



Review article

Biocomputation assisted data base of plant stress responsive traits: Current progress and future challenges

Dibyendu Talukdar*

Dept. of Botany, R.P.M. College (University of Calcutta), Uttarpara, Hooghly, West Bengal-712258, India

*Corresponding Author: dibyendutalukdar9@gmail.com

[Accepted: 16 August 2015]

Abstract: Plants are sessile organism and therefore, they are continuously exposed to diverse types of biotic and abiotic stress factors. Understanding of plant stress response mechanisms and application of knowledge derived from integrated experimental approaches gradually develop holistic approaches in plant stress tolerance mechanism. The stunning advancements in ‘Omics’-based technologies and information technology steered development of a broadening field of science, termed bioinformatics or computational biology in which biological phenomena are meaningfully explained through information technology. Plants respond to environmental stresses through cascade of events, occurring at cellular, metabolic and molecular level but in fine tune during tolerance mechanisms. During the last decade, the stress responsive factors have been extensively studied in model and crop plants, but the global data is highly sparse in different data bases. Present endeavor, thus, provides a comprehensive review on plant stress responsive database generated through assistance of biocomputation in different plants. Furthermore, names of different stress factors and data base links were also provided. The progress achieved, so far and upcoming challenges of stress responsive data base were also discussed.

Keywords: Computational Biology - Transcription factors - Comparative genomics - Bioinformatics tools - *Arabidopsis thaliana*.

[Cite as: Talukdar D (2015) Biocomputation assisted data base of plant stress responsive traits: Current progress and future challenges. *Tropical Plant Research* 2(2): 143–153]

INTRODUCTION

Plants are sessile organisms, and therefore, they have to withstand arrays of both biotic and abiotic stresses. Stress response is the general term for defining the interaction between plants and the extreme environmental conditions. The term ‘Stress’ is difficult to define. However, it is a type of metabolic derailment which occurred due to cellular and metabolic modulations in plant homeostasis (Kumar *et al.* 2014). Plants have developed integrated and interactive mechanisms to detect precise environmental changes, allowing optimal responses to adverse conditions (Atkinson & Urwin 2012). Stress response and tolerance towards these stresses are governed by complex biological pathways and regulatory events involving multiple molecular components. The effects of abiotic or biotic stress may occur singularly or in combination and induce cellular damage at multiple levels of plant growth and development, and may induce varying degrees of alterations in cellular, metabolic and molecular events, ultimately manifested in reduction of growth and development (Chinnusamy *et al.* 2004). Prominent crop plants like rice, wheat, maize, grain legumes, vegetables, spices and other economically important plants like oil-yielding crops, fiber-yielding crops, and plants with medicinal and aromatic values suffered varying degrees of biotic and abiotic stresses (Mahajan & Tuteja 2005). This complex mechanism is mediated through distinct signal transduction cascades, which in turn activate stress-responsive genes, consequently leading to survival by transcriptional reprogramming during plant growth and development (Xiong & Zhu 2001, Talukdar & Talukdar 2013, 2014). Understanding the molecular pathways and regulatory networks which influence various facets of stress-responsive events in plants is crucial for developing stress-tolerant or stress-adaptive varieties of plants. Furthermore, interactions of several biological signaling molecules particularly, H₂O₂, nitric oxide and hydrogen sulfide with prominent antioxidant molecules like glutathione and ascorbate and enzymatic defense components in modulation of stress response traits need to be studied

holistically for development of stress-tolerant crops (Chakrabarty *et al.* 2009, Talukdar 2011, 2012, 2013a, d, 2015b, Tripathi *et al.* 2012a, b).

The practice of integrating morphological, physiological, molecular, biochemical and genetic information has long been applied to diverse fields of plant biological research (Yuan *et al.* 2008), inducing the origin of system biology which landmarks the interactions among biological components using models and/or networks to integrate genes, metabolites, proteins, regulatory elements and other biological components through data mining, data modeling and visualization, meta-analysis, multivariate analysis, and dynamic as well as static networks (You 2004, Yuan *et al.* 2008, Bhardwaj & Somvanshi 2015). Considering the large datasets continuously generated through different ‘omics’ technologies such as genomics, proteomics, transcriptomics, interactomics and metabolomics, there is an urgent need for system integration in plant biology. Biocomputational approaches offer a powerful platform of bioinformatics to gather information by integrating various public data sets and sensitive prediction algorithms, helping to understand the major cellular and molecular activities involved in stress response, adaptation and tolerance (Fernandez-Suarez & Birney 2008, Perez-Rodriguez *et al.* 2010, Cramer *et al.* 2011, Kinsella *et al.* 2011, Sucaet & Deva 2011, Spooner *et al.* 2012; Talukdar 2015a). While mathematical methods are based on description and analysis of intra and intercellular processes and events through systematic mathematical equations, bio computation methods mainly serve for creation of algorithms to simulate biological processes, and to construct and visualize them. The experimental techniques employed in systems biology tend to have high-throughput capabilities by the employment of techniques, such as protein–protein interactions, transcriptional regulations (protein–DNA interactions) and genetic interactions. This enables the researchers to determine the abundance or activity of numerous biological components at the same time (Bhardwaj & Somvanshi 2015). In *Arabidopsis thaliana*, data base of biochemical pathway has been developed (Caspi *et al.* 2010, Mueller *et al.* 2003), but stress responsive factors are not distinct. The web-accessible database providing information on plant stress responsive traits at metabolic and genomic levels proves to be an invaluable resource for crop improvement, particularly when the climate change phenomena in the backdrop of immense population pressure are looming large. However, in most of the cases the information is sparse and is not easily accessible. This review, thus, aims to integrate and reveal the current progress and upcoming challenges of using biocomputational tools in developing data base of plant stress responsive features.

