
A Comprehensive Overview on Application of Bioinformatics and Computational Statistics in Rice Genomics Toward an Amalgamated Approach for Improving Acquaintance Base

Jahangir Imam, Mukesh Nitin, Neha Nancy Toppo, Nimai Prasad Mandal, Yogesh Kumar, Mukund Variar, Rajib Bandopadhyay, and Pratyoosh Shukla

Abstract

Rice (*Oryza sativa* L.) is a major crop in the world and provides the staple food for over half of the world population. From thousands of years of cultivation and breeding to recent genomics and systems biology approach, rice has been the focus of agriculture and plant research. Modern scientific research depends on computer technology to organize and analyze large datasets. Rice informatics – a relatively new discipline – has been developing rapidly as a subdiscipline of bioinformatics. Rice informatics devotes to leveraging the power of nature’s experiment of breeding and evolution to extract key findings from sequence and experimental data. Recent advances in high-throughput genotyping and sequencing technologies have changed the landscape of data collection

J. Imam

Biotechnology Laboratory, Central Rainfed Upland Rice Research Station (CRRI), Hazaribagh 825301, Jharkhand, India

Enzyme Technology and Protein Bioinformatics Laboratory, Department of Microbiology, Maharshi Dayanand University, Rohtak 124001, Haryana, India

M. Nitin • N.N. Toppo • N.P. Mandal • Y. Kumar
M. Variar (✉)

Biotechnology Laboratory, Central Rainfed Upland Rice Research Station (CRRI), Hazaribagh 825301, Jharkhand, India

e-mail: mukund.variar@gmail.com

R. Bandopadhyay

Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi 835215, Jharkhand, India

P. Shukla (✉)

Enzyme Technology and Protein Bioinformatics Laboratory, Department of Microbiology, Maharshi Dayanand University, Rohtak 124001, Haryana, India

e-mail: pratyoosh.shukla@gmail.com

and its analysis by using friendly database access and information retrieval. It focuses on developing and applying database tools and computationally intensive techniques and statistical software (e.g., pattern recognition, data mining, machine learning algorithms, R-statistical, MATLAB, and visualization) which give the opportunity to quickly and efficiently study heap of genomics information, chemical structure, and model generation study. Over recent years, various newly emerged diseases to rice varieties have an increasing concern to agriculturists and pathologists. The establishments of International Rice Information System, Rice Genome Research Project, Integrated Rice Genome Explorer, and Rice Proteome Databases are important initiatives for rice improvement using in silico software (e.g., homology modeling using SWISS Model, Modeler, and Autodock); the recent ongoing research on rice protein and its role in metabolic pathways works is being done around the world. Rice informatics has already started showing its profound impact on agricultural research and developments.

Keywords

Rice informatics • IRIS • Rice genomics resources • Rice proteomic resources • Databases • Rice bioinformatics software

Abbreviations

		MOsDB	MIPS Rice (<i>Oryza sativa</i>) database
		RiceGAAS	Rice Genome Automated Annotation System
RAP-DB	Rice Annotation Project Database	RED	Rice Expression Databases
ESTs	Expressed Sequence Tag	RMOS	Rice Microarray Opening Site
SAS	Statistical Analysis Software	RAD	Rice Array Database
IRRI	International Rice Research Institute	CREP	Collection of Rice Expression Profiles
ICIS	International Crop Information System	RAN	RiceArrayNet
WWW	World Wide Web	RGKbase	Rice Genome Knowledgebase
IRIS	International Rice Information System	MATLAB	MATrixLABoratory
GMS	Genealogy Management System	GRNN	Generalized Regression Neural Network
NCBI	National Center for Biotechnology Information	BPNN	Back Propagation Neural Network
EMBL	European Molecular Biology Laboratory	PDB	Protein Data Bank
EBI	European Bioinformatics Institute		
DDBJ	DNA Data Bank of Japan		
TIGR	The Institute of Genomic Research		
IRFGC	International Rice Functional Genomics Consortium		
GWAS	Genome Wide Association Study		
OMAP	Oryza Mapping Alignment Project		

1 Introduction

Rice is a major staple food crop for almost half of the world population. Among several agricultural crops, rice is considered as one of the most important crop plants for bioinformatics and computational biology research as it has become the model monocot plant having a number of

biological characteristics and recent research advancement in the field of genetics, breeding, genomics, germplasm collection and maintenance, systems biology, and functional genomics. During the last three decades, advancement in biotechnology led to the acceleration in many rice research programs particularly in breeding, selection of superior genotypes, large-scale cDNA analysis, genetic mapping, and genome sequencing (Khush and Brar 1998; Sasaki and Burr 2000). At the same time, this progression in biotechnology research paved the way for a new era, i.e., rice informatics in rice research, and opened new opportunities and direction for the improvement of rice crop which will address the issues concerning global problems on food security. The *japonica* rice cultivar Nipponbare genome sequencing project was completed in 2005 by consortium research of 10 countries, and Rice Annotation Project Database (RAP-DB) was developed to provide an accurate annotation of the rice genome through HTTP access (IRGSP 2005; Itoh et al. 2007). Parallel with rice genome sequence work and its related genomics resources, advancement in rice breeding research and development of molecular marker resources has helped the researchers to accelerate the identification, isolation, and incorporation of agronomically important genes and QTLs (Ashikari et al. 2005; Konishi et al. 2006; Ma et al. 2006, 2007; Kurakawa et al. 2007).

Recent advances in rice research are associated with the emergence of high-throughput data from large-scale sequencing, expression profiling of thousands of genes, phenotyping, and strategies on transcriptomics, proteomics, and metabolomics (Nagamura and Antonio 2010). In addition, large-scale collections of bioresources, such as mass-produced mutant lines and clones of full-length cDNAs and their integrative relevant databases, are now available (Brady and Provart 2009; Kuromori et al. 2009; Seki and Shinozaki 2009). The vast accumulation of genomics data from these strategies has culminated the need and importance of transforming these data into easily accessible and understandable form to the researcher, which can be ultimately studied and interpreted into useful biological information (Lewis et al. 2000). For this robust infrastructure

for organizing data, computational methods for analysis and interfaces for integration and retrieval of various types of data through user-friendly databases have been developed (Nagamura and Antonio 2010).

The application of bioinformatics has triggered the research in rice sciences with speed. This contributed to the easy and convenient way of data handling and data analysis much faster than traditional approach. The potential of the Internet in access of most up-to-date information on scholarly content, communication with colleagues, engaging two-way process of communication between researchers, and publishing materials more easily has been visualized in the advancement of the rice-related information and technologies (Ram and Rao 2012). Many bioinformatics resources are now available to the researchers around the world through the World Wide Web. Nowadays, researchers can easily post their research findings on the Web or compare their discoveries with previous results. The easy access and sharing of data between institutions has increased the opportunity for collaboration and thus dramatically fastened the research work and the development in the field of rice science. These developments are highlighted through the availability of databases, web servers, articles, and research organizations working in this area.

In the field of agriculture, the main focus is rice research, and the last 20 years belong to advancement of sequencing technology. In every stage of rice research, global problem of food security has been the burning issue, and every research activity from morphological and physiological to application of biotechnology to marker-assisted rice improvement to the development in bioinformatics technologies has been focused and addressed. The advancement in bioinformatics rice research played a pivotal role in this. Later, large-scale DNA analysis, genetic mapping, and genome sequencing have resulted in a tremendous increase in computer-generated information on rice genome (Sasaki and Burr 2000). During the same period, the advancement of sequencing technologies such as Expressed Sequence Tags (EST) research changed the path

of genetic expression of rice. EST project collectively represents about 1,251,304 entries of GenBank available on NCBI dbEST (http://ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) [GenBank Release 1.3.2011] (Ram and Rao 2012). In recent years, with the development of new software and statistical analyses, the physiology experiments of rice (*Oryza sativa* L.) were performed by analyzing ANOVA and Tukey's HSD mean comparison using Rv. 2.8.0 (Swamy et al. 2013). Rv. 2.8.0 is a free statistical software for computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows, and MacOS. MATLAB is a data analysis and visualization package software. Agriculture scientists use it for climate change analysis and in probabilistic model designing as it can cope up with large gridded dataset quite easily. Statistical Analysis Software (SAS) is also one of the widely used statistical packages for analyzing statistical data in crop science. Rice plants respond to different stresses via a number of mechanisms. Availability of rice genome sequences, large amount of information were generated from genomics and proteomics studies and in silico computational bioinformatics tools set a new platform for the management of environmental stresses in rice. In silico docking between the two proteins showed a significant protein–protein interaction between rice EDS1 and PAD4, suggesting that they form a dimeric protein complex, which, similar to that in *Arabidopsis*, is perhaps also important for triggering the salicylic acid signaling pathway in plants (Singh and Shah 2012).

