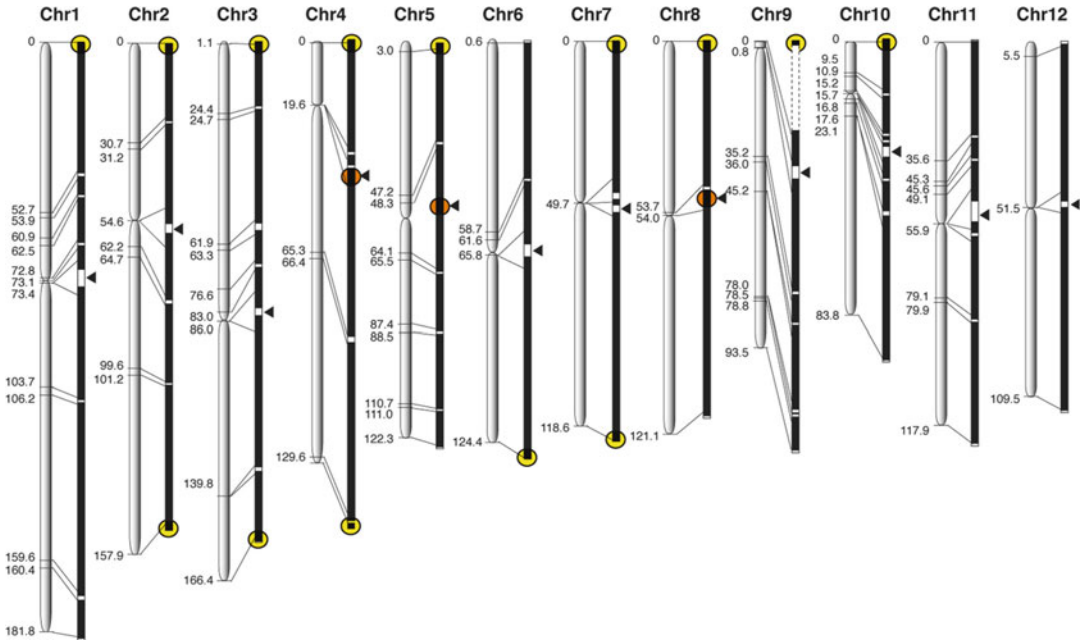

Fig. 1.1 Physical map of the 12 rice chromosomes. For each chromosome (Chr1–12), the genetic map is shown on the *left* and the physical map on the *right*. The position of markers flanking the gaps of the physical map which are shown in *white* is indicated on the genetic map. The nucleolar organizer on chromosome 9 is represented with a *dotted line*. Constrictions in the genetic maps and *arrowheads* to the right of physical maps represent the chromosomal positions of centromeres. The centromeres of chromosomes 4, 5, and 8 (*circled*) represent the first fully sequenced centromere reported in plants. The telomere-specific repetitive sequence unit, CCCTAAA, has been clarified in 14 telomeres. The genetic map is scaled to genetic distance in centimorgan (cM) and the physical map corresponds to the relative physical length

The IRGSP genome assembly was unified with the sequence assembly from the MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/). Additionally, the assembly has been reconstructed using an optimal BAC tiling path that included a BAC-optimal map [37]. The unified high quality genome assembly, Os-Nipponbare-Reference-IRGSP-1.0 (IRGSP-1.0), can be downloaded at http://rapdb.dna.affrc.go.jp/ and http://rice.plantbiology.msu.edu/.

5 Annotation of the Os-Nipponbare-Reference-IRGSP-1.0

The annotation of the Nipponbare genome assembly has been available through the RAP-DB including automatically predicted and manually curated gene models [24, 29]. Re-annotation of the unified IRGSP-1.0 genome assembly was then conducted. The primary gene structures were determined by the cDNA-mapping method described previously [4, 29]. We used 207,343 major monocotyledon cDNA sequences registered in the international DNA databases (EMBL/DDBA/GenBank) including 81,129 cDNAs obtained from *Zea mays*, 39,676 from *Oryza sativa* (*japonica*), 30,270 from *Hordeum vulgare*, 26,321 from *Triticum aestivum*, and 11,789 from *O. sativa* (*indica*). Most of these sequences cover the full-length transcript [1, 13, 14, 18, 21, 23, 27, 28], so that it is expected that the complete exon–intron structure can be reconstructed on the genome. The protein sequences of plants in UniProt [30] and RefSeq [25] were also aligned to the genome so that the number of missing genes in the annotation data set could be