BIOCOMPUTATION IN PLANT STRESS RESPONSIVENESS: The Arabidopsis and rice model

Genomics and proteomics technologies have revolutionized 21st century plant biology research with the advent of ultra-high-throughput experimental platforms, optimized assay systems and advanced bioinformatics approaches (Mochida & Shinozaki 2010). DNA microarray is a high-throughput technology extensively used to investigate plant model organisms such as *Arabidopsis* and rice varieties to detect expression levels of multiple transcripts quantitatively in parallel. On the other hand, tightly regulated protein interaction networks or interactomics mediate cellular responses to environmental cues and direct the implementation of developmental programs. The use of mathematical and computer modeling methods allows investigation of processes and events that are difficult to study even using highly efficient experimental methods (Bhardwaj & Somvanshi 2015, Yuan *et al.* 2008). The signaling cascade of mitogen-activated protein kinase (MAPK) was analyzed by biocomputer-based approaches while carbohydrates and N-glycosylation pathways were biocomputerized in *Pichia pastoris* (Srivastava *et al.* 2013). There are few databases that have been designed for stress responsive genes in plants. *Arabidopsis thaliana* (thale cress), a member of the Brassicaceae family, has become a widely used model for the study of plant biology because of its small genome size, short generation time, facile genetics, and ease of transformation. Since the completion of the genome sequencing in *Arabidopsis* (Arabidopsis Genome Initiative 2000) and *Oryza sativa* L. genomes (Yu *et al.* 2002, Zhao *et al.* 2004), several stress responsive data bases are being developed in both plant species, some of which are given below-

i) Arabidopsis Stress Responsive Gene Database (ASRGDB data base)

The Arabidopsis Stress Responsive Gene Database or ASRGDB database (<http://srgdb.bicpu.edu.in/>), developed by School of Life Sciences, Pondicherry University, India as freely web-accessible platform, was constructed and configured by using typical LAMP (Linux, Apache, MySQL, and PHP) server with data search, use and retrieval facilities (Borkotoky *et al.* 2013). Database design and interface has been developed using PHP and MySQL v. 5.5 whereas BLAST searches were carried out using PERL scripts. It listed around 44 types

of different stress factors related to *Arabidopsis thaliana*, and contains 636 gene entries related to stress response with their related information like gene ID, nucleotide and protein sequences, cross-response, and so forth. The database is based exclusively on published stress responsive genes associated with the *Arabidopsis*, and include BLAST search interface for both nucleotide and proteins. Genomic and proteomic data for the collected gene have been obtained from the TAIR (The Arabidopsis Information Resources), while gene expression data were obtained from Genevestigator (Zimmermann *et al.* 2004) and *Arabidopsis* Gene Family Profiler (aGFP) (Dupl'akov'a *et al.* 2007). The database is freely web-accessible.

ii) STIFDB and STIFDB2 data base

The mechanism of stress-responsiveness in plants involves complex regulation of multiple genes and transcription factors. About 10 specific families of transcription factors were found involved in *A. thaliana* while six specific families of transcription factors were known to be participated in rice responding to diverse abiotic stresses. Notable families involved are ABI3/VP1, AP2/EREBP, ARF, BHLH/myc, bZIP, HSF, MYB, NAC, and WRKY for salinity, drought, abscisic acid, cold, excess light, and oxidative stress. STIFDB or the Stress-responsive Transcription factor database is a database that curated information of stress-responsive genes and transcription factor binding sites for abiotic stress-responsive genes in *A. thaliana*. STIFDB was developed on a MySQL backend and the web interface was developed using HTML and JavaScript. Perl-CGI programs for searching putative binding sites and for performing STIF prediction were used for the development of search, query and retrieval system. Stress profiles have been created for each gene that indicates the associated stress signals. Gene descriptions from TAIR (Lamesch *et al.* 2012), Rice Annotation Project (RAP) database (Ouynag *et al.* 2007), rice DRTF (transcription factors) (Gao *et al.* 2006) and transcription factor-related information from Database of Arabidopsis Transcription Factors (DATF) (Guo *et al.* 2005) were also added. The STIFDB provides features like, 1. TFmap-a graphical representation of the upstream regions of the stress genes in *Arabidopsis thaliana* with the predicted and the validated transcription factor binding sites marked along with their Z-Scores; 2. TAIR ID; 3. Gene Ontology- Obtained from TAIR, gene ontology annotations are to help the users to understand the known functional associations of genes in STIFDB; 4. Gene description- a short description of genes along with predicted domain associations from InterPro database; 5. Gene names- using standard gene names or its aliases reported in TAIR, users can access STIFDB; 6. Chromosome position-Users can detect exact location of the relevant stress gene among the five *A. thaliana* chromosomes; 7. Reference-publications and related resources; 8. Transcription Factor Family Name- Family whose binding site sequence has been located/predicted on a given promoter sequence; 9. Binding Site Information-it refers to the core binding sequence to which a transcription factor binds; 10. Orientation of Binding Sites- it refers to the DNA strand on which the transcription factor-binding site has been located which is either on the forward strand or on the reverse DNA strand; 11. Stress signals- type of stress, which regulates the transcription factor/s; 12. Algorithms scores, etc (Shameer *et al.* 2009). Several studies involving plant stress-responsiveness (Kang *et al.* 2011, Sanghera *et al.* 2011, Babitha *et al.* 2013) and biocomputational approaches (Mishra *et al.* 2009, Georgii *et al.* 2012) on stress-specific gene regulation have utilized the data compiled in STIFDB (Shameer *et al.* 2009).

The STIFDB2 is the up-dated large compendium of STIFDB, providing