2 Rice (*Oryza sativa*): Model Species for Monocot Plants

Poaceae family is one of the most important among the monocots which mostly includes agricultural crop species like maize, wheat, barley, sugarcane, sorghum, and rice. These species share extensive synteny across their genomes, allowing for one of the species to serve as the base for comparative genomics and bioinformatics study within the family (Moore et al. 1995).

Rice is one of the major monocot plants among them and used as a model species for the various biotechnological and bioinformatics research. Rice represents the most suitable species for genomics and bioinformatics research, which is the main reason for selecting it as a model species for its small genome size (~431 Mb). The second reason is the availability of genetic and molecular resources. After the complete genome sequencing of *japonica* rice variety Nipponbare, the research in rice has been revolutionized.

3 Rice Information System: The Rice Informatics

The computer and information technology has revolutionized the research in biotechnology and bioinformatics. Recent technologies in molecular biology and germplasm conservation have speeded up the sequence, genetic, and phenotypic information analysis (Ram and Rao 2012). The IRRI (International Rice Research Institute) in collaboration with CIMMYT (International Maize and Wheat Improvement Center) established the International Crop Information System (ICIS) project. In this project, scientists work for varietal improvement with the use of new and advanced bioinformatics software and also work for the development of software to facilitate and fasten the research and establishing the links between information from different crops like rice and maize. By this, the ICIS is maintaining huge datasets of rice which is publicly available and easily accessible to the scientists.

4 Web Tools and Resources on Rice

The results of huge data and information generated through acquiring knowledge of genomics resources in rice improvement has posed the problem in front of agricultural scientists to maintain data for future use and manipulate the data. Such activities have created the provision for the maintenance of data in the form of databases. Rice

research in global world and organizations working in the area of rice have developed and maintained such databases globally. Databases, software tools, web servers, etc. are available for data management and are used for solving problems related to rice, whether it would be the handling of molecular level activities or the production of a disease-resistant variety of rice (Ram and Rao 2012). The World Wide Web (www) provides a mechanism for extraordinary information sharing among the researchers as many bioinformatics resources are now available all over the world through www.

4.1 The International Rice Information System (IRIS)

The International Rice Information System (IRIS) is the rice implementation of the International Crop Information System (ICIS), a database system for the management and integration of global information on genetic resources and germplasm improvement (Bruskiewich et al. 2003). In 1995, the international agricultural research centers CIMMYT and IRRI partnered with other CGIAR centers to establish a project to develop an International Crop Information System (ICIS; Fox and Skovmand 1996) to overcome these deficiencies in crop data management for a wide range of crops. Several CGIAR centers, national agricultural research systems, and advanced research institutes are collaborating to develop ICIS as a generic system that will accommodate all data sources for any crop and breeding system. There are basically two objectives of ICIS: first, to integrate different data types in both private and public datasets into a single information system and, second, to provide specialist views and applications that operate on this integrated platform. After successful completion, ICIS will support a range of activities from germplasm conservation, evaluation, functional genomics, allele mining, breeding, testing, and release. The Genealogy Management System (GMS) of ICIS is the core database which ensures unique identification of germplasm, management of nomenclature (including homonyms and synonyms), and

retention of all germplasm development information. The ICIS system is fast, user-friendly, PC based, and is available on CD-ROM and also available online (<http://www.iris.irri.org/>). Mainly the ICIS (IRIS) is designed to allow biologists to manage local data and query and view their own data fully integrated with global public information. One of the innovative features of ICIS is that it permits independent users to integrate their own local data with public central data. The IRIS is being developed under the open-source ICIS project. Code is freely available to anyone, and the latest information about the ICIS project can be accessed at <http://www.icis.cgiar.org>.

4.2 Database Development

Computer-based databases are new innovations in the field of molecular biology, biotechnology, and bioinformatics, finding its scope in usage and online accessibility of information. The three major database sources at GenBank at the National Center for Biotechnology Information (NCBI) (<http://ncbi.nlm.nih.gov>), the European Molecular Biology Laboratory (EMBL) at European Bioinformatics Institute [(EBI) <http://www.ebi.ac.uk/>], and DNA Data Bank of Japan (DDBJ) (<http://ddbj.nig.ac.jp>) are imparting major roles in the management of biological information. The basic sequenced data submitted to these institutions can be mirrored to other institutions automatically on a routine basis to suffice for basic data. Beyond sequence data, the range of pertinent functional genomics, proteomics, structured experimental data, and associated data are being utilized by various organizations in order to develop different kinds of databases in various categories. A list of such databases in rice is shown in Table 1.

4.3 Structural and Functional Genomics Databases

Rice genomics databases are the major sources of information that could be used in understanding the genetic and molecular basis of all biological

Table 1 Major rice structural and functional genomics databases available

Sr. no.	Database	Description/major features	URL
1	DTRF	Database of Rice Transcription Factors	http://drtf.cbi.pku.edu.cn/
2	RAP-DB	<i>Oryza sativa</i> ssp. <i>Japonica</i> cv. Nipponbare genome sequence and annotation	http://rapdb.dna.affrc.go.jp/
3	Rice Genome Annotation	<i>Oryza sativa</i> ssp. <i>Japonica</i> cv. Nipponbare genome sequence and annotation	http://rice.plantbiology.msu.edu/
4	RISe	<i>Oryza sativa</i> ssp. <i>Indica</i> (93-11) and <i>japonica</i> (Nipponbare) sequence contigs and annotation	http://rice.genomics.org.cn/
5	OryzaSNP	SNP data from <i>Oryza</i> species and cultivars	http://oryzasnp.org/
6	Rice Haplotype Map	SNP data from 517 rice landraces	http://www.ncgr.ac.cn/RiceHapMap
7	Koshihikari genome	<i>Oryza sativa</i> ssp. <i>Japonica</i> cv. Koshihikari genome sequence	http://koshigenome.dna.affrc.go.jp/
8	OMAP	<i>Oryza</i> Map Alignment Project	http://www.omap.org/
9	dbEST	Rice ESTs	http://www.ncbi.nlm.nih.gov/
10	KOME	Japonica rice full-length cDNA sequences	http://cdna01.dna.affrc.go.jp/cDNA/
11	RICD	Indica rice full-length cDNA sequences	http://www.ncgr.ac.cn/ricd
12	Rice MPSS	Rice small RNA MPSS data	http://mpss.udel.edu/rice/
13	Rice PARE	Parallel Analysis of RNA Ends	http://mpss.udel.edu/rice/
14	Rice SBS	Rice small RNA SBS data	http://mpss.udel.edu/rice/
15	Indica MPSS	Indica Rice small RNA MPSS data	http://mpss.udel.edu/rice/
16	Tos17 Insertion Mutant Database	Tos17 insertion mutant panel	http://tos.nias.affrc.go.jp/
17	Rice Mutant Database	T-DNA insertion lines	http://rmd.ncpgr.cn/
18	TRIM	Rice insertional mutant lines	http://trim.sinica.edu.tw/index.php
19	SHIP	Shanghai T-DNA Insertion Population	http://ship.plantsignal.cn/index.do
20	Oryza Tag Line	T-DNA and Ds flanking sequence tags	http://urgi.versailles.inra.fr/OryzaTagLine/
21	IR64 Rice Mutant Database	Mutant lines for mutagenesis of IR64	http://irfgc.irri.org/cgi-bin/gbrowse/IR64_deletion_mutants/
22	Rice Array DB	NSF Rice Oligonucleotide Array Project	http://www.ricearray.org/
23	Rice Atlas	Rice cellular expression profile database	www.yalescientific.org/2010/02/the-rice-atlas/
24	OryzaExpress	Geo-based rice gene expression profiles	http:// bioinf.mind.meiji.ac.jp/OryzaExpress/
25	CREP	Collection of rice gene expression profiles	http://crep.ncpgr.cn/crep-cgi/home.pl
26	RiceXpro	Rice Expression Profile Database	http://ricexpro.dna.affrc.go.jp/
27	NIAS Genebank	30,000 rice accessions	http://www.gene.affrc.go.jp/
28	National Plant Germplasm	USDA germplasm database	http://www.ars-grin.gov/npgs/
29	IRIS	International Rice Information System	http://www.iris.irri.org/germplasm/
30	Gramene	Grass genome data anchored on rice	http://www.gramene.org/
31	Oryzabase	Oryza genetics database	http://www.shigen.nig.ac.jp/rice/oryzabase/
32	GreenPhyl	Phylogenomics database	http://greenphyl.cirad.fr/v2/cgi-bin/index.cgi

