Research Journal of Biotechnology Vol. **16 (3)** March **(2021)**
Res. J. Biotech

Review Paper:

Insight exploration of drought tolerant genes in *Oryza sativa*

Vasanthi S. and Febin Prabhu Dass J.*
Department of Biotechnology, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore (Tamil Nadu), 632014, INDIA
*mail2febin@gmail.com

Abstract

Rice (Oryza sativa) is a monocotyledonous plant belonging to the Poaceae grass family. Humans have domesticated this model plant species several million years ago as a major food crop. Now the crop accounts for about half of the global food production. The crop growth is highly affected due to exploitation of cultivable lands together with deterioration of environmental determinants. Both the biotic and abiotic stress press the need for exploration of the genomic content to develop a high yielding stress resistant strain. Although a good amount of progress has been shown in this field of study, a quick initiative could be started with the help of data exploration and analysis. Especially the known candidate drought resistance pathway genes are analyzed using the available rice databases and computational tools.

The drought resistant abscisic acid pathway genes are explored with respect to morphology, taxonomy, level of gene expression, transposons and proteome information with other interacting proteins, sub-cellular location with the evolutionary aspect. Rice drought tolerance gene information retrieved and interpreted from the different data resources could aid in bridging the unknown relationships existing at the different data level. This study may be further enhanced by the quality and quantity improvement of traits under various environmental stresses.

Keywords: Plant Molecular technique, Biotic and Abiotic stress, Drought, Abscisic acid, Traits, Environmental determinants, Data resources, Drought tolerance gene, Abscisic acid pathway genes, *Oryza sativa*.

Introduction

The entire world population depends on agriculture for their daily consumption and nutritional supplements. Humans obtain 85% of their calories from 20 plant species and 60% from wheat, rice and maize. Agriculture acts as a backbone of Indian economy which plays crucial role in the socio-economic sphere of the country[1]. It is now in the headlines because high food prices are increasing food insecurity and poverty. In the coming years, it is highly imminent to increase food productivity and production in developing countries[2]. The historical prominence of the grasses as an object of botanical research reflects their almost ubiquitous bio-geographical presence and their pervasive economic importance since the very beginning of human civilizations. Approximately one-third of the world's dry land is covered by species of Poaceae and the majority of the world's human population relies predominantly on cereal grasses such as rice, corn, or wheat for their daily consumption and agricultural productivity[3]. Of these domesticates, maize (*Zea mays*), rice (*Oryza sativa*) and wheat (*Triticum aestivum*) fall within the same Poaceae family[4].

However, taxonomical difference is observed between Panicoideae (maize), Bambusoideae (rice) and Pooideae (wheat) at the sub-family level[4]. In recent times, the grass family (Poaceae) has attained abundant focus in both scientific and phylogenetic analysis[3].

Rice (*Oryza sativa*) is one of the leading food crops in the world. As such, it is a staple food of over half of the world's population, mostly in Asia, Africa and South America[5]. Rice is the second most cultivated cereal after wheat[6]. Conventional breeding methods were used to select the genotype based on the expressed phenotype[7]. However, the current research focused on improvement of trait both qualitatively and quantitatively. Crop growth is highly sensitive to many biotic and abiotic factors wherein drought is a challenging yield limiting determinant for farming community[8]. Drought stress refers to insufficient water supply to plants in order to meet daily transpiration requirements that ultimately lead to many physiological responses such as leaf curling, narrowing of root[9].

Drought stress also makes numerous changes in the metabolic pathway and structural protein of the plant targeting down-regulation of functional gene responsible for the quantity and quality of traits[10]. Hence there is demand in the field of plant molecular biology and genetic engineering to isolate, down-regulate gene and bring attempt to improve their stress tolerance through gene transfer and regulate inducible gene[11]. The advancement of several plant molecular biology databases, especially for *Oryza sativa*, a multiple representation of biological data could well enhance the research towards achieving this desired goal.

The objective of plant biological databases is to store, organize and share sequence data which would enable researchers to retrieve, visualize, analyze and draw a worth full hypothesis towards genomic information[12]. Compilation of rice genome sequence information aims to identify genes, proteins responsive towards environmental stress both under molecular level of transcriptional factors[13].

There are more than 30 rice databases available with varied information about specific species, taxonomic classifications, morphology and gene expression at different developmental stages including transposable elements, motifs/domains for predicted gene of interest, biochemical pathway of functional gene, phylogeny among different species, gene families, candidate genes, proteomic information, gene expression from transcript profiling.

Databases

In this work, we collate and present biological databases of rice crop along with genes responsible for abscisic acid signaling pathway activating closure of stomata during drought condition. The abscisic acid pathway genes were explored under each differentiated sub-classes of databases which would include taxonomic classification of Poaceae family mentioning their tribes and sub-family, morphological and molecular level information to illustrate diagrammatically the stomata in open and close position pertain to drought due to abscisic acid synthesis, quantitative and qualitative trait locus information, interacting partner genes, phylogenetic tree of abscisic acid signaling genes present within grass family, pathway representation of genes involved in abscisic acid synthesis, comparison of gene expression level under drought and normal condition along with three normal expression levels of random housekeeping genes, transposons positioning in the genomic sequence of species under grass family, abscisic acid pathway signaling genes.

In this work we have collected six drought tolerant genes present in rice genome that helps in signaling of abscisic acid synthesis pathway during drought. These genes were retrieved from NCBI database for analysis and interpretation under each sub-class of collected and reviewed rice biological databases: (1) Taxonomy databases (2) Morphology databases (3) QTL databases (4) Proteome databases (5) Phylogeny databases (6) Pathway databases (7) Gene expression databases and (8) Transposon databases.

Taxonomy Databases: Plant taxonomic databases are central repository to detect nomenclature, classification and nucleotide sequence of plant species including Oryza sativa[14], which comes under Poaceae family. Hence, the taxonomic classification of Poaceae family is represented in fig. 1.

One of the well annotated rice genome database, Gramene Oryza database contains a clear interpretation of rice belonging to genus Oryza, tribe Oryzeae, subfamily Bambusoideae or Ehrhartoideae, family Poaceae or Gramineae [15]. Aside with the taxonomic information, 26 Oryza species genomic data is provided in Gramene database which is derived and updated from Plants Database (http://plants.usda.gov, 17 October 2006). Gramene database is cross linked to NCBI-*Oryza*, Germplasm Resources Information Network (GRIN) Taxonomy-*Oryza*. The above includes genetic structure and diversity in *Oryza sativa L*., population structure of African cultivated rice (*Oryza Glaberrima*), Wild Rice Taxonomy from IRRI.

