Available online at www.sciencedirect.com

SciVerse ScienceDirect

ELSEVIER

Current Opinion in Plant Biology

Genomic resources for functional analyses of the rice genome

Ying Yang, Yan Li and Changyin Wu

With the availability of the rice genome sequence, rice research communities are entering a new era of plant functional genomics. The last decade has seen rapid worldwide progress on establishing platforms for rice functional genomic research. These platforms offer practical toolkits and genomic resources for high-throughput identification of genes and pathways. In this review, we summarize available genomic resources for functional analyses of the rice genome. These genomic resources include high-quality bacterial artificial chromosome libraries, large-scale expression sequence tags, full-length cDNA collections, large amounts of data on global expression profiles, various mutant libraries and integrated bioinformatics databases. We not only present the current status of genomic resources but also discuss their usage in elucidating gene functions of the rice genome.

Addresses
National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China

Corresponding author: Wu, Changyin (cywu@mail.hzau.edu.cn)

Current Opinion in Plant Biology 2013, **16**:157–163

This review comes from a themed issue on **Genome studies and molecular genetics**

Edited by **Qifa Zhang** and **Rod Wing**

For a complete overview see the Issue and the Editorial

Available online 6th April 2013

http://dx.doi.org/10.1016/j.pbi.2013.03.010

Introduction

Rice (*Oryza sativa* L.) is an important agricultural staple that feeds more than half of the world's population. Because of the population explosion and cropland reduction, scientists have used rice biotechnology to propose the creation of green super rice (GSR), which has ideal traits including high and stable production yields, good grain quality and enhanced resistance to biotic and abiotic stresses [1]. In addition, rice is the model plant of monocot functional genomics. Hence, it is important to study the biological function of the rice genome and clarify the function of genes possessing important agronomic traits to assist in breeding GSR [2••].

In recent years, genomic resources for functional analyses of the rice genome have been developed rapidly. The completion of the high-quality rice genome sequence (International Rice Genome Sequencing Project, IRGSP, 2005) has greatly accelerated platform construction for rice functional genomics research, including high-quality bacterial artificial chromosome (BAC) libraries, large-scale expression sequence tags (ESTs), full-length cDNA collections, large amounts of data on whole-genome expression and various mutant libraries [3•,4,5••,6–9]. Many rice bioinformatics websites, accessed worldwide, have integrated these genomic resource databases. On the basis of this progress, an International Rice Functional Genomics Project (IRFGP) has been proposed with a goal to assign every gene's biological function by the year 2020 [2••]. In this review, we summarize the current status of genomic resources and their application in functional analyses of the rice genome.

Rice germplasm resources

Germplasm not only provides broad genetic diversity in rice breeding but also serves as a resource for gene identification in functional genomics. In the last century, the dwarf germplasm and cytoplasmic male sterile lines from *Oryza rufipogon* have been widely used to develop dwarf cultivars and hybrid rice combinations that contribute greatly to increased yield of rice, respectively. With the development of the next sequencing technology, it is feasible to exploit agronomically important genes by resequencing using genetically highly diverse germplasm, including some unique wild rice species.

The rice germplasm resource is extremely rich worldwide. Rice belongs to the genus *Oryza*, containing 23 species with 10 genome types, of which two are cultivated (*O. sativa* and *O. glaberrima*) and 21 are wild [10]. The Asian cultivated rice species *O. sativa* contains two subspecies: *japonica* and *indica*. The International Rice Genebank at the International Rice Research Institute (IRRI) in the Philippines holds the world's largest and most diverse collection of rice genetic resources. Most accessions are Asian cultivated rice varieties (*O. sativa*) or all known wild *Oryza* species, but significant collections of *O. glaberrima*, the West African cultigen, are also available. The West Africa Rice Development Association (WARDA) is responsible for conservation of the African cultivated rice and wild rice. The International Center for Tropical Agriculture (CIAT) is responsible for the rice germplasm of Latin America [11•]. IRRI conserves about 110 000 rice germplasm resources (http://irri.org/our-science/genetic-diversity-conservation). In China, over 70 000 varieties have been collected and cataloged (http://icgr.caas.net.cn/). To capture the largest amount of genetic diversity with a minimum number of accessions, a core collection comprising 932 accessions and a mini-core collection containing 189 accessions have been constructed [12,13]. In November 2011, the project of resequencing 3000 rice core collections for fully exploiting the elite genes was launched.