(continued)

Table 1 (continued)

Sr. no.	Database	Description/major features	URL
33	Rice Phylomics	Rice phylomics analysis data	http://phylomics.ucdavis.edu/
34	RKD	Rice Kinase Database	http://phylomics.ucdavis.edu/kinase/
35	Rice GT	Rice Glycosyltransferase Database	http://phylomics.ucdavis.edu/cellwalls/gt/
36	SALAD	Clustering of conserved amino acid sequences	http://salad.dna.affrc.go.jp/
37	Q-TARO	Rice QTL mapping information	http://qtaro.abr.affrc.go.jp/
38	Rice TOGO Browser	Integrated rice genomics browser	http://agri-trait.dna.affrc.go.jp/
39	INE	Integrated Rice Genome Explorer	http://ine.dna.affrc.go.jp/giot/
40	MOsDB	MIPS <i>Oryza sativa</i> database	http://mips.gsf.de/proj/plant/jsf/rice/index.jsp
41	OryGenesDB	Rice genes, T-DNA, and Ds flanking sequence tags	http://orygenesdb.cirad.fr/
42	Rice Annotation DB	Contig data for manual annotation of rice genome	http://ricedb.plantenergy.uwa.edu.au/
43	Rice pipeline	Unification tool for rice databases	http://cdna01.dna.affrc.go.jp/PIPE
44	Rice Proteome DB	Rice Proteome Database	http://gene64.dna.affrc.go.jp/RPD/main_en.html
45	RiceGAAS	Rice Genome Automated Annotation System	http://ricegaas.dna.affrc.go.jp/
46	Rice Genome Project/IRGSP	The International Rice Genome Sequencing Project (IRGSP), a consortium of publicly funded laboratories	http://rgp.dna.affrc.go.jp/E/IRGSP/
47	TIGR Rice	The TIGR Rice Genome Project BLAST, a server for collection of databases for use in searching with BLAST programs	http://blast.jevi.org/euk-blast/index.cgi/project=osa1
48	IRFGC	International Rice Functional Genomics Consortium	http://irfgc.irri.org/
49	RED	Rice Expression Database	http://red.dna.affrc.go.jp/cDNA/
50	Yale Plant Genomics	Yale plant genomics	http://plantgenomics.biology.yale.edu/
51	Geneinvestigator	Multi-organism microarray database and expression meta-analysis tool	http://www.geneinvestigator.com/
52	AgriTOGO	A database system for information and resources generated from various genome projects of the Ministry of Agriculture, Forestry and Fisheries, Japan	http://togo.dna.affrc.go.jp/
53	VanshanuDhan	Vanshanu rice gene database	http://www.nrcpb.org/

processes including many economically important traits, which are the main concern of breeders. The basic information that can be deciphered from the genome sequence is indispensable in the development of new cultivars with target traits such as high yield, biotic/abiotic stress resistance, good eating quality, etc. The availability of wide range of genetic and molecular markers made rice as an important species for genomics analysis. With the completion of

the rice genome sequence in 2004, a standard annotation is necessary so that the information from the genome sequence can be fully utilized and understood. This led to the establishment of a platform for structural and functional characterization of the rice genome. The Rice Annotation Project Database (RAP-DB) provides sequence and annotation data for rice genome. RAP-DB is a hub for *Oryza sativa* ssp. *japonica* genome information. This web-based tool provides

information on rice genome sequence from the Nipponbare subspecies of rice and annotation of the 12 rice chromosomes. The RAP-DB contains the IRGSP genome sequence (build 3 assembly) (IRGSP 2005) and the RAP loci with corresponding locus IDs representing the annotated genes. The primary concept of the RAP-DB is to provide simple access for the IRGSP genome sequence and the RAP annotation. RAP-DB has two different types of annotation viewers, BLAST and BLAT search, and other useful features. The Institute of Genomic Research (TIGR) works in the area of rice genome sequence with new data to improve the quality of the annotation. The TIGR Rice Genome Project BLAST server has a collection of databases for use in searching with the BLAST programs *blastn*, *blastx*, *tblastn*, and *tblastx*. The TIGR Rice Pseudomolecules database allows user to search against the latest version of the 12 TIGR rice pseudomolecules. These databases have been conceptualized with the aim of providing a comprehensive analysis of the rice genome and include both structural annotations to identify the genomics elements and functional annotations to attach biological meaning to the sequence data.

The International Rice Functional Genomics Consortium (IRFGC) has re-sequenced the 100 Mb of gene-rich genomics sequences to determine SNP variation from 20 diverse rice varieties and Landraces commonly used in breeding programs internationally and has the impressive genotypic and phenotypic diversity of domesticated rice (McNally et al. 2009). Zhao et al. (2011) showed the detailed results of a Genome Wide Association Study (GWAS) based on 44,100 SNP variants across 413 diverse accessions of *O. sativa* collected from 82 countries, which has been systematically phenotyped for 34 traits. SNP variation has been analyzed in 177 Japanese rice accessions, which are categorized into three groups: landraces, improved cultivars developed from 1931 to 1974 (the early breeding phase), and improved cultivars developed from 1975 to 2005 (the late breeding phase) (Yonemaru et al. 2012). *Oryza* Mapping Alignment Project (OMAP, Wing et al. 2005) has been initiated to

characterize the genome of wild rice species and already released the genome sequence of *Oryza glaberrima*, an African species of domesticated rice. Due to genome sequencing advancement during recent years, the researchers are able to access huge information on genetic variation present within the genus *Oryza*, within the two major subspecies, and among diverse rice cultivars and landraces that have great potential for improvement of cultivated rice (Nagamura and Antonio 2010). Along with complete genome sequencing, short ESTs (Expressed Sequence Tags), full-length cDNA sequence databases in dbEST, and KOME are also useful for evaluation of gene expression and variation. Furthermore, short read sequences including microRNAs (miRNAs), small interfering RNAs (siRNAs), transacting siRNAs, and heterochromatic siRNAs can be accessed via the rice databases at the University of Delaware (Simon et al. 2008). In particular, the Rice MPSS database is a repository of small RNA sequences with detailed information on sense and antisense expression of rice annotated genes (Kan et al. 2007). The genome sequences and all short transcript resources immensely augment the information on rice and will help in understanding the genetic control of agronomically important traits.