Morphology Databases: Plant morphology analysis is 211 yr old originated by Goethe in 1790. It is a discipline that has largely been in practice. Plant morphology denotes both vegetative and reproductive plant parts including both root and shoot system (Er and Er 1997) to examine the developmental pattern of plants. Apart from the standard morphological features that differentiates between and within species, environmental factors in particular drought also could change morphological characters both external as well internal tabulated in table 1.

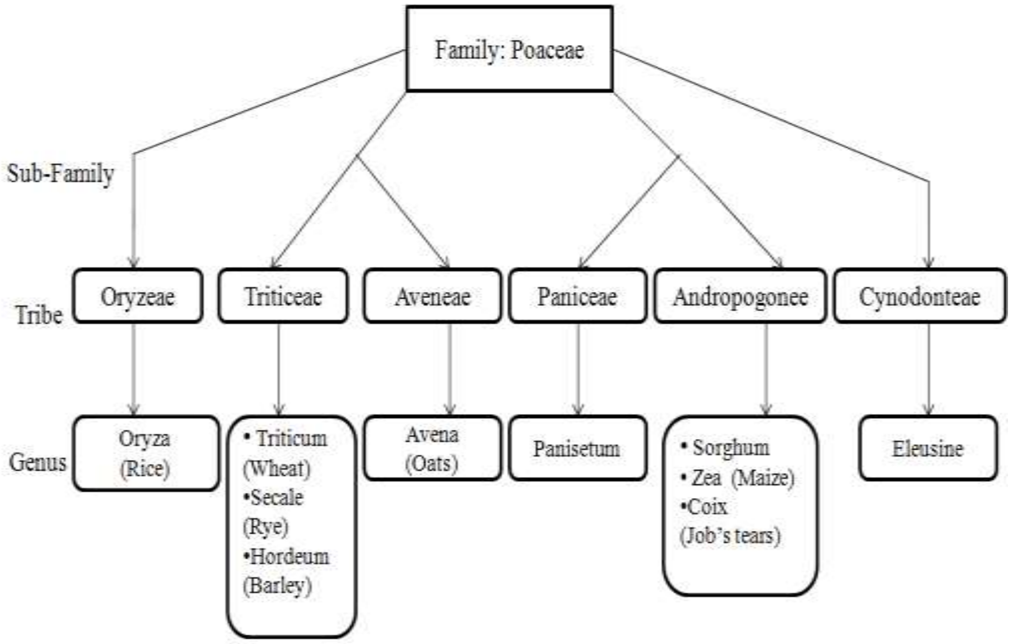

Fig. 1: Taxonomy of the Poaceae family

In specific, synthesis of abscisic acid in plant cell during drought enables closure of stomata and inhibits water loss within the cells through transpiration process which is depicted in fig. 2.

Morphological information carrying diagrammatic illustration of awn, rhizomes, ligules and rice cultivation manual involving protocol for seeding, germination, transplantation, artificial crossing, emasculation, pollination, harvest, seed preservation and map representation of rice crop distribution around the world is also provided in Oryzabase and Gramene databases.

Oryzabase is a unique rice database that shows morphological and gene expression characteristics of rice at different developmental stages and in various mutants. The biological information in Oryzabase is both unique and basic and it is composed of four sections: Development/Anatomy, Trait Genes, Mutants and Basic Biological Information [16].

Oryzabase has gathered information on nearly 20,000 useful strains. Design, data collection and construction of the database are organized by Rice Genetic Resources Committee together with the members of the Rice Gene Nomenclature Committee of Rice Genetics Cooperative (RGC). They also provide complete details on different rice varieties inclusive of their origin of cultivation[17] as mentioned in table 2.

In India, rice cultivation takes place densely in Southern part of India at Tamil Nadu State. Tamil Nadu Agricultural University (TNAU) maintains TNAUrice[39] database that consists of 250 rice records collected from Paddy Breeding Station (PBS) in Tamil Nadu. The varieties included in the database are from the following research stations located at 12 different places in the State of Tamil Nadu. All these rice varieties could be well understood by 29 descriptors such as name, parentage, duration, season, area of adoption, yield, habit, leaf sheath, axial, junction, auricle, ligules, septum, leaf blade, flag leaf, fertile glumes, apiculus, awns, panicle, exsertion, grain length and weight, rice color, milling, hulling percentage and protein content. The TNAU rice database was structured using Microsoft Structured Query Language (MS-SQL). The user interface of the database was enabled with Microsoft ASP.NET. However, genome sequence is yet to be updated.

Table 1
Morphological changes of rice crop under normal and drought condition

S.N.	Normal	Drought
1	Increase in Leaf area	Decrease in Leaf area
2	Healthy plant biomass	Reduction in plant biomass
3	Good yield and productivity	Reduction in yield
4	Elongated leaf and stem	Leaf rolling
5	Increase in plant height	Reduction in plant height
6	Increased panicle length	Reduced panicle length
7	Increased root elongation	Reduced root elongation

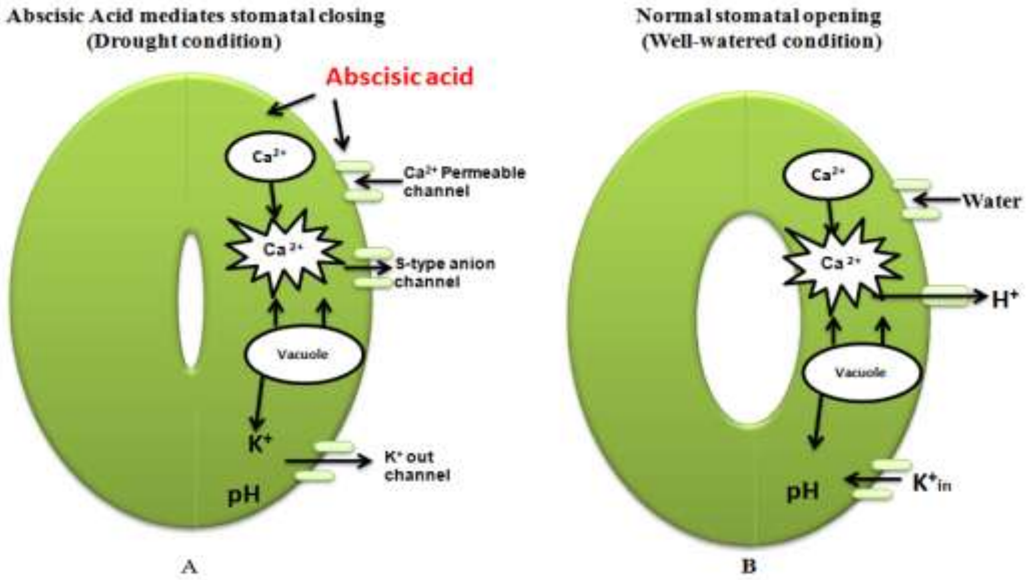

Fig. 2: (A) During water stress that stimulates abscisic acid production, K$^+$ and other solutes exit guard cells. This causes water to leave closing the pore.
(B) Water enters through osmosis by aquaporin channels, the water-swollen guard cells separate, opening the pores.