Table 1.1 Statistics of annotated loci in the Build 5 and IRGSP-1.0 assemblies

	Build 5.0	IRGSP-1.0
Protein coding loci	31,232	33,276
Nonprotein coding loci	1,515	2,191
Ab initio predictions[a]	2,034	2,405
Total	34,781	37,872

[a]The primary gene structures were determined by ab initio prediction methods and all of them are possibly protein coding

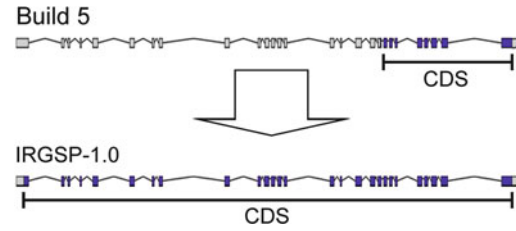

Fig. 1.2 Improved CDS regions by error-corrections in the IRGSP-1.0 assembly

maintained at a minimum. To further add genes without supporting cDNA or protein sequence information, ab initio gene predictions by GlimmerHMM, GeneMark.hmm, and GeneZilla were conducted [19, 20] and the prediction results were integrated by JIGSAW [2]. We compared the predicted genes with 6,700,357 ESTs of major cereals by BLASTN [3]. Predictions were employed in the final annotation data set for all annotations supported by the expression evidence of ESTs.

This comprehensive annotation resulted in identification and prediction of 37,872 loci (Table 1.1). In the previous assembly (Build 5), a total of 34,781 loci were predicted corresponding to 31,232 protein coding loci, 1,515 nonprotein coding loci, and 2,034 ab initio predictions. More than 3,000 novel loci were found in the current assembly because thousands of full-length cDNAs of maize and barley were newly added to the cDNA-mapping data set. In many cases, these novel loci were located in the opposite strand of other loci suggesting that they might be possible antisense transcripts. However it is also possible that these loci are transcribed specifically in non-rice species or are experimental artifacts that should be subjected to further validation.

The genome-wide error correction of the IRGSP-1.0 assembly improved the annotation. For example, a protein coding region of a locus (Os10g0477800) was truncated in the Build 5 assembly (Fig. 1.2), but it turned out that this coding-frame disruption was artificial. In fact, the coding region could be extended to the genuine amino-terminal in the IRGSP-1.0 assembly and the intact frame was recovered. Since the rice genome was deciphered by the Sanger method on the basis of a precise physical map, the sequence quality was thought to be high enough. Nonetheless, to cope with the continuous increase of sequence data from thousands of cultivars lined-up for sequencing, the reference genome assembly should be of exceptionally high quality to facilitate accurate comparative analysis among different cultivars.

6 Concluding Remarks

Rice biology is now in the midst of a genomics revolution with the cheap, fast, and ubiquitous sequencing of thousands of cultivars and lines that represent the *Oryza* germplasm resources. In the next few years or so, the public databases will be flooded with rice sequence data of almost all cultivars and lines grown around the world. A high quality rice genome sequence will always play a pivotal role in analysis of these genome sequence data. Available to researchers worldwide, the rice genome reference sequence provides an unprecedented biological resource to the scientific community that will serve as a basis for research and discovery of novel genes and, ultimately, introgression of these genes into cultivars grown in different cultivation conditions. The sequence already is having an impact on finding genes associated with many agronomic traits in rice. Other rice genome sequence projects focusing on various rice cultivars will enable detailed comparisons among cultivars, species, and wild relatives. It is therefore important that the Os-Nipponbare-Reference-IRGSP-1.0 genome assembly is of the highest quality. The next challenge will focus on how to

use all the information in actual breeding programs to facilitate the breeding of new varieties that are stronger, faster-growing, and higher-yielding than varieties currently available, to guarantee a stable food supply for mankind.