The MIPS Rice (*Oryza sativa*) database (MOsDB) provides a comprehensive data collection dedicated to the genome information of rice. MOsDB integrates data from two publicly available rice genomics sequences: *O. sativa* L. ssp. *indica* and *O. sativa* L. ssp. *japonica*. MOsDB provides an integrated resource for associated data analysis like internal and external annotation information as well as a complex characterization of all annotated rice genes. It includes an up-to-date access to publicly available rice genomics sequences and various search options. MOsDB is continuously expanding to include increasing range of data type and the growing amount of information on the rice genome. The RiceGAAS (Rice Genome Automated Annotation System) is also extensively used to identify various structural and functional components (Sakata et al. 2002). It has been developed to

execute a reliable and up-to-date analysis of the genome sequence as well as to store and retrieve the results of annotation. The RiceGAAS system does the functional analysis by collecting rice genome sequences from GenBank and then executing gene prediction, analysis of exons, splice sites, repeats, and transfer RNA based on algorithm which combines multiple gene prediction programs with homology search results. RiceGAAS system consists of 14 analysis programs. These include BLAST for homology search against protein database and rice EST database, GENSCAN and RiceHMM for gene domain prediction, MZEF for exon prediction, SplicePredictor for splice site prediction, and more (Sakata et al. 2002). Thus, RiceGAAS provides a systematic and comprehensive annotation of accumulated sequence data.

OryGenDB is another database for rice functional genomics. The database is an interactive tool for rice reverse genetics. Insertion mutants of rice genes are cataloged by flanking sequence tag (FST) information that can be readily accessed by this database. Oryzabase is an integrated database of rice genome resources. Oryzabase has been created for a comprehensive view of rice (*Oryza sativa*) as a model monocot plant by integrating biological data with molecular genomics information (<http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp>). The database contains information about rice development and anatomy, rice mutants, and genetic resources, especially for wild species of rice. Several genetic, physical, and expression maps with full genome and cDNA sequences are also combined with biological data in Oryzabase. This provides a useful tool for gaining greater knowledge about the life cycle of rice, the relationship between phenotype and gene function, and rice genetic diversity (Kurata and Yamazaki 2006).

Database has also been developed to screen and phenotype mutant lines in rice which can be used for selecting lines with specific morphological and physiological features. Some of these mutant databases are Rice Tos17 Insertion Mutant Database with about 50,000 Tos17 insertion mutant lines from *japonica* rice cultivar Nipponbare (Miyao et al. 2007), Rice Mutant Database

(RMD) containing 134,346 rice T-DNA insertion lines (Zhang et al. 2006), Taiwan Rice Insertional Mutants (TRIM) with 55,000 T-DNA insertion lines (Chern et al. 2007), Shanghai T-DNA Insertion Population Database (SHIP) containing 65,000 T-DNA insertion lines, Oryza Tag Line with 46,000 T-DNA and Ds insertion lines (Larmande et al. 2008), and IR64 Rice Mutant Database with phenotype information for irradiation and chemical mutants of IR64 cultivar (Wu et al. 2005). These databases mostly provide information on flanking sequences of the disrupted genes and thus allow users to screen the available mutants.

Another major part in bioinformatics is the analysis of huge amount of genome expression data, which is generated by microarray and SAGE (Gerstein and Jansen 2000). For analysis of such vast amount of data, many analysis tools and databases have been developed like Rice Expression Databases (RED) and Rice Microarray Opening Site (RMOS), Rice Array Database (RAD), Rice Atlas, OryzaExpress, Collection of Rice Expression Profiles (CREP), RiceArrayNet (RAN), RiceChip.Org, Rice MPSS, RicePLEX, MGOS (*Magnaporthe grisea*, *Oryza sativa*) database, and Rice Gene Expression Profile database (RiceXPro). These databases provide analysis tools that allow comparison of expression profiles from different samples based on specific criteria. Moreover, these databases provide Gene Expression Networks and various kinds of omics information including genome annotation, metabolic pathways, and gene expression.

For the better analysis of genome assemblies and annotation, a database, namely, Rice Genome Knowledgebase (RGKbase) – an annotation database for rice comparative genomics and evolutionary biology – has been introduced (Wang et al. 2012). RGKbase has three major components: (1) integrated data curation for rice genomics and molecular biology, (2) user-friendly viewers, and (3) bioinformatics tools for compositional and synteny analyses. Currently RGKbase includes data from five rice cultivars and species, and new datasets are continuously introduced in it. A very important genomics database, known as Rice TOGO

Browser, a component database of AgriTOGO, is an integrated database on rice functional and applied genomics. Rice TOGO Browser can be accessed through a user-friendly web interface that provides three search options, namely, keyword search, region search, and trait search, to retrieve information on specific genes, sequences, genetic markers, and phenotypes associated with a specific region of the genome (Nagamura et al. 2010).

4.4 Rice Proteomics Databases

As genome sequencing of rice has been completed, proteome analysis, which is the detailed investigation of the functions, functional networks, and 3-D structures of proteins, has gained increasing attention. Many rice proteomics databases are also available that are important resources for better understanding of protein functions in cellular system, protein-protein interaction, and downstream protein functions. Today, there are lists of proteome-related databases and tools available for rice proteomics analysis. This includes ExPASy Proteomics tools, Compute pI/Mw tool in ExPASy, ProteinProspector, pIans MW calculation service of aBi, SWISS-PROT and TrEMBL, ProteinProspector (UCSF), Rockefeller Univ Prowl (search engine), Mascot, EMBL, PeptideSearch, Swiss-protExPaSy, YPD (Proteome Inc.), and Sherpa. OryzaPG-DB, a Rice Proteome Database based on shotgun proteogenomics, incorporates the genomics features of experimental shotgun proteomics data. This version of the database was created from the results of 27 nanoLC-MS/MS and runs on a hybrid ion trap-orbitrap mass spectrometer, which offers high accuracy for analyzing tryptic digests from undifferentiated cultured rice cells. Approximately 3,200 genes were covered by these peptides and 40 of them contained novel genomics features. OryzaPG is the first proteogenomics-based database of the rice proteome, providing peptide-based expression profiles, together with the corresponding genomics origin, including the annotation of novelty for each peptide (Helmy et al. 2011).

The Rice Proteome Database is the first detailed database to describe the proteome of

rice. The Rice Proteome Database contains 23 reference maps based on 2D-PAGE of proteins from various rice tissues and subcellular compartments. These reference maps comprise 13,129 identified proteins, and the amino acid sequences of 5,092 proteins are entered in the database. Major proteins involved in growth or stress responses were identified using the proteome approach. Rice Proteome Database contains the calculated properties of each protein such as molecular weight, isoelectric point, and expression, experimentally determined properties such as amino acid sequences obtained using protein sequencers and mass spectrometry, and the results of database searches such as sequence homologies. The database is searchable by keyword, accession number, protein name, isoelectric point, molecular weight, and amino acid sequence or by selection of a spot on one of the 2D-PAGE reference maps (Komatsu et al. 2004). The information obtained from the Rice Proteome Database will aid in cloning the genes for and predicting the function of unknown proteins.

The important proteome databases and web resources popular in the rice research are listed in Table 2. These given databases are classified into different categories that have been used for rice proteomics databases comparison.

5 Statistical Rice Informatics Using R-Software

The new science of statistical rice bioinformatics aims to modify available classical and nonclassical statistical methods, develop new methodologies, and analyze the databases to improve the understanding of the complex biological phenomena. This interdisciplinary science requires understanding of mathematical and statistical knowledge in the biological sciences and its applications (Mathur 2010). Statistical programming language R was developed by R Development Core Team (2009), specifically for our purposes as shown in Fig. 1. This will solve the practical issues to follow the stream of reasoning. R can be used as a simple calculator and also for complex statistical analysis.