Table 2
Rice varieties and their origin of cultivation[23]

Rice Variety	Area of Cultivation
O.sativa, O.granulata, O.meyeriana, O.ridley	Asia
O.glaberrima	West Africa
O.rhizomatous	Sri Lanka
O.rufipogon	Tropic and Sub-tropic zones of Asia, Oceania
O.barthii	Deep-water area of African savanna
O.longistaminata, O.eichingeri, O.punctata, O.brachyantha	Africa
O.meridionalis, O.australiensis	Northern parts of Australia
O.glumipatula, O .latifolia, O.alta	Wet places of Central and South America
O.minuta	Philippines

QTL Database: QTLs originally reported on individual genetic maps that have been systematically aligned to the rice sequence where they can be searched as standard genomic features. Therefore, mutations identified in functional genomics population and association mapping panels can be aligned with QTL regions to facilitate fine mapping and validation of gene–phenotype associations. Further, QTL analyses also enable to identify the genetic basis of evolutionarily and ecologically relevant traits[18]. Under evolutionary studies, QTL of the actual genes clarifies the genetic architecture of traits that are related to fitness and that differ between recently derived species.

Gramene database[19] contains 6293 rice QTLs. These QTLs are anchored and directly associated with the markers mapped on the genome. MCDRP[20] consists of digitized experimental data for over 2400 rice proteins from 540 published research articles. More than 15,000 individual experiments containing over 90,000 data-points have been digitized with the help of manual data curation models. OGRO[21], a gene information table and a genome viewer has been designed as a web application. Similarly, Q-TARO[22] database consists of a relational database and two web applications, QTL Information Table and QTL Genome Viewer. The OGRO and Q-TARO web applications were implemented using Perl scripts and CGI modules with MySQL database application.

RiceQTLPro[23] were collected from the Korea rice (*Oryza sativa*) genome project of National Academy of Agricultural Science and the New Bio-Green 21 project. The platform was developed using MYSQL and Oracle relational database management system for its easy querying and speed of accessibility.

PlantGM [24] database is a web-based system that will provide information about genetic markers in rice and Chinese cabbage. PlantGM has three major constituents and functions namely map search, marker search and QTL search. The platform was developed using MYSQL in JAVA applications. Data was stored in an Oracle relational database management system (RDBMS). It shares rice QTL data that serves as a repository for international rice research community. The locus information of trait provides researchers to know the exact gene position responsible for both quantitative and qualitative trait improvement of plant crops. However, the mapped QTLs from the different research projects vary in the different mapping populations and it is difficult to utilize them for crop improvement.

Proteome Database: Plant proteins are sources of many peptides with diverse biological activity. Study on protein-protein interactions is critical element of biological systems and the analysis of interaction partners can provide valuable hints about unknown protein functions[25]. Understanding gene organization and protein interaction enables researcher to establish cellular contribution and their phenotypes[26].

Hence, compilation of protein information in biological databases plays an important key role in prediction of genomic function, system biology and for designing novel protein (Genomics, Design and Design 2004).

OryzaPG-DB contains peptides obtained from shotgun-based proteomics with the corresponding protein and transcript information. The OryzaPG web interface is developed using HTML, Java and PHP as server side script and implemented using the MySQL server. Other database containing such similar information is Gramene, where each protein is annotated with InterPro domain. RiceDB has been designed as a multilevel network with genomics to proteomics data that would aid the researchers in this field. The software and website have been developed using JAVA runtime environment and SCALA programming language with Lift web framework.

Rice kinase database consists of representative kinase from six kinase groups (STE, TKL, CMGC, AGC, CK1 and CAMK) that can be used to perform reiterative TBLASTN searches against two databases: National Center for Biotechnology Information and Knowledge-based Oryza Molecular biological Encyclopedia full-length cDNA database.

In addition to QTLs, MCDRP database also contains manually curated proteins from vast published literature.

Magnaporthe grisea Oryza sativa (MGOS) database contains experiments interaction data between *Oryza sativa* and *Magnaporthe grisea*, where the later species is a fungal pathogen that causes the rice blast disease. MGOS also use MySQL application with a CGI/Perl interface; Rice Genome Knowledgebase (RGKbase) provides user with proteome information specifically for rice obtained from Ensembl Plant Database.

Table 3 provides relating protein synthesized which would be responsible for drought tolerance along with their interacting gene and sub-cellular location.

Phylogeny Database: Phylogenies are important for addressing various biological questions such as relationships among species or genes, origin and migration patterns of species. Phylogenetic framework construction has explored the evolutionary patterns of many morphological and biochemical characters including complex pathways, defense and resistance mechanisms. A major goal is to nurture increased communication between phylogeneticists and physiologists/molecular geneticists/bioinformaticians by providing references for those requiring a phylogenetic backbone for analyses.

In such case, knowledge on rice phylogeny allows plant researchers to trace evolutionary histories of genes, estimate divergence times of genes or organisms. Few available computational databases provide best foundation for researchers to retrieve molecular data and build phylogenetic trees for functional predictions.

Gramene's Compara database[27] has new tools to generate gene trees and predict the orthologs and paralogs between five monocot genomes (*Oryza sativa japonica, Oryza sativa indica, Oryza glaberrima, Brachypodium distachyon* and *Sorghum bicolor*). In the current release of Gramene 2010 database[27], gene tree was rebuilt for four rice varieties (*Oryza sativa, Oryza japonica, Oryza sativa Indica* and *Oryza glaberrima*) to infer the ancestral gene and conserved region associated within them. As a part of phylogenetic analysis (Fig. 3), six genes (TRAB1, RAB21, LTP, OsPYL, SnRK, OsNPK1) responsible for drought resistant in Oryza species were retrieved to construct phylogenetic tree to know their sequence evolution, ancestral species and degree of closeness within grass family (Poaceae).