References

1. Alexandrov NN, Brover VV, Freidin S, Troukhan ME, Tatarinova TV, Zhang H et al (2009) Insights into corn genes derived from large-scale cDNA sequencing. Plant Mol Biol 69:179–194
2. Allen JE, Salzberg SL (2005) JIGSAW: integration of multiple sources of evidence for gene prediction. Bioinformatics 15:3596–3603
3. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402
4. Amano N, Tanaka T, Numa H, Sakai H, Itoh T (2010) Efficient plant gene identification based on interspecies mapping of full-length cDNAs. DNA Res 17:271–279
5. Barry GF (2001) The use of the Monsanto draft rice genome sequence in research. Plant Physiol 125: 1164–1165
6. Durbin RM, Altshuler DL, Durbin RM, Abecasis GAR, Bentley DR, Chakravarti A, Clark AG, Collins FS et al (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073
7. Genome 10K Community of Scientists (2009) Genome 10K: a proposal to obtain whole-genome sequence for 10000 vertebrate species. J Hered 100:659–674
8. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100
9. Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y et al (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479–494
10. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967
11. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q et al (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44:32–39
12. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800
13. Jia J, Fu J, Zheng J, Zhou X, Huai J, Wang J et al (2006) Annotation and expression profile analysis of 2073 full-length cDNAs from stress-induced maize (*Zea mays* L.) seedlings. Plant J 48:710–727
14. Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 301:376–379
15. Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J et al (1994) A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nat Genet 8:365–372
16. Kurata N, Umehara Y, Tanoue H, Sasaki T (1997) Physical mapping of the rice genome with YAC clones. Plant Mol Biol 35:101–113
17. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
18. Liu X, Lu T, Yu S, Li Y, Huang Y, Huang T et al (2007) A collection of 10,096 *indica* rice full-length cDNAs reveals highly expressed sequence divergence between *Oryza sativa indica* and *japonica* subspecies. Plant Mol Biol 65:403–415
19. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 33:6494–6506
20. Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics 20:2878–2879
21. Matsumoto T, Tanaka T, Sakai H, Amano N, Kanamori H, Kurita K et al (2011) Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. Plant Physiol 156:20–28
22. McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci U S A 106:12273–12278
23. Mochida K, Yoshida T, Sakurai T, Ogihara Y, Shinozaki K (2009) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. Plant Physiol 150:1135–1146
24. Ohyanagi H, Tanaka T, Sakai H, Shigemoto Y, Yamaguchi K, Habara T et al (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. Nucleic Acids Res 34:D741–D744
25. Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res 40:D130–D135
26. Sasaki T, Burr B (2000) International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr Opin Plant Biol 3:138–141
27. Sato K, Shin-I T, Seki M, Shinozaki K, Yoshida H, Takeda K et al (2009) Development of 5006 full-length CDNAs in barley: a tool for accessing cereal genomics resources. DNA Res 16:81–89
28. Soderlund C, Descour A, Kudrna D, Bomhoff M, Boyd L, Currie J et al (2009) Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. PLoS Genet 5:e1000740

29. Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, Numa H et al (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. Nucleic Acids Res 36:D1028–D1033
30. UniProt Consortium (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40:D71–D75
31. Weigel D, Mott R (2009) The 1001 Genomes Project for *Arabidopsis thaliana*. Genome Biol 10:107
32. Wing RA, Ammiraju J, Luo M, Kim H, Yu Y, Kudrna D 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
33. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X et al (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30: 105–111
34. Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35:135–144
35. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214
36. Zhao K, Tung CW, Eizenga G, Wright MH, Ali ML, Price AH et al (2011) Genome-wide association mapping reveals rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467
37. Zhou S, Bechner MC, Place M, Churas P, Pape L, Leong S et al (2007) Validation of rice genome sequence by optimal mapping. BMC Genomics 8:278

The Wild Relative of Rice: Genomes and Genomics

2

Paul L. Sanchez, Rod A. Wing, and Darshan S. Brar

The wild species of the genus *Oryza* serve as a virtually untapped reservoir of genetic diversity that can be used to improve the world's most important food crop—rice. The genus is composed of two domesticated (*O. sativa* and *O. glaberrima*) and 22 wild species [68] and represents between 15 and 25 million years of evolutionary diversification.

In this chapter we will describe the current status of the genetic and genomic applications of the genus *Oryza* toward the penultimate goal of helping to solve the 9 billion people question—i.e., how can we grow enough food to feed more than 9 billion human inhabitants under 40 years [52]?

P.L. Sanchez, Ph.D.
School of Plant Sciences, Arizona Genomics Institute (AGI), The University of Arizona, 1657 East Helen Street, Tucson, AZ 85721, USA

R.A. Wing, Ph.D. (✉)
The Arizona Genomics Institute, School of Plant Sciences & Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
e-mail: rwing@ag.arizona.edu