Table 2 Major rice proteomics databases available worldwide

Sr. no.	Database	Description/major features	URL
1	WORLD-2DPAGE	Index to 2-D PAGE databases and services	http://www.expasy.ch/ch2d/2d-index.html
2	SWISS-2DPAGE	Annotated databases of 2-D PAGE and SDS-PAGE	http://www.expasy.ch/ch2d/2d-top.html
3	YPM	Yeast Proteome Map	http://www.ibgc.u-bordeaux2.fr/YPM
4	YEAST 2D-PAGE	Database of Yeast 2D-PAGE	www.world-2dpage.expasy.org/list/
5	ECO2DBASE	2-D database of <i>E. Coli</i>	http://pcsf.brcf.med.umich.edu/eco2dbase
6	Sub2D	2-D database of <i>Bacillus subtilis</i>	http://pc13mi.biologie.uni-greifswald.de/
7	Cyano2Dbase	2-D database of cyanobacteria	http://www.kazusa.or.jp/tech/sazuka/cyano/teproteome.html
8	Abeerden 2-D db	<i>Haemophilus</i> 2-D database	http://www.abdn.ac.uk/-mmb023/2dhome.html
9	Fly 2-D db	Database of <i>Drosophila</i> genes and genomes	http://ty.cmb.ki.se/
10	ECO2DBASE	2-D database of <i>E. coli</i>	www.expasy.org/cgi-bin/dbxref?ECO2DBASE
11	HSC-2DPAGE	2-DE gel protein database at Harefield hospital	www.doc.ic.ac.uk/vip/hsc-2dpage/
12	SIENA-2DPAGE	2D-PAGE database from University of Siena, Italy	http://www.bio-mol.unisi.it/2d/2d.html
13	PHCI-2DPAGE	2D-PAGE database of parasite–host cell interaction	http://www.gram.au.dk/
14	Rice Proteome Database	Rice 2D-PAGE and proteomics database	http://gene64.dna.affrc.go.jp/RPD

R is free software and comes with absolutely no warranty. It also generates codes which are actively in use by researchers for crop science to evaluate their experimental data. A rice strip-plot experiment with three replications, variety as the horizontal strip, and nitrogen fertilizer as the vertical strip was calculated using R-Software as shown in Fig. 2 (Gomez and Gomez 1984).

```
dat<- gomez.stripplot
# Gomez figure 3.7
desplot(gen~x*y, data=dat,
out1=rep, num=nitro, cex=1)
# Gomez table 3.12
tapply(dat$yield, dat$rep, sum)
tapply(dat$yield, dat$gen, sum)
tapply(dat$yield, dat$nitro,
sum)
# Gomez table 3.15. Anova table
for strip-plot
dat<- transform(dat,
nf=factor(nitro))
```

```
m1 <- aov(yield ~ gen * nf +
Error(rep + rep:gen + rep:nf),
data=dat)
summary(m1)
>library(agridat)
>png(filename="gomez.
stripplot_%03d_large.png",
width=1000, height=800)
>### Name: gomez.stripplot
>### Title: Rice strip-plot
experiment
>### Aliases: gomez.stripplot
>
>### ** Examples
>
>
>dat<- gomez.stripplot
>
># Gomez figure 3.7
>desplot(gen~x*y, data=dat,
out1=rep, num=nitro, cex=1)
>
```

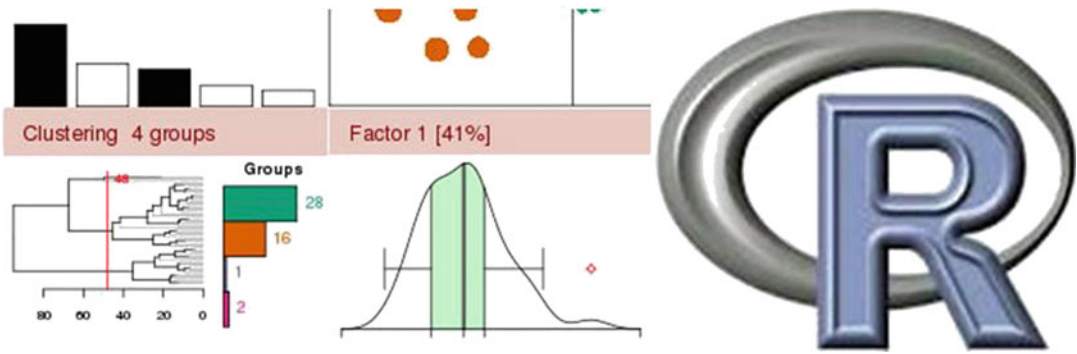


Fig. 1 R version 2.15.1 (2012-06-22). “Roasted Marshmallows.” Copyright (C) 2012. The R Foundation for Statistical Computing ISBN 3-900051-07-0. Platform: x86_64-unknown-linux-gnu (64-bit)

Fig. 2 Rice strip-plot experiment (Gomez and Gomez 1984)

	dat								
rep	1	3	2	3	2	1	3	1	2
gen	1	3	2	3	2	1	3	1	2
nitro	1	3	2	3	2	1	3	1	2
1 0	1	3	2	3	2	1	3	1	2
2 60	1	3	2	3	2	1	3	1	2
3 120	1	3	2	3	2	1	3	1	2

```

> # Gomez table 3.12
>tapply(dat$yield, dat$rep,
sum)
  R1  R2  R3
84700 100438 100519
>tapply(dat$yield, dat$gen,
sum)
  G1  G2  G3  G4  G5  G6
48755 56578 54721 50121 47241
28241
>tapply(dat$yield, dat$nitro,
sum)
  0  60  120
72371 98608 114678
>
> # Gomez table 3.15.Anova table
for strip-plot
>dat<-transform(dat,
nf=factor(nitro))
> m1 <- aov(yield ~ gen * nf +
Error(rep + rep:gen + rep:
nf), data=dat)
>summary(m1)

```

```

Error: rep
      Df SumSq Mean Sq F value Pr(>F)
Residuals 2 9220962 4610481

Error: rep:gen
      Df Sum Sq MeanSq F value Pr(>F)
gen  5 57100201 11420040  7.653
0.00337 **
Residuals 10 14922619 1492262
---
Signif. codes:  0 '***' 0.001
                '**' 0.01  '*' 0.05  '.' 0.1  ' ' 1

Error: rep:nf
      Df Sum Sq MeanSq F value Pr(>F)
nf  2 50676061 25338031  34.07
0.00307 **
Residuals 4 2974908 743727
---
Signif. codes:  0 '***' 0.001
                '**' 0.01  '*' 0.05  '.' 0.1  ' ' 1

Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
gen:nf  10 23877979 2387798
5.801 0.000427 ***
Residuals 20 8232917 411646
---
Signif. codes:  0 '***' 0.001
                '**' 0.01  '*' 0.05  '.' 0.1  ' ' 1
>dev.off()
null device
      1
>
Result

```

Thiel et al. (2009) used R-statistical package to test three regression methods: (i) linear least squares regression, (ii) local polynomial regression fitting using *loess* with the *degree* = 1 (the degree of polynomials) and *span* = 0.7 (the smoothing parameter), and (iii) local polynomial regression fitting with the same set of parameters as (ii) to normalize the observed map length differences of up to 24 % among chromosomes of both datasets using the previously defined anchor points in order to support the evidence

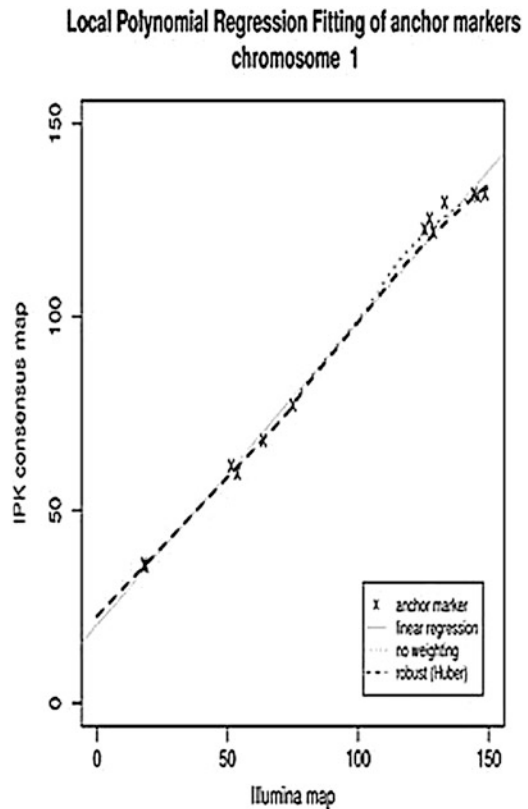


Fig. 3 Local polynomial regressions fitting of anchor markers of chromosome 1 using R-statistical package (Thiel et al. 2009)

and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice as shown in Fig. 3.