Pathway Database: The biological function of all living cells is a result of many interacting molecules[28]. Knowledge about these interacting molecules is necessary to computationally generate many species-specific pathways. Likewise, various pathway databases for rice have been generated and managed to provide necessary information on enzymes, compounds and genes involved in each specific reactions[29].

Table 3
Abscisic acid pathway genes and their interacting partner genes along with their sub-cellular location identified using CELLO software.

Drought Tolerant Genes				Interacting Genes			
Gene Name	**Description**	**Category**	**Sub-cellular location**	**Gene Name**	**Description**	**Category**	**Sub-cellular location**
bZIP23	Oryza sativa Indica Group cultivar IR20 bZIP23 gene	Promoter region	Cytoplasm	bZIP19	basic leucine zipper 19	Promoter region	Cytoplasm
TRAB1	Oryza sativa Indica Group cultivar IR-29 abscisic acid inducible bZIP protein	Inducing protein	Periplasmic Membrane	OsVP1	B3 domain-containing protein VP1	Inducing protein	Periplasmic Membrane
RAB21	Water-stress inducible protein	Inducing protein	Extracellular Membrane	OsAHL	AT-hook content nuclear localized protein)	Inducing protein	Extracellular Membrane
LTP	Lipid Transfer Protein	Lipid protein	Periplasmic Membrane	PAL	Peptidoglycan-associated lipoprotein	Lipid protein	Periplasmic Membrane
OsPYL	Oryza sativa cultivar Dong-jin cytosolic	Abscisic acid receptor	Cytoplasm	OsPP2C	Protein phosphatase 2C	Abscisic acid receptor	Cytoplasm
SnRK2	Oryza sativa Indica Group cultivar Nagina 22	CAMK like protein	Cytoplasm	ABI1	Abscisic acid interactor protein	Interactor protein	Cytoplasm

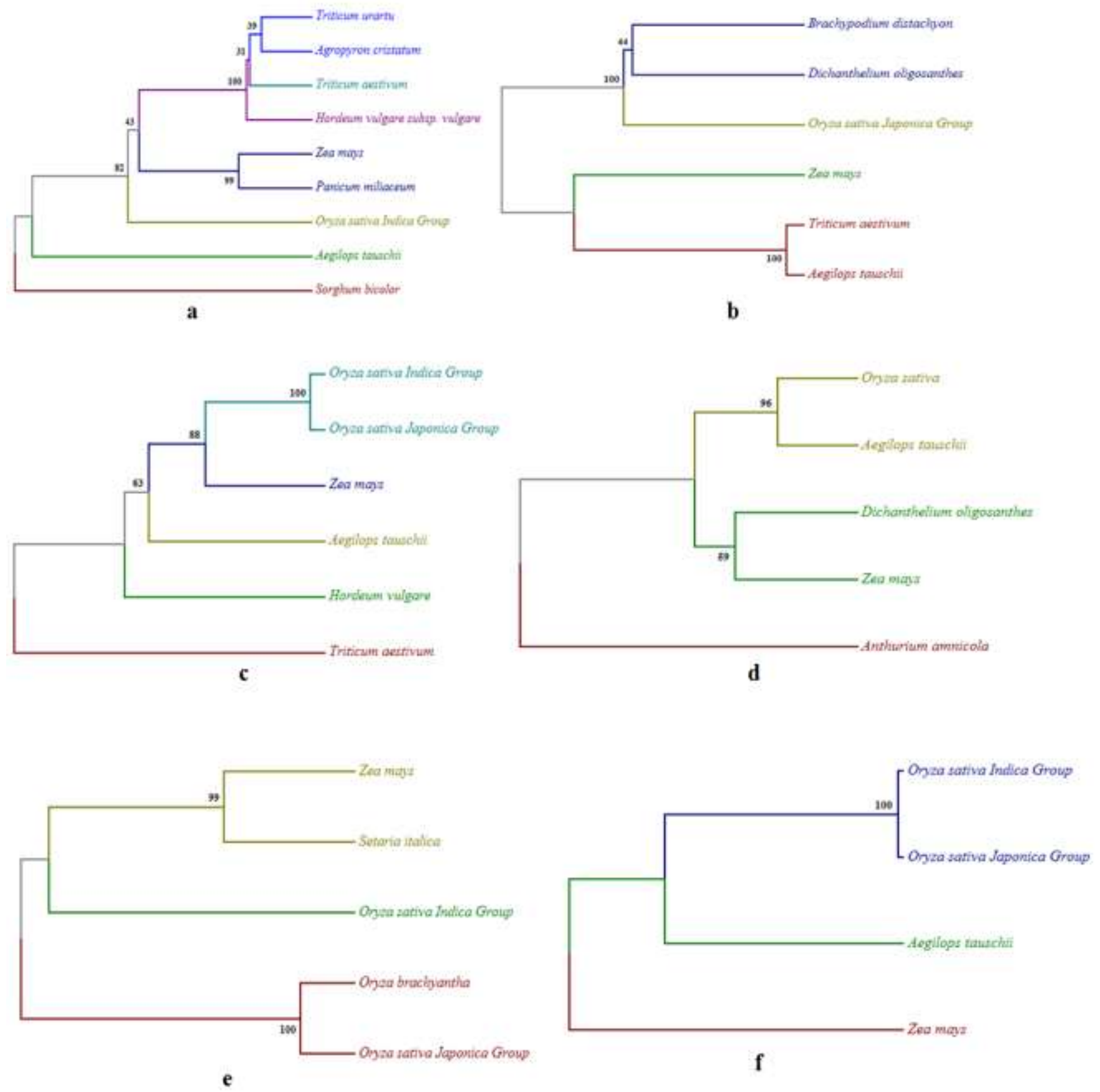

Fig. 3: (a) Phylogeny for TRAB1 gene in poaceae family, (b) Evolutionary analysis for RAB21 gene present in poaceae family, (c) Phylogenetic tree generated for LTP gene in poaceae family, (d) Phylogeny for OsPYL gene in poaceae family, (e) Phylogeny for SnRK gene among poaceae family and (f) Evolutionary analysis for OsNPK1 gene present in poaceae family.