D.S. Brar, Ph.D.
Punjab Agricultural University (PAU), Ludhiana, Punjab, India

International Rice Research Institute (IRRI), Los Baños, Laguna, Philippines

1 The Genus *Oryza*: Broadening the Gene Pool of Rice—Exploitation of Diversity of the Wild Species Germplasm

The genus *Oryza* includes two cultivated ($2n = 24$, AA) and 22 wild species ($2n = 24$, 48) representing the AA, BB, CC, BBCC, CCDD, EE, FF, GG, KKLL, and HHJJ genome types (Table 2.1). Figures 2.1 and 2.2 show a phylogenetic tree of the genus (inferred from [3, 6, 25, 45]) and a photograph of 12 of these species at the same developmental stage, respectively. These wild *Oryza* species are, in fact, grass-like plants which are phenotypically inferior in agronomic traits—such as poor plant type, low grain yield, poor grain type, and are shattering in nature [1]. The wild species exhibit tremendous diversity in morphological traits, height, tillering, flowering, growth habit, panicle, leaf, culm, and seed characteristics (Fig. 2.2), and adaptation to different habitats and agronomic traits (Table 2.1).

2 Gene Transfer from Wild Species into Rice

The International Rice Research Institute's (IRRI) Rice Gene Bank and The National Institute of Genetics' *Oryza* base, combined, maintain more than 4,000 accessions of wild *Oryza* species and 1,500 accessions of cultivated

Q. Zhang and R.A. Wing (eds.), *Genetics and Genomics of Rice*, Plant Genetics and Genomics: Crops and Models 5, DOI 10.1007/978-1-4614-7903-1_2, © Springer Science+Business Media New York 2013

Table 2.1 Chromosome number, genomic composition, and distribution of *Oryza* species, *Oryza*-related genera, and their useful traits

Species	2*n*	Genome	Number of accessions	Distribution	Useful traits
O. sativa complex					
O. sativa L.	24	AA	96,564	Worldwide	Cultigen, high yielding
O. glaberrima Steud.	24	A^gA^g	1,562	West Africa	Cultigen; tolerance to drought, acidity, iron toxicity, P-deficiency; resistance to BB, blast, RYMV, African gall midge, nematodes, weed competitiveness
O. nivara Sharma et Shastry	24	AA	1,260	Tropical and subtropical Asia	Resistance to grassy stunt virus, BB
O. rufipogon Griff.	24	AA	858	Tropical and subtropical Asia, tropical Australia	Resistance to BB, blast, BPH, tungro virus; moderately tolerant to Shb, tolerance to aluminum and soil acidity, increased elongation under deep water; source of CMS and yield-enhancing loci
O. breviligulata A. Chev. et Roehr. *O. barthii*	24	A^gA^g	218	Africa	Resistance to GLH, BB; drought avoidance; tolerance to heat and drought
O. longistaminata A. Chev et Roehr	24	A^lA^l	203	Africa	Resistance to BB, nematodes, stemborer, drought avoidance
O. meridionalis Ng	24	A^mA^m	56	Tropical Australia	Elongation ability; drought avoidance; tolerance to heat and drought
O. glumaepatula Steud.	24	$A^{gp}A^{gp}$	54	South and Central America	Elongation ability; source of CMS; tolerance to heat
O. officinalis complex					
O. punctata Kotschy ex Steud.	24, 48	BB, BBCC	71	Africa	Resistance to BPH, BB, zigzag leafhopper; tolerance to heat and drought
O. minuta J.S. Presl. ex C.B. Presl.	48	BBCC	63	Philippines and Papua New Guinea	Resistance to BB, blast, BPH, GLH
O. officinalis Wall ex Watt	24	CC	265	Tropical and subtropical Asia, tropical Australia	Resistance to thrips, BPH, GLH, WPH, BB, stem rot; tolerance to heat
O. rhizomatis Vaughan	24	CC	19	Sri Lanka	Drought avoidance, resistance to blast; tolerance to heat
O. eichingeri A. Peter	24	CC	30	South Asia and East Africa	Resistance to BPH, WBPH, GLH
O. latifolia Desv.	48	CCDD	40	South and Central America	Resistance to BPH, BB, high biomass production
O. alta Swallen	48	CCDD	6	South and Central America	Resistance to striped stemborer; high biomass production
O. grandiglumis (Doell) Prod.	48	CCDD	10	South and Central America	High biomass production
O. australiensis Domin.	24	EE	36	Tropical Australia	Resistance to BPH, BB, blast; drought avoidance; tolerance to heat and drought

(continued)

Table 2.1 (continued)