6 MATLAB Computing Tool in Rice Science

MATLAB stands for **MAT**rix**LAB**oratory, which is a state-of-the-art mathematical software package that integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Unlike other mathematical packages, such as MAPLE or MATHEMATICA, MATLAB cannot perform symbolic manipulations without the use of additional toolboxes. It remains, however, one of the leading software packages for numerical computation. MATLAB application is widely used in

The screenshot shows the RB-PRED web server interface. At the top, it says "RB - PRED" and "A SVM-based server for rice blast prediction dedicated to the farming community". The main form is titled "SUBMIT WEATHER CONDITIONS TO PREDICT RICE BLAST SEVERITY". It contains several input fields: Temperature (Max): 00.00, Temperature (Min): 00.00, Relative Humidity (Max): 00.00, Relative Humidity (Min): 00.00, Rainfall: 00.00, and Rainy days/week: 0. There is an "OR" section with "Upload from file:" and a "Browse..." button. Below that is a "Result display:" section with two radio buttons: "Cross-location prediction" (unselected) and "Cross-year prediction" (selected). At the bottom of the form are "Submit" and "Reset" buttons. The left sidebar has links: HOME, SUBMIT, ALGORITHM, DEVELOPERS, CONTACT, OTHER SERVERS at BIC, and MIRROR SERVER at CSK HPAU, PALAMPUR, INDIA. The right sidebar has an "IMAGE GALLERY" with four images of rice fields. The bottom navigation bar includes: HOME, SUBMIT, ALGORITHM, BIOINFORMATICS CENTRE, IITCC, TEAM, CONTACT, and OTHER SERVERS.

Fig. 4 An overview of submission form for online prediction of rice BLAST severity with “RB-Pred” web server (Kaundal et al. 2006)

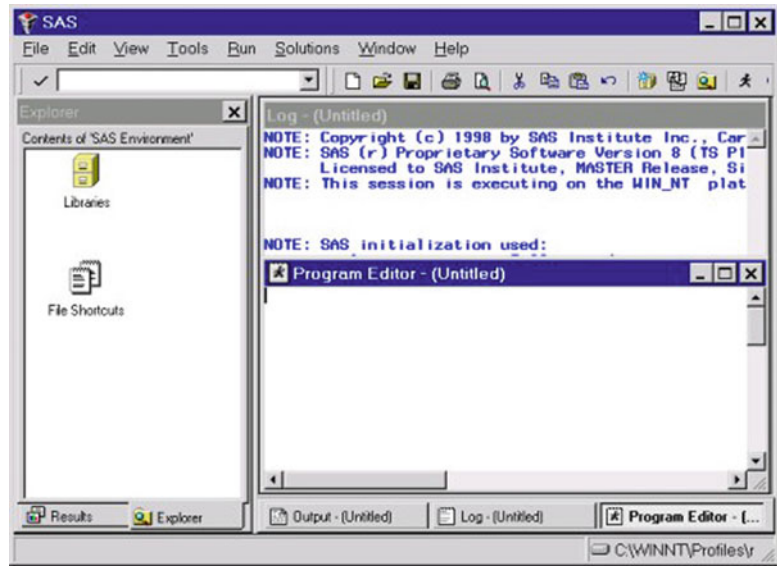
the field of rice sciences for designing probabilistic model to predict the time window of various rice diseases. The Generalized Regression Neural Network (GRNN) models were developed using the MATLAB (the MathWorks Inc., Natick, MA) software for designing a second type of neural network which do not require the optimization of multiple parameters that is required in feed-forward Back Propagation Neural Network (BPNN) and help in designing probabilistic model web server for forecasting rice BLAST disease prediction (Kaundal et al. 2006) as shown in Fig. 4.

7 SAS (Statistical Analysis Software) Application in Rice

SAS is an integrated system of software solutions that enables us to perform the various tasks like data entry, retrieval, management, report writing,

graphics design, statistical and mathematical analysis, business forecasting, decision support, operations research, project management, and applications development. SAS runs on IBM mainframes, Unix, Linux, OpenVMS Alpha, and Microsoft Windows. Code is “almost” transparently moved between these environments. Older versions have supported PC DOS, the Apple Macintosh, VMS, VM/CMS, PrimeOS, Data General AOS, and OS/2 as shown in Fig. 5.

SAS is also one of the widely used statistical software for analyzing statistical data in crop sciences. The statistical analysis that was carried out using SAS obtained the first three places, respectively, based on Version 8.2. A sample SAS program that can be used to superiority along with a high stability level (Table 3) while carrying out the analysis for three-month maturity group for Bg2845, Bg2834, and Bg300 obtained the first three places 2001/02 wet season

Fig. 5 SAS Windowing Environment (Cary 2001)**Table 3** ANOVA of split plot design for AVT-2 (BT) trial (Cyprien and Kumar 2012)

Source of variation	D.f.	ANOVA SS	Mean square	F value	Pr > F
N	2	0.08	0.04	2.54	0.19

(Samita et al. 2010). Varieties and their interactions were tested by using SAS software. ANOVA of SAS was used to display the results. Table 3 shows the results of ANOVA of split plot design for grain yield (kg/plot) in AVT-2 (BT) trial and observed significant variation in grain yield (Cyprien and Kumar 2012).

8 Computational Modeling in Rice Science

Protein–protein interaction is one of the crucial ways to decipher the functions of proteins and to understand their role in complex pathways at cellular level. The most accurate structural characterization of proteins is provided by X-ray crystallography and NMR spectroscopy. Due to certain technical difficulties and labor intensiveness of these methods, the number of protein structures solved by experimental methods lags far behind the accumulation of protein sequences. By the end of 2007, there were 44,272 protein structures deposited in the Protein

Data Bank (PDB) (www.rcsb.org) (Berman et al. 2000) – accounting for just one percent of sequences in the UniProtKB database (<http://www.ebi.ac.uk/swissprot>). The advancements in the field of bioinformatics have given us efficient tools to understand several biological processes at the molecular level. In recent times, significant progress has been made in computational modeling of protein structures and molecular docking, which holds great promise in prediction of protein–protein interactions. Docking is the computational scheme that attempts to find the best matching between two molecules: a receptor and ligand (Halperin et al. 2002). Protein–protein docking is one of the potential means to study the structure of protein–protein complexes such as antibody–antigen complexes (Gray 2006; Sivasubramanian et al. 2006; Sharma 2008). There are several docking software now introduced into the market such as GOLD, Autodock, Hex etc. Certain online server resources like SWISS Model (<http://swissmodel.expasy.org/>) and Modbase (<http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi>) are also being used