The genome sequences of the *japonica* variety Nipponbare and the *indica* variety 93-11 have been published and used extensively as reference genome sequences in rice [14,15]. With the development of new sequencing technology and availability of core collections, it is now feasible to establish an open-source translational research platform for genome-wide association studies (GWAS) in rice. GWAS have contributed tremendously to discovering genes controlling natural variations of complex traits [16•,17••].

cDNA, EST libraries and BAC libraries

The BAC libraries serve as a research platform for physical mapping, positional cloning, integration of genetic maps, comparative studies and whole-genome sequencing. High-quality BAC libraries of 12 *Oryza* species (11 wild *Oryza* species as well as the African cultivated species *O. glaberrima*) representing the 10 genome types were constructed for the Oryza Map Alignment Project (OMAP, http://www.omap.org) [7,18]. BAC-end sequences and fingerprints were produced from these BAC libraries, and phase I physical maps were constructed for the 12 *Oryza* species [19]. Recently, Lin *et al.* constructed a physical map of *O. sativa* ssp. *japonica* cv. ZH11 with ZH11 BAC clones. These data plus BAC-end sequence alignment are available on the Internet website (http://GResource.hzau.edu.cn) [20].

cDNA and ESTs provide verified and valuable information about gene structure and genome-coding capacity [21]. ESTs derived from various kinds of tissues, including tissues from organisms in a range of developmental stages or under stress, have been developed by several laboratories, and more than 1.2 million ESTs from rice have been registered in the NCBI GenBank (http://www.ncbi.nhm.nih.gov/dbEST/). These ESTs can also be used for new gene discovery, probe designs in microarrays and molecular markers. About 60 000 full-length cDNA clones are now available and the full-length cDNA libraries for both *indica* and *japonica* rice have been constructed [4,6,8,22]. The Rice FL-cDNA project was launched in 2000. More than 380 000 full-length cDNA clones from different tissues of *japonica* cv. Nipponbare with or without stress treatment were collected. Details of each clone are shown on the Knowledge-based Oryza Molecular Biological Encyclopedia (KOME, http://cdna01.dna.affrc.go.jp/cDNA/). Recently, the National Center for Gene Research (http://www.ncgr.ac.cn/ricd) developed information on approximately 21 960 unique 5′ ESTs and 10 081 full-length cDNAs isolated from *O. sativa* L. *indica* cv. Guang-lu-ai 4 and on approximately 27 130 unique 5′ ESTs and 12 727 full-length cDNAs isolated from *O. sativa* L. *indica* cv. Minghui 63 [6,8,22]. In addition, 2045 full-length cDNAs from *O. rufipogon* were isolated. These data are available worldwide to aid rapid discovery of new genes [23].

Large-scale gene expression profiles in rice

A major approach of cloning genes is the functional characterization of genes that show differential expression mostly from microarray analyses. In recent years, large amounts of data on rice whole-genome expression have been generated [24,25•,26]. The hybridization-based methods, such as microarrays and GeneChips, and the sequencing-based technology, such as serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS), have been established for acquiring large-scale gene expression profiles in rice [27,28]. The transcript abundance of genes in different

Table 1

Rice expression profiles databases

Library	Description	Database website	Subspecies
CREP	Rice expression profiles of tissues covering the entire life cycle	http://crep.ncpgr.cn	*indica*
Yale Virtual Center for Cellular Expression Profiling of Rice	Whole-genome transcriptional profiles of diverse cell types	http://bioinformatics.med.yale.edu/riceatlas/	*japonica*
RiceXPro	Rice expression profile database of whole developmental tissues	http://ricexpro.dna.affrc.go.jp/	*japonica*
RAD	Rice array database	http://www.ricearray.org/index.shtml	*indica* and *japonica*
RiceGE	Rice functional genomic express database	http://signal.salk.edu/cgi-bin/RiceGE	*indica* and *japonica*
RED	Rice expression database	http://red.dna.affrc.go.jp/RED/	*japonica*
RMOS	Rice microarray opening site	http://cdna01.dna.affrc.go.jp/RMOS/	*japonica*
Yale Plant Genomics	Gene expression from tiling path arrays	http://plantgenomics.biology.yale.edu/	*japonica*
Rice MPSS	Rice massive parallel signature sequencing gene expression database	http://mpss.udel.edu/rice/mpss_index.php	*indica* and *japonica*