Species	2*n*	Genome	Number of accessions	Distribution	Useful traits
O. meyeriana complex					
O. granulata Nees et Arn. ex Watt	24	GG	24	South and South Asia	Shade tolerance, adaptation to aerobic soil
O. meyeriana (Zoll. et (Mor. ex Steud.) Baill.)	24	GG	11	Southeast Asia	Shade tolerance; adaptation to aerobic soil
O. ridleyi complex					
O. longiglumis Jansen	48	HHJJ	6	Irian Jaya, Indonesia, and Papua New Guinea	Resistance to blast, BB
O. ridleyi Hook. F.	48	HHJJ	15	South Asia	Resistance to blast, BB, tungro virus, stem borer, whorl maggot
Unclassified					
O. brachyantha A. Chev. et Roehr	24	FF	19	Africa	Resistance to BB, yellow stemborer, leaf folder, whorl maggot; tolerance to laterite soil
O. schlechteri Pilger	48	KKLL	1	Papua New Guinea	Stoloniferous
O. coarctata Tateoka	48	KKLL	1	Asian Coastal Area	Tolerance to salinity, stoloniferous
Leersia perrieri A. Camus	24	UNKNOWN	1	Africa	Shade tolerance, stoloniferous

BPH brown plant hopper, *GLH* green leaf hopper, *WBPH* whitebacked plant hopper, *BB* bacterial blight, *Shb* sheath blight, *CMS* cytoplasmic male sterility, *RYMV* rice yellow mottle virus

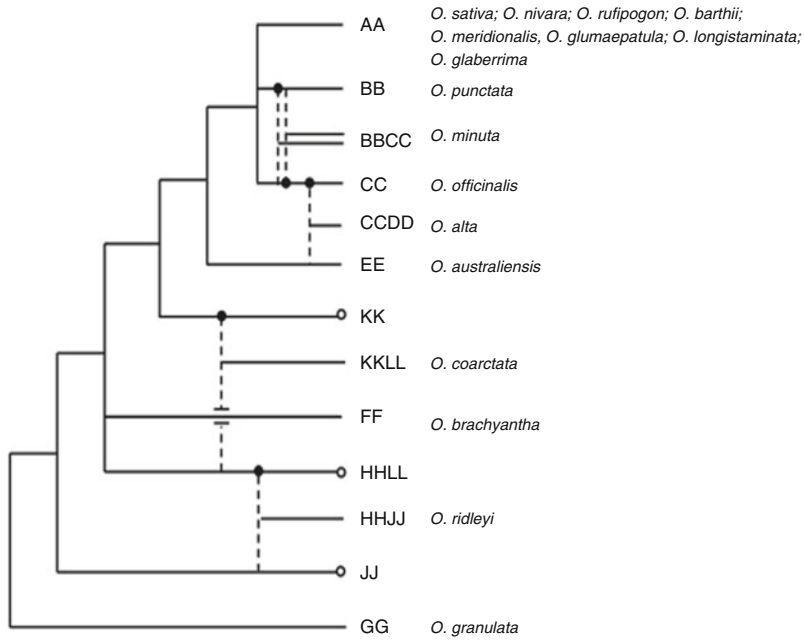

Fig. 2.1 *Oryza* phylogenetic tree. Evolutionary relationships of the *Oryza* genome were inferred from Ammiraju et al. [3, 6], Ge et al. [25], and Lu et al. [46]. *Dashed line* indicates origins of allotetraploids; *filled circle* indicates maternal parents; *open circle* indicates unidentified diploid species