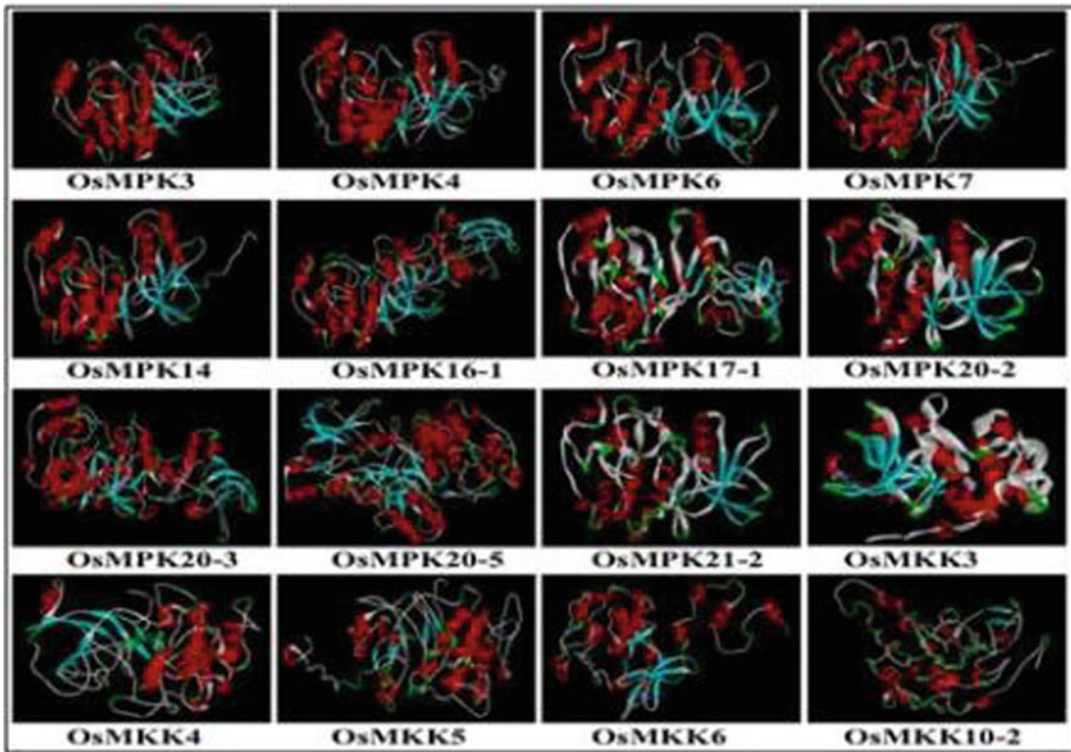


Fig. 6 Theoretical 3-D models of rice MAPKKs and MAPKs built by homology modeling. Structure of 11 rice MAP kinases (OsMPK3, OsMPK4, OsMPK6, OsMPK7, OsMPK14, OsMPK16-1, OsMPK17-1, OsMPK20-2, OsMPK20-3, OsMPK20-5, and OsMPK21-2) and 5 MAP

kinase kinases (OsMCK3, OsMCK4, OsMCK5, OsMCK6, OsMCK10-2) are shown. The *red* region represents the alpha helices and *sky blue* regions the beta sheets, *green* colored regions depict the turns, whereas the *gray* color represents the loops (Wankhede et al. 2013)

in homology modeling. Other structure validation programs such as Procheck, WHAT IF, VERIFY3D, and ERRAT were employed for further validation of models.

In the absence of crystallographic structures for rice MAPKs and MAPKK, the homology modeling approach was employed to determine a reasonable 3-D structure of these proteins based on the known structure of the template proteins. The 3-D structures were further used as an input for protein–protein docking using ZDOCK and RDOCK programs, to predict MAPKK–MAPK interactions. Simultaneously, Y2H analyses were used to study rice MAPKK–MAPK protein–protein interaction networks. A direct comparison of computational prediction and Y2H analyses of MAPKK and MAPK was made to assess the reliability of computational docking for prediction of

protein–protein interactions (Wankhede et al. 2013). All the 3-D models were refined with the help of loop refinement (Modeler and looper algorithm based) and side chain refinement protocols. The modeled 3-D structure of each of eleven MAPKs and five MAPKKs are shown in Fig. 6.

The overall stereochemical quality of the modeled 3-D structure of proteins was evaluated by using Ramachandran plot which is based on psi (Ca–C bond) and phi (N–Ca bond) angles of the protein and provides information about the number of amino acid residues present in allowed and disallowed regions. All the modeled proteins showed maximum residues in the most favored region followed by allowed region and least in the generously allowed regions in the Ramachandran plot (Fig. 7).

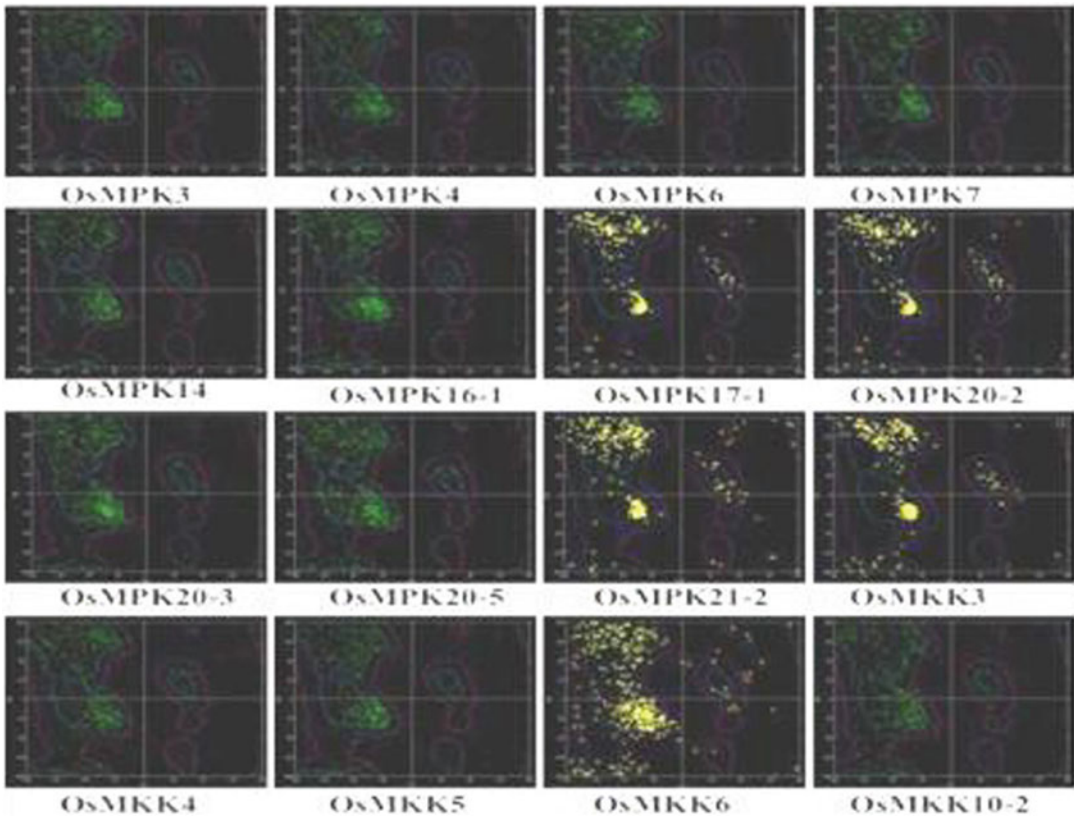


Fig. 7 Ramachandran plot analysis of theoretical 3-D structure of rice MAPKs and MAPKKs. The 3-D structures of 11 rice MAP kinases (OsMPK3, OsMPK4, OsMPK6, OsMPK7, OsMPK14, OsMPK16-1, OsMPK17-1, OsMPK20-2, OsMPK20-3, OsMPK20-5, and OsMPK21-2) and 5 MAP kinase kinases (OsMKK3, OsMKK4, OsMKK5, OsMKK6, OsMKK10-2) were validated using Ramachandran plot. The *green dots*

yellow dots show the amino acids that are in the most favored regions and additionally allowed region while *red dots* show the amino acids that are in generously allowed region or disallowed regions. The regions covered by *sky blue line* show most favored regions, while the regions covered by *pink line* show additionally allowed regions. Other regions of the plot show the generously allowed or disallowed region loops (Wankhede et al. 2013)

9 Conclusion

Rice is one of the economically important monocot crops in the world. The field of bioinformatics fulfills the needs and provides information on demand to the researcher in all fields of rice research. Bioinformatics is generally applicable to all branches of agriculture and is a boon for varietal improvement. It helps in understanding many agronomic traits involved in crop productivity. With the advancement in structural and functional genomics and proteomics rice databases, the research at both molecular and phenotypic level will advance further. Integrated Bioinformatics Information Resource Access

(iBIRA) is an initiative to associate bioinformatics researchers with bioinformatics resources at a single platform. Every day new databases, web server, and software tools are coming up to fulfill the need of the researchers. The challenge for rice informatics is how to translate it into a logical end.