In Gramene, the pathway information linked to Metacyc and RiceCyc contains 282 pathways, 1529 reactions, 43 172 polypeptides (10 387 enzymes) and 1162 compounds. Similarly, the MCDRP (Manually Curated Database of rice proteins) carries more than 200 metabolic pathways with the current release consisting of experimental data of 2400 rice proteins from several research articles[30]. In case of Rice Genome Knowledgebase (RGK) pathway annotations for multiple functional genomes are done[31]. RGKbase current includes data from five rice cultivars and species: Nipponbare (japonica), 93-11 (indica), PA64s (indica), the African rice (*Oryza glaberrima*) and a wild rice species *(Oryza brachyantha).*

Candidate genes OsPYL, PP2C, SnRK2 focused in this analysis were found responsible for abscisic acid signaling pathway during drought condition using KEGG database[35] wherein, other genes KUP6, SLAC1, QUAC1, KAT1, AtrbohF, ABF, AKS are also involved in this signaling mechanism which together enables closure of stomata. Fig. 4 shows genes responsible for the signaling pathway of abscisic acid.

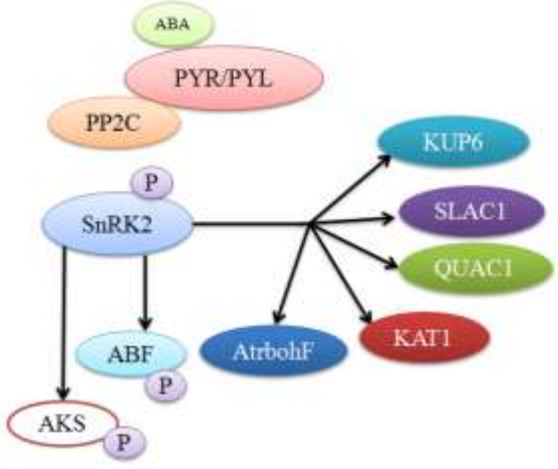

Fig. 4: Genes involved in abscisic acid signaling pathway (KEGG pathway database, Ruth Finkelstein 2013)

Gene Expression Databases: In the late 90's and 2000's, DNA array technology progressed rapidly. In addition to increase in the knowledge on DNA sequences of multiple genomes, raw information is necessary to assure that arrays could be made to represent each gene in a whole genome [32]. Microarray studies through the application of genetic and molecular biology information have allowed researchers to study and represent arrays as each single gene from a whole genome sequence, also to discover gene expression evolution and its levels, key players in metabolic pathways and to assign probable function to genes[32].

Microarrays are an attractive platform for genotyping applications in plant genetics and crop improvement[33]. It provides knowledge on up-regulated and down-regulated gene during any specific biotic and abiotic stress[34]. The advancement in microarray technology permits researchers to monitor the expression profile of thousands of genes simultaneously making it as an indispensible tool to study transcriptome evolution. The sensibility behind these studies is that orthologous gene fetches similar physiological functions that depict conserved expression profiles.

Analysis of genes enables control over genetic information that is turned into various biological products such as therapeutic proteins. Hence, molecular breeders focused this area to store, organize and manage gene expression data of various plants including Oryza species. Numerous Oryza gene expression database available online such as OryzaPG followed the usual standards of web application development and the Java scripts employed are cross-browser scripts[34].

OryzaPG-DB contains peptides obtained from shotgun-based proteomics with their product ion spectra as well as updated annotations, side by side with the corresponding protein, cDNA, transcript and genomic sequences and information. RiceDB provides useful ways to examine transcript expression patterns in rice. Additionally, hundreds of microarrays were analyzed in-house and expression information was collated to indicate whether or not gene(s) are expressed[35].

RiceXpro was initially conceptualized with the aim to provide a platform for monitoring gene expression of the rice plant under the natural field conditions. The first release (version 1.5) contained two data sets in the field/development category corresponding to gene expression profiling based on 48 different tissues and organs at various developmental stages and continuous profiling of mature leaves from transplanting until harvesting[36]. Hence, the updated release includes a total of 530 microarray data representing 23 data sets.

RiceArrayNet provides information on co-expression between genes in terms of correlation coefficients (r values) using the accumulated data and expression data from 183 microarrays that are obtained from the samples of either wild-type or mutant rice organs such as leaf, root, flower and callus at various developmental stages.

RAP-DB contains large number of rice-expressed sequence tags and functional genomics resources which have been produced worldwide. The team had thoroughly updated the genome annotation by manual curation of all the functional descriptions of rice genes. The latest version of the RAP-DB contains a variety of annotation data as follows: clone positions, structures and functions of 31,439 genes validated by cDNAs, RNA genes detected by massively parallel signature sequencing (MPSS) technology and sequence similarity, flanking sequences of mutant lines and transposable elements[37].

Rice Oligonucleotide Array Database (ROAD) is a useful reference data for elucidating rice gene expression and function. A total of 105 rice microarray experiments (1,867 hybridizations) were collected from NCBI GEO and annotated here. This database was constructed with PHP and MySQL and run on Windows server[38].

Rice FREND consists of co-expression analysis of 27201 genes performed against 815 microarray data derived from expression profiling of various organs and tissues at different developmental stages. It serves as a platform for efficient prediction of gene functions and gene regulatory networks based on gene expression similarity[39]. MGOS (*Magnaporthe grisea Oryza sativa*) web-based database contains data from *Oryza sativa* and *Magnaporthe grisea* interaction[40] allows researchers to explore gene expressions (Table 4). MGOS use MySQL databases with a CGI/Perl interface. PAVE and the Genome Browser are separate MySQL databases for modularity, though they work seamlessly with the main MGOS database.

Expression Atlas database provides gene and protein expression data in different organism parts during their developmental stages, diseases and other experimental conditions. The database consists of high quality micro array and RNA-sequencing experiments from Array Express database that have been manually curated. Knowledge on gene expression data would provide researchers to gain insights towards biological mechanisms[41].

Comparative analysis was done for six abscisic acid synthesis pathway gene using Expression Atlas database. Three genes out of six and their concerned expression values were documented in the database. Hence, those three abscisic acid signaling pathway genes RAB21[42], LTP[43], OsNPK1[44] were taken for this analysis to retrieve their expression values under drought and normal condition. We also selected three random housekeeping genes albumin, β-tubulin (*TUB*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*)[45], along with their normal expression values and generated a histogram (Fig. 5) based on these values for better understanding for this comparison.

Hence, this showed increase in the level of expression in RAB21, LTP and OsNPK1 genes during drought than in normal condition.