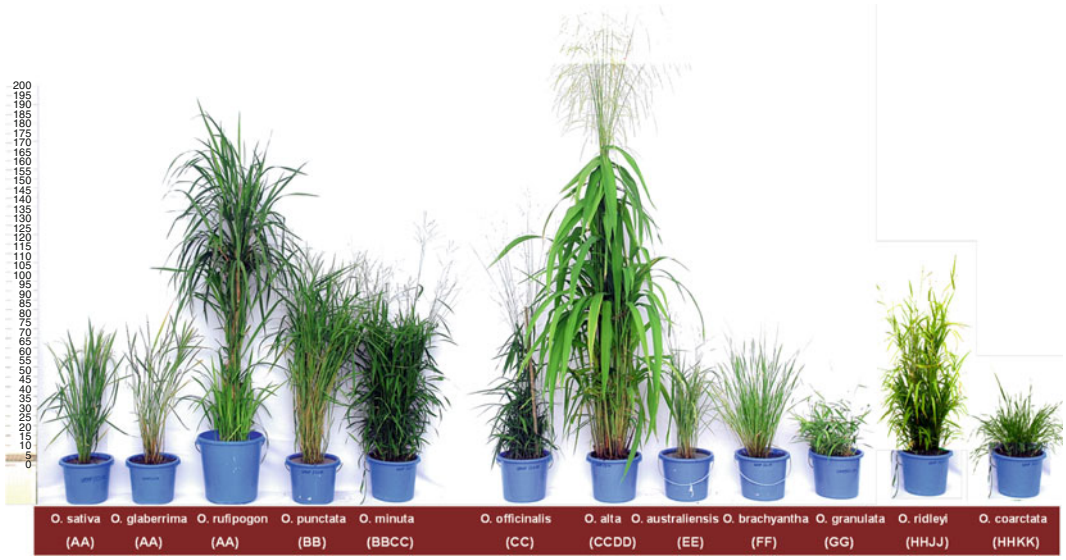

Fig. 2.2 The genus *Oryza*: 12 representative species

African rice (*O. glaberrima*). These wild species are reservoirs of many useful genes, particularly for resistance to major biotic and abiotic stresses (Table 2.1). However, these wild species are associated with several weedy traits, such as grain shattering, poor plant type, poor grain characteristics, and low seed yield. Besides, several incompatibility barriers limit the transfer of useful genes from wild species into cultivated species [14, 15]. The major consideration in alien gene transfer is to selectively transfer agronomically important genes from wild species while avoiding linkage drag. To achieve precise transfer of genes from wild species, strategies involving a combination of conventional plant breeding methods with tissue culture and molecular approaches have become important [16, 17]. Advances in tissue culture, molecular marker technology, genomics, and fluorescence in situ hybridization have opened new opportunities to tap alien genetic variability from distant *Oryza* genomes through interspecific hybridization.

3 Strategy for Alien Gene Transfer into Cultivated Rice

The strategy used to transfer genes from wild species into rice depends on the nature of the target trait(s), relatedness of the wild species, and incompatibility barriers. Several protocols are available to overcome such barriers [16]. Some of the steps involved in gene transfer include: (1) *Search for useful genetic variability for target traits*—this involves the screening of wild species to identify specific accession(s) that possess useful genetic variability. (2) *Production of hybrids and alien introgression lines* (*AIL*): interspecific hybrids are produced between elite breeding lines with the wild species carrying the desired traits. Such hybrids are produced through direct crosses between rice and AA genome wild species. However, embryo rescue is required to produce hybrids and backcross progenies (introgression lines) between rice and all the wild species of *Oryza* except AA genome species. (3) *Evaluation of introgression lines for transfer of target traits*: AILs generated through backcrossing are evaluated for the transfer of target traits. This involves extensive laboratory, greenhouse, and field testing. (4) *Molecular mapping of genes/QTLs*: molecular markers are developed to track the desired alien trait(s) for marker-assisted selection (MAS).

Following the above strategy, a number of genes have been transferred from wild species into rice (Table 2.2) and varieties have been released for commercial cultivation (Table 2.3). MAS has been practiced and some varieties have become available (Table 2.4).

Table 2.2 Introgression of genes from wild *Oryza* species into rice

	Donor *Oryza* species		
Trait	Wild species	Gene	Genome
Grassy stunt resistance	*O. nivara*	*GS*	AA
Bacterial blight resistance	*O. rufipogon*	*Xa23*	AA
	O. longistaminata	*Xa21*	AA
	O. nivara	*Xa38*	AA
	O. officinalis	*Xa29*(t)	CC
	O. minuta	*Xa27*	BBCC
	O. latifolia	Unknown	CCDD
	O. australiensis	Unknown	EE
	O. brachyantha	Unknown	FF
Blast resistance	*O. glaberrima*[a]	Unknown	AA
	O. rufipogon	Unknown	AA
	O. minuta	*Pi9*	BBCC
	O. australiensis	*Pi40*	EE
Brown planthopper resistance	*O. officinalis*	*bph11*, *bph12*	CC
	O. eichingeri	*Bph14*, *Bph15*	CC
	O. minuta	*Bph20*, *Bph21*	BBCC
	O. latifolia	Unknown	CCDD
	O. australiensis	*Bph10*, *Bph18*	EE
Whitebacked planthopper resistance	*O. officinalis*	*Wbph7*(t), *Wbph8*(t)	CC
	O. latifolia	Unknown	CCDD
Cytoplasmic male sterility	*O. sativa* f. *spontanea*	Unknown	AA
	O. perennis	Unknown	AA
	O. glumaepatula	Unknown	AA
	O. rufipogon	Unknown	AA
Tungro tolerance	*O. rufipogon*	Unknown	AA
Tolerance to iron toxicity	*O. rufipogon*	Unknown	AA
	O. glaberimma[a]	Unknown	AA
Heat and/or drought-related traits	*O. australiensis*[b]	Unknown	EE
	O. barthii[b]	Unknown	AA
	O. glaberimma[a]	QTLs	AA
	O. glumaepatula[b]	Unknown	AA
	O. meridionalis[b]	Unknown	AA
	O. officinalis[b]	Unknown	CC
	O. punctata[b]	Unknown	BB
	O. rhizomatis[b]	Unknown	CC
Tolerance to aluminum toxicity	*O. rufipogon*	QTL	AA
Tolerance to acidic conditions	*O. glaberrima*[a]	Unknown	AA
	O. rufipogon	Unknown	AA
Tolerance to P-deficiency	*O. rufipogon*	Unknown	AA
	O. glaberimma[a]	Unknown	AA
Yield-enhancing loci	*O. rufipogon*	QTL, *yld1*, *yld2*	AA
Yellow stemborer (larval mortality)	*O. longistaminata*	QTL	AA
Increased elongation ability	*O. rufipogon*	Unknown	AA