References

- Ashikari M, Sakakibara H, Lin S, Yamamoto T et al (2005) Cytokinin oxidase regulates rice grain production. *Science* 309:741–745
- Berman HM, Westbrook J, Feng Z, Gilliland G et al (2000) The protein data bank. *Nucleic Acids Res* 28:235–242

- Brady SM, Provart NJ (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell* 21:1034–1051
- Bruskiewich RM, Cosco AB, Eusebio W et al (2003) Linking genotype to phenotype: the International Rice Information System (IRIS). *Bioinformatics* 19:163–165
- Cary NC (2001) Step-by-step programming with base SAS® software. SAS Institute Inc., Cary
- Chern C, Fan M, Yu S, Hour S et al (2007) A rice phenomics study-phenotype scoring and seed propagation of a T-DNA insertion-induced rice mutant population. *Plant Mol Biol* 265:427–438
- Cyprien M, Kumar V (2012) A comparative statistical analysis of rice cultivars data. *J Reliab Stat Stud* 5:143–161
- Fox PN, Skovmand B (1996) The International Crop Information System (ICIS)-connects genebank to breeder to farmer's field. In: Cooper M, Hammer GL (eds) *Plant adaptation and crop improvement*. CAB International, Wallingford
- Gerstein M, Jansen R (2000) The current excitement in bio-informatics-analysis of whole-genome expression data: how does it relate to protein structure and function. *Curr Opin Struct Biol* 10:574–584
- Gomez KA, Gomez AA (1984) *Statistical procedures for agricultural research*. Wiley Interscience, New York
- Gray JJ (2006) High resolution protein-protein docking. *Curr Opin Struct Biol* 16:183–193
- Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 47:409–443
- Helmy M, Tomita M, Ishihama Y (2011) OryzaPG-DB: rice proteome database based on shotgun proteogenomics. *BMC Plant Biol* 11:63
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436:793–800
- Itoh T, Tanaka T, Barrero RA, Yamasaki C et al (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res* 17:175–183
- Kan N, Venu RC, Cheng L, Belo A et al (2007) An expression atlas of rice mRNAs and small RNAs. *Nat Biotechnol* 25:473–477
- Kaundal R, Kapoor AS, Raghava GPS (2006) Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC Bioinformatics* 7:485. doi:10.1186/1471-2105-7-485
- Khush GS, Brar DS (1998) The application of biotechnology to rice. In: Ives C, Bedford B (eds) *Agricultural biotechnology in international development*. CAB International, Wallingford
- Komatsu S, Kojima K, Suzuki K, Ozaki K, Higo K (2004) Rice Proteome Database based on two-dimensional polyacrylamide gel electrophoresis: its status in 2003. *Nucleic Acids Res* 32:388–392
- Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. *Science* 312:1392–1396
- Kurakawa T, Ueda N, Maekawa M, Kobayashi K et al (2007) Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* 445:652–655
- Kurata N, Yamazaki Y (2006) Oryzabase. An integrated biological and genome information database for rice. *Plant Physiol* 140:12–17
- Kuromori T, Takahashi S, Kondou Y, Shinozaki K, Matsui M (2009) Phenome analysis in plant species using loss-of-function and gain-of-function mutants. *Plant Cell Physiol* 50:1215–1231
- Larmande P, Gay C, Lorieux M, Perin C et al (2008) Oryza Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library. *Nucleic Acids Res* 36:1022–1027
- Lewis S, Ashburner M, Reese MG (2000) Annotating eukaryote genomes. *Curr Opin Struct Biol* 10:349–354
- Ma JF, Tamai K, Yamaji N, Mitani N et al (2006) A silicon transporter in rice. *Nature* 440:688–691
- Ma JF, Yamaji N, Mitani N et al (2007) An efflux transporter of silicon in rice. *Nature* 448:209–212
- Mathur SK (2010) *Statistical bioinformatics: with R*. Elsevier, Boston
- McNally KL, Childs KL, Bohnert R et al (2009) Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci U S A* 106:12273–12278
- Miyao A, Iwasaki Y, Kitano H, Itoh JI, Maekawa M, Murata K, Yatou O, Nagato Y, Hirochika H (2007) A large-scale collection of phenotypic data describing an insertion mutant population to facilitate functional analysis of rice genes. *Plant Mol Biol* 63:625–635
- Moore G, Devos KM, Wang Z, Gale MD (1995) Grasses, line up and form a circle. *Curr Biol* 5:737–739
- Nagamura Y, Antonio BA (2010) Current status of rice informatics resources and breeding applications. *Breed Sci* 60:549–555
- Nagamura Y, Antonio BA, Sato Y, Miyao A, Namiki N, Yonemaru J, Minami H, Kamatsuki K, Shimura K, Shimizu Y, Hirochika H (2010) Rice TOGO Browser: a platform to retrieve integrated information on rice functional and applied genomics. *Plant Cell Physiol* 52:230–237
- Ram S, Rao LN (2012) Global information resources on rice for research and development. *Rice Sci* 19:327–334
- Sakata K, Nagamura Y, Numa H, Antonio BA et al (2002) RiceGAAS: an automated annotation system and database for rice genome sequence. *Nucleic Acids Res* 30:98–102
- Samita S, Anputhas M, De DS (2010) Selection of rice varieties for recommendation in Sri Lanka: a complex-free approach. *World J of Agric Sci* 6:189–194
- 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
- Seki M, Shinozaki K (2009) Functional genomics using RIKEN *Arabidopsis thaliana* full-length cDNAs. *J Plant Res* 122:355–366

- Sharma B (2008) Structure and mechanism of a transmission blocking vaccine candidate protein Pfs25 from *P. falciparum*: a molecular modeling and docking study. *In Silico Biol* 8:193–206
- Simon SA, Zhai J, Zeng J, Meyers BC (2008) The cornucopia of small RNAs in plant genomes. *Rice* 1:52–62
- Singh I, Shah K (2012) In silico study of interaction between rice proteins enhanced disease susceptibility and phytoalexin deficient, the regulators of salicylic acid signalling pathway. *J Biosci* 37:563–571
- Sivasubramanian A, Chao G, Pressler HM, Wittrup KD, Gray JJ (2006) Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis. *Structure* 14:401–414
- Swamy BPM, Ahmed HU, Henry A, Mauleon R, Dixit S et al (2013) Genetic, physiological, and gene expression analyses reveal that multiple QTL enhance yield of rice mega-variety IR64 under drought. *PLoS ONE* 8:e62795. doi:[10.1371/journal.pone.0062795](https://doi.org/10.1371/journal.pone.0062795)
- Thiel T, Graner A, Waugh R, Grosse I, Close TJ, Stein N (2009) Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. *BMC Evol Biol* 9:209. doi:[10.1186/1471-2148-9-209](https://doi.org/10.1186/1471-2148-9-209)
- Wang D, Xia Y, Li X, Hou L, Yu J (2012) The Rice Genome Knowledgebase (RGKbase): an annotation database for rice comparative genomics and evolutionary biology. *Nucleic Acids Res* 41:1199–1205
- Wankhede DP, Misra M, Singh P, Sinha AK (2013) Rice mitogen activated protein kinase kinase and mitogen activated protein kinase interaction network revealed by in-silico docking and yeast two-hybrid approaches. *PLoS ONE* 8:e65011. doi:[10.1371/journal.pone.0065011](https://doi.org/10.1371/journal.pone.0065011)
- Wing RA, Ammiraju JS, Luo M, Kim H et al (2005) The *oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol Biol* 59:53–62
- Wu J, Wu C, Lei C, Baraoidan M, Boredos A et al (2005) Chemical and irradiation induced mutants of *indica* rice IR64 for forward and reverse genetics. *Plant Mol Biol* 59:85–97
- Yonemaru J, Yamamoto T, Ebana K, Yamamoto E, Nagasaki H, Shibaya T, Yano M (2012) Genome-wide haplotype changes produced by artificial selection during modern rice breeding in Japan. *PLoS ONE* 7:e32982. doi:[10.1371/journal.pone.0032982](https://doi.org/10.1371/journal.pone.0032982)
- Zhang J, Li C, Wu C, Xiong L, Chen G, Zhang Q, Wang S (2006) RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Res* 34:745–748
- Zhao K, Tung CW, Eizenga GC et al (2011) Genome-wide association mapping reveals a rice genetic architecture of complex traits in *oryza sativa*. *Nat Commun* 2:467. doi:[10.1038/ncomms1467](https://doi.org/10.1038/ncomms1467)