Transposon Databases: Transposable elements (TEs) are mobile DNA elements frequently found in plant genome and are suggested to play an important role in evolution, mutations, gene expression, chromosome rearrangement[46] and transposition mechanisms. TEs are classified into two classes based on their mode of transposition; class I elements are also known as retrotransposons that propagated from RNA intermediates and class II elements include DNA transposons[47].

Therefore, TEs have been known to be a major area of interest worldwide in order to track the evolutionary history of an organism[48]. Many studies have disclosed that TEs are preferably useful tools that engage in various processes counting genome evolution, gene expression regulation and genetic instability.

In rice, efforts to annotate and distribute repetitive DNA sequence led to discovery of more specialized database such as BGI-RIS[49] is a highly integrated resource database for rice data storage, retrieval, visualization and analysis and has several secondary databases such as the repeat databases for identification of repetitive sequences and tagging repeat classes of the grass family.

Its back end is an Oracle9i relational database and the front end consists of a set of JSP scripts running on TomCat web server. Rice transposable Elements database[8] is set up on a Linux-Apache-MySQL-Perl (LAMP) system, JavaScript libraries including jQuery (1.10.2), jQueryUI (1.10.4) and some additional plugins to perform dynamic web services.

Table 4
List of rice genome expression databases

S.N.	Databases	Description
1.	DRTF	Information about rice transcriptional factor[19]
2.	MGOS	Data from *Oryza sativa* and Magnoporthe (fungal pathogen) that causes rice blast disease.[44]
3.	RAP-DB	Browses gene, marker, protein, restriction sites for 12 *Oryza* species
4.	Rice Array Net	Provides co-expression information between genes in terms of correlation coefficients from microarray data
5.	OryzaPG	Retrieve gene, protein, cDNA and transcript expression for individual chromosome.[20]
6.	ROAD	Facilitates the extraction of gene expression profiles and identification of microarray element.[7]
7.	Rice DB	Integrates and compares rice with Arabidopsis thaliana based on gene expression[34]
8.	RiceXpro	Comprehensive information on transcriptome of rice encompassing the entire growth cycle under various experimental conditions[43]
9.	Rice FREND	Gene function prediction and regulatory networks based on gene expression data[42]
10.	CARMO	Intensive functional study in rice genome using gene expression level[51]

CARMO: Comprehensive Annotation of Rice Multi-Omics database, **MGOS**: Magnoporthe grisea Oryza sativa,
ROAD: Rice Oligonucleotide Array Database.
RiceFREND: Rice Functionally Related gene Expression Network Database, **RAP-DB**: Rice Annotation Project Database, **DRTF**: Database of Rice Transcriptional Factor, **Rice DB**: Rice Database, **Oryza PG**: Oryza Proteogenomics Database.

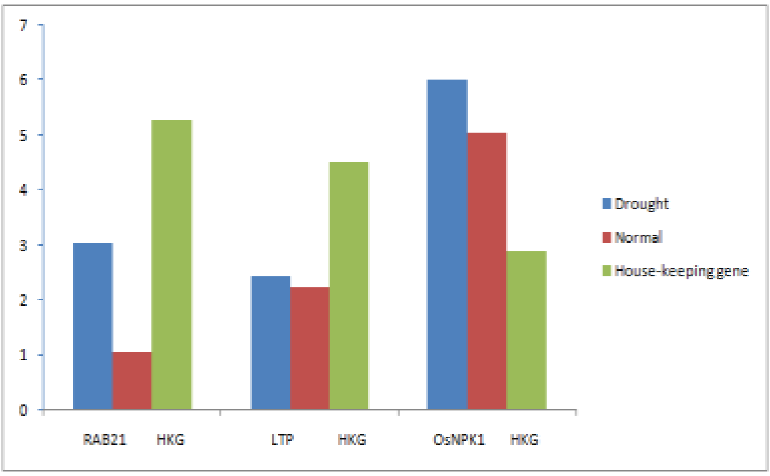

Fig. 5: Gene expression level comparison between abscisic acid synthesis pathway genes during drought and normal condition along with randomly selected house-keeping gene

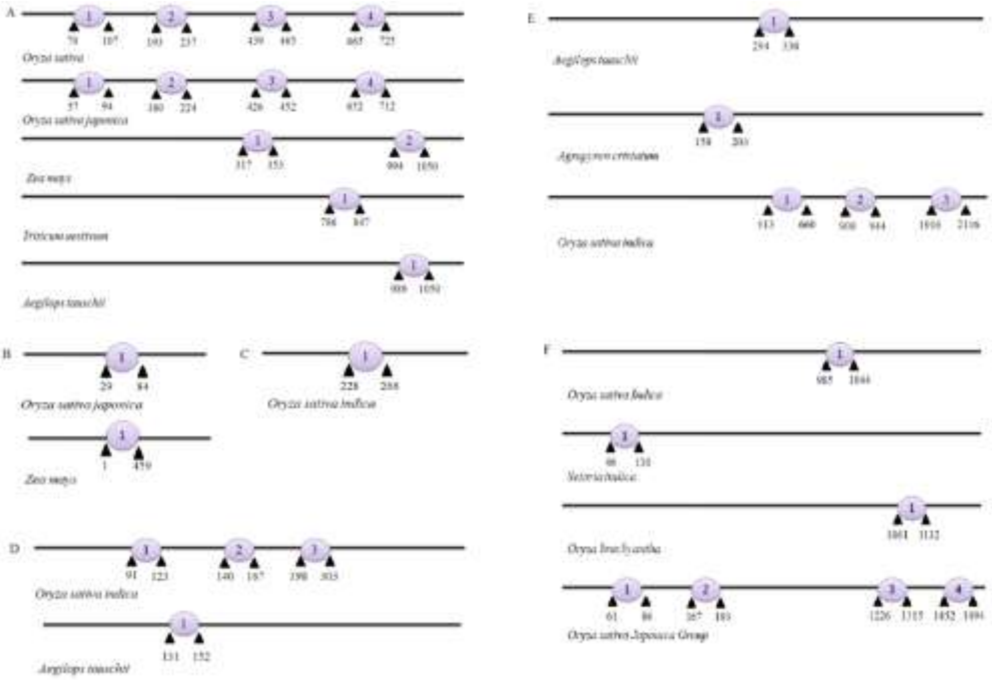

Fig. 6: (A) Transposons positions in TRAB1 gene within Poaceae family, (B) Transposon positions in OsNPK1gene of *Oryza sativa* japonica and *zea mays* genome, (C) Transposon positions in LTP gene present in *Oryza sativa* indica genome, (D) Transposon positions in OsPYL gene present in *Oryza sativa* indica, *Aegilops tauschii* genome, (E) Transposon positions in RAB21 gene of *Aegilops tauschii, Agropyron cristatum, Oryza sativa* indica and (F)Transposon positions in SnRK gene of *Oryza sativa* Indica, *Setaria italic, Oryza brachyantha* and Oryza *sativa* japonica.