[a]*O. glaberrima*—African rice species. Modified from Brar and Khush [17]
[b]*O. australiensis*, *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. officinalis*, *O. punctata*, *O. rhizomatis*. Sanchez et al. (unpublished data)

Table 2.3 Rice varieties developed through wide hybridization

Key trait	Wild species	Varieties released	Country
Grassy stunt resistance	*O. nivara*	Many rice varieties	Rice growing countries in Asia
BPH resistance	*O. officinalis*	MTL 98, MTL 103 MTL 105, MTL114	Vietnam
Acid sulfate tolerance	*O. rufipogon*	AS 996	Vietnam
Salinity tolerance	*O. rufipogon*	BRRIdhan55 (As996)	Bangladesh
Tungro resistance	*O. rufipogon*	Matatag 9	Philippines
Bacterial blight resistance	*O. longistaminata*	NSICRc 112	Philippines
Blast resistance	*O. rufipogon*	Dhanarasi	India
	O. glaberrima[a]	Yun Dao	YAAS, China
High yield, earliness, weed competitive ability, and tolerance to abiotic stresses	*O. glaberrima*[a]	Many Nerica lines/varieties	African countries
Tolerance to heat	*O. meridionalis*	Arizona Rice-1[b]	USA
		Arizona Rice-2[b]	USA

Modified from Brar and Singh [18] and Sanchez et al. (unpublished data)
[a]*O. glaberrima*—African rice species
[b]*Arizona Rice-1 and 2*—varieties to be released in 2013, Sanchez et al. (unpublished data)

Table 2.4 Rice varieties developed through MAS carrying *Xa21* gene from *O. longistaminata* and *Bph18* from *O. australiensis*

Inbreds/hybrids	Year	Resistance gene(s)	Institute/country
NSICRc 142 (Tubigan 7)	2006	*Xa4*+*Xa21*[a]	PhilRice, Philippines
NSICRc 154 (Tubigan 11)	2007	*Xa4*+*Xa21*[a]	PhilRice, Philippines
Improved Sambha Mahsuri	2007	*Xa5*+*xa13*+*Xa21*[a]	India
Improved Pusa Basmati 1	2007	*Xa5*+*xa13*+*Xa21*[a]	India
Xieyou 218	2002	*Xa21*[a]	China
Zhongyou 218	2002	*Xa21*[a]	China
Guodao 1	2002	*Xa4*+*xa5*+*xa13*+*Xa21*[a]	China
Guodao 3	2004	*Xa4*+*xa5*+*xa13*+*Xa21*[a]	China
Neizyou	2002	*Xa4*+*xa5*+*xa13*+*Xa21*[a]	China
Ilyou 8006	2005	*Xa4*+*xa5*+*xa13*+*Xa21*[a]	China
Ilyou 218	2005	*Xa21*[a]	China
ZhongbaiYou 1	2006	*Xa21*[a]	China
Suweon 523	2011	*Bph18*	Korea

Modified from Brar and Singh [18]
[a]*Xa21* gene has also been transferred into many elite inbreds and parental lines of hybrids by several institutes in India, Philippines, and Thailand