In the current version it contains 265,549 Transposable elements and repeat sequence entries from 11 Oryza and one out-group species and RGKbase[50] implemented Linux-Apache-Mysql-Php infrastructure to realize dynamic invocation and optimization and HTML/javascript to write the static and simple user-responsive webpage.

Drought responsive abscisic acid signaling pathway genes TRAB1, OsNPK1, LTP, OsPYL, RAB21, SnRK2 under the poaceae family were analyzed for the presence of transposons. The transposons locations were identified within each of the above mentioned genes using REPEATMASKER[51]. These results were depicted in fig. 6 which includes transposon position mentioned in numeric format that would serve researchers to focus on specific positions of mutation and number of transposons present within each selected genes that has high or low copy number variation with evolutionary significance[52]. Transposons contribute to the development of gene diversity in natural population to shape the plant genome.

Conclusion

Rice crop satisfies nearly half the world's population who are dependent on this major crop for daily food consumption and as nutrition. Between the biotic and abiotic stress, the crop production is majorly affected by drought. The advent of cutting edge molecular biology techniques and deposition of the experimental information in data repositories maintained globally draws information exploration. The drought tolerance abscisic acid signaling pathway genes across the Oryza species and within the poaceae family are analyzed using the following databases: taxonomy, morphological database information for the drought genes, significance of QTLs, interacting genes and their pathway, evolutionary relationship of the selected genes within grass family (poaceae), level of gene expression at normal and drought time from expression database and transposons present within these genes.

The above analysis may be applied in plant breeding and has evolutionary implication on yield productivity under severe stress condition like drought. This cumulative exploratory analysis on drought resistance pathway genes from different data systems present in rice databases provides researchers a strong platform and a novel insight to link the rice crop genome completely from species to gene expression level during biotic, abiotic stresses and make them tolerable towards various crop growth limiting factors.

Acknowledgement

Authors would like to acknowledge the facilities provided by Vellore Institute of Technology for carrying out this work.

References

1. Anderle Pascale, Gene Expression Databases and Data Mining, *BioTechniques*, **34**, S36-S44 **(2003)**

2. Athews S., Arah M., Rocky S. and Elizabeth A.K., Phylogenetic structure in the grass family (Poaceae) : evidence from the nuclear gene phytochrome B, *American Journal of Botany*, **87(1)**, 96–107 **(2000)**

3. Bevitori R., Oliveira M.B. and Lanna A.C., Selection of Optimized Candidate Reference Genes for qRT-PCR Normalization in Rice (Oryza Sativa L.) during Magnaporthe Oryzae Infection and Drought, *Genetic and Molecular Research*, **13(4)**, 9795–9805 **(2014)**

4. Braun Pascal, Jelle V.L., Geert D.J. and Claire L., Plant Protein Interactomes, 1–27 **(2013)**

5. Bai Zetao et al, The Impact and Origin of Copy Number Variations in the Oryza Species, *BMC Genomics*, **17,** 1–12 **(2016)**

6. Borthakur Anwesha and Pardeep S., Agricultural Research in India: An Exploratory Study, *International Journal of Social Science and Interdisciplinary Research*, **1(9)**, 59-74 **(2012)**

7. Cao Peijian et al, The Rice Oligonucleotide Array Database : An Atlas of Rice Gene Expression, Rice, 1–9 **(2012)**

8. Copetti Dario et al, RiTE Database : A Resource Database for Genus-Wide Rice Genomics and Evolutionary Biology, *BMC Genomics*, doi: 10.1186/s12864-015-1762-3 **(2015)**

9. Donald K., The science of plant morphology : definition, history and role in modern biology, *American Journalof Botany*, **88(10),** 1711–1741 **(2001)**

10. Dethier, Jean J. and Alexandra E., Agriculture and Development : A Brief Review of the Literature, *Economic Systems*, **9(3),** 175-205 **(2012)**

11. Farooq M., Hussain M., Abdul W. and Siddique M., Drought Stress in Plants : An Overview, *Plant Responses to Drought Stress*, **1,** 1-33 **(2012)**

12. Filippis D., Bioinformatic Tools in Crop Improvement, Crop Improvement, Sydney, 49-58 **(2007)**

13. Youens-Clark K., Buckler E. and Casstevens T., Gramene database in 2010: updates and extensions, *Nucleic Acids Res*, **39**, 1085–1094 **(2011)**

14. Yamamoto E., Yonemaru J., Yamamoto T. and Yano M., OGRO: The Overview of functionally characterized Genes in Rice online database, *Rice*, **5**, 26 **(2012)**

15. Yonemaru J. and Yamamoto T., Fukuoka S., Q-TARO: QTL Annotation Rice Online Database, *Rice*, **3**, 194–203 **(2010)**

16. Youens C.K., Buckler E., Casstevens T., Chen C., Declerck G., Derwent P., Dharmawardhana P., Jaiswal P., Kersey P., Karthikeyan A.S., Lu J., McCouch S.R., Ren L., Spooner W., Stein J.C., Thomason J., Wei S. and Ware D., Gramene database in 2010: updates and extensions, *Nucleic Acids Research*, **1**, 1085–94 **(2011)**

17. Green M.L. and Karp P.D., The Outcomes of Pathway Database Computations Depend on Pathway Ontology, *Nucleic Acids Research*, **34(13)**, 3687–97 **(2006)**

18. Galbraith David W. and Jeremy E., Applications of Microarrays for Crop Improvement : Here, There and Everywhere, *Bioscience*, **60**, 337–48 **(2010)**

19. Gao Dongying, Transposons Play an Important Role in the Evolution and Diversification of Centromeres among Closely Related Species, *Frontiers in Plant Science*, **6**, 1-11 **(2015)**

20. Helmy Mohamed, Masaru T. and Yasushi I., OryzaPG-DB : Rice Proteome Database Based on Shotgun Proteogenomics, *BMC Plant Biology*, **11**, 63 **(2011)**

21. Jun-Ichi I and Ken-Ichi N., Rice Plant Development : From Zygote to Spikelet, *Plant and Cell Physiology*, **46(1)**, 23–47 **(2005)**