4 Examples of Crosses Between Wild AA Genome Species and Cultivated Rice

To date many chromosome segmental substitution lines (CSSLs) or BILs in rice have and are being developed and new varieties are continuously being released at different research stations around the world (Table 2.3). Crosses between cultivated rice (*O. sativa*, $2n=24$, AA) and AA genome wild species can be easily made. Hybrids between *O. sativa* and *O. rufipogon* are partially fertile; however, *O. sativa*×*O. glaberrima* and *O. sativa*×*O. longistaminata* F_1s are highly sterile. Among the classical examples are the introgression of a gene for grassy stunt virus resistance from *O. nivara* to cultivated rice

varieties [40] and the transfer of a CMS source from wild rice, *O. sativa* f. *spontanea*, to develop CMS lines for commercial hybrid rice production [44]. Other useful genes, such as *Xa21* for BB resistance, were transferred into rice from *O. longistaminata*, and new CMS sources from *O. perennis* and *O. glumaepatula*. Genes for tungro virus tolerance and tolerance to acid sulfate soil conditions have been transferred from *O. rufipogon* into *indica* rice cultivars. Some of the breeding lines with genes introgressed from wild species have been released as varieties (Table 2.3). Ram et al. [55, 56] transferred broad spectrum blast resistance from *O. rufipogon* and also released a variety, Dhanarasi. Some of the alien genes have been tagged with molecular markers and used in MAS (Table 2.4).

At Kyushu University in Japan, Yoshimura et al. [74] developed a series of introgression lines using *O. glaberrima*, *O. glumaepatula*, *O. meridionalis*, *O. nivara*, and *O. rufipogon* accessions as donor parents in *O. sativa* cv. Taichung 65 background [20, 21, 43, 46, 62, 63]. Using these introgression lines, alleles associated with a number of desirable traits were identified, such as awn character [46], days to heading [59] and seed shattering [58], and green leafhopper resistance [23, 24].

In the USA, a University of Arizona, USDA, ARS, and University of Arkansas collaborative study is developing four BIL libraries using several *O. barthii* accessions as donors and *O. sativa* cv. LaGrue and M-202 as the recurrent parents (Eizenga and Sanchez, unpublished). These introgression lines will be tested for heat and drought tolerance in Arizona. At USDA, ARS, Stuttgart, AR, under the RiceCAP program, three BIL libraries are being developed using *O. nivara* and *O. meridionalis* as donors and *O. sativa* cv. Bengal and Lemont as the recipient parents. The introgression lines will be used to map sheath blight and blast resistance genes (Eizenga, personal communication). A collaboration between Cornell University, USDA, ARS Stuttgart, AR, and the University of Arkansas developed introgression lines using three diverse *O. rufipogon*/*O. nivara* accessions as donors and *O. sativa* cv. IR64 and Cybonnet as recurrent parents [67].

At Huazhong Agricultural University in China, backcrossing programs are underway to develop 14 CSSL/IL libraries using seven AA genome *Oryza* species accessions (6 wild and 1 *O. glaberrima*) that have BAC-end sequence as part of OMAP ([69]; Y. Sibin, personal communication) as donors and *O. sativa* cv. Zhenshan 97B and 93-11 as recurrent parents. The donor wild AA genome accessions include *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. nivara*, and *O. rufipogon*.

The CIAT/IRD (International Center for Tropical Agriculture and Institut de Recherche pour le Développement) rice genetics and genomics group lead a Generation Challenge Project (GCP) that is developing four libraries of CSSLs with the wild species *O. barthii*, *O. glumaepatula*, *O. meridionalis*, and *O. rufipogon* as donors, all sharing the same genetic background of the *tropical japonica* cultivar Curinga. The GCP-associated partners with this effort are Cornell University (USA), Fedearroz (Colombia), Embrapa-CNPAF (Brazil), and AfricaRice (Benin) (see http://www.generationcp.org/arm/ARM06/day_2/Lorieux_part_1.pdf; http://www.generationcp.org/arm/ARM06/day_2/Lorieux_part_2.pdf; [1]). Development of introgression lines from the *O. sativa* × *O. glumaepatula* interspecific cross [57] was also undertaken in the GCP initiative.

5 Introgression from *O. glaberrima* into *O. sativa*

Cultivars of Asian rice *O. sativa* are high yielding, whereas African rice, *O. glaberrima*, is low yielding. However, *O. glaberrima* has several desirable traits, such as resistance to rice yellow mottle virus (RYMV), African gall midge, and nematodes, and tolerance to drought, acidity, and iron toxicity. Another important feature of *O. glaberrima* is its strong weed competitiveness. Thus, interspecific hybridization among Asian and African species offers tremendous potential for combining the high productivity of *O. sativa* with tolerance to biotic and abiotic stresses of *O. glaberrima*. F_1 hybrids between *O. sativa* and