22. Kim Joung S. et al, A Pepper MSRB2 Gene Confers Drought Tolerance in Rice through the Protection of Chloroplast-Targeted Genes, *Plos One*, **9(3)**, 1-17 **(2014)**

23. Kurata Nori and Yukiko Y., Oryzabase, An Integrated Biological and Genome Information Database for Rice, *Plant Physiology*, **140**, 12–17 **(2006)**

24. Kim Chang K., RiceQTLPro : An Integrated Database for Quantitative Trait Loci Marker Mapping in Rice, *Plant Bioinformation*, **10(10),** 0–2 **(2014)**

25. Kim Chang K., Bioinformation PlantGM : A Database for Genetic Markers in Rice (*Oryza Sativa*) and Chinese Cabbage, *Bioinformation,* **3(6),** 1–2 **(2008)**

26. Kanehisa Minoru and Susumu G., KEGG : Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Research,* **28(1)**, 27–30 **(2000)**

27. Klingstro Tomas, Protein Protein Interaction and Pathway Databases , a Graphical Review, *Briefings in Bioinformatics,* **12(6)**, 702–13 **(2011)**

28. Liang Chengzhi., Gramene : A Growing Plant Comparative Genomics Resource, *Nucleic Acids Research*, **36**, 947–953 **(2008)**

29. Mitra Jiban, Genetics and Genetic Improvement of Drought Resistance in Crop Plants, *Current Science,* **80(6)**, 758-763 **(2001)**

30. Matsuoka et al, Whole Chloroplast Genome Comparison of Rice, Maize and Wheat : Implications for Chloroplast Gene Diversification and Phylogeny of Cereals, *Molecular Biology and Evolution*, **19(12)**, 2084–91 **(2002)**

31. Miriam V. and Danka V., Transposons – the Useful Genetic Tools, *Biologia,* **3**, 309–18 **(2004)**

32. Ning Jing, Siyuan L., Honghong H. and Lizhong X., Systematic Analysis of NPK1 -like Genes in Rice Reveals a Stress-Inducible Gene Cluster Co-Localized with a Quantitative Trait Locus of Drought Resistance, 535–46 **(2008)**

33. Nezhadahmadi Arash, Zakaria H.P. and Golam F., Drought Tolerance in Wheat, *The Scientific World Journal*, **10(11)**, 1-13 **(2013)**

34. Narsai Reena, Rice DB : An Oryza Information Portal Linking Annotation , Subcellular Location , Function , Expression, Regulation and Evolutionary Information for Rice and Arabidopsis, *The Plant Journal*, **10**, 57–73 **(2013)**

35. Ogata Hiroyuki, KEGG : Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Research,* **27(1)**, 29–34 **(1999)**

36. Oystein H., Digest : Toward Predicting Evolutionary Response to Environmental Change : The Power of Integrated Experimental and Genetic Studies, *International Journal of Organic Evolution*, **71(3)**, 800-801 **(2017)**

37. Rolf S. and Rolfrutishaus E., The Fundamental Relevance of Morphology and Morphogenesis to Plant Research, *International Journal of Plant Science,* **1**, 571–82**(1997)**

38. Rafi S and Wusirika R., Machine Learning Approaches Distinguish Multiple Stress Conditions Using Stress-Responsive Genes and Identify Candidate Genes for Broad, *Plant Physiology,* **164(1)**, 481–495 **(2014)**

39. Ramalingam J., Arul L. and Sathishkumar N., TNAURice: Database on rice varieties released from Tamil Nadu Agricultural University, *Bioinformation,* **5**, 264 **(2010)**

40. Rajendran D., Stanislaus A.C., Krishnaraj T. and Savarimuthu I., Genetic Engineering of Crop Plants for Drought Tolerance : Role of Transcription Factors, *South Indian Journal of Biological Sciences*, **2(2)**, 272–86 **(2016)**

41. Raghuvanshi S., Gour P. and Joseph S.V., Manually Curated Database of Rice proteins (MCDRP), a database of digitized experimental data on rice, *Current Plant Biology,* **7(8),** 26-30 **(2016)**

42. Sato Y., Namiki N. and Takehisa H., RiceFREND: a platform for retrieving coexpressed gene networks in rice, *Nucleic Acids Res,* **41**, 1214–1221 **(2013)**

43. Sato Y. et al, RiceXPro Version 3.0: expanding the informatics resource for rice transcriptome, *Nucleic Acids Res,* doi: 10.1093/nar/gks1125, **41**, D1206–D1213 **(2013)**

44. Soderlund C., Haller K. and Pampanwar V., MGOS: A resource for studying Magnaporthe grisea and Oryza sativa interactions, *Mol Plant Microbe Interact,* **19**, 1055–1061 **(2006)**

45. Subramanian Aravind et al, Gene Set Enrichment Analysis : A Knowledge-Based Approach for Interpreting Genome-Wide, PNAS, 15545-15550 **(2005)**

46. Sasaki Takuji and Benjamin B., International Rice Genome Sequencing Project : The Effort to Completely Sequence the Rice Genome, *Rice Science*, **12(1)**, 138–41 **(2000)**

47. Shaik Rafi and Wusirika R., Genes and Co-Expression Modules Common to Drought and Bacterial Stress Responses in Arabidopsis and Rice, *Plos One,* **8(10)**, 1–16 **(2013)**

48. Sheng Z., Wenbin Y., Lishan Y., Hongjuan F., Congting Y., Xuesong X., Yuanhaowei J., Weixu L., Guoli J. and Xiaohui W., PlantAPAdb: A Comprehensive Database for Alternative Polyadenylation Sites in Plants, *Plant Physiology*, **182**, 228-242 **(2020)**

49. Scott F., The NCBI Taxonomy Database, *Nucleic Acids Research*, doi:10.1093/nar/grk1178 **(2012)**

50. Vinod M.S. and Sharma N., Candidate Genes for Drought Tolerance and Improved Productivity in Rice (Oryza Sativa L .), *Journal of Bioscience,* **31(1),** 69–74 **(2006)**

51. Wang D., Xia Y. and Li X., The Rice Genome Knowledgebase (RGKbase): an annotation database for rice comparative genomics and evolutionary biology, *Nucleic Acids Res,* **41**, 1199–1205 **(2012)**

52. Zhao W., BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics, *Nucleic Acids Res,* **32**, 377D–382 **(2004)**.

(Received 20th April 2020, accepted 21st June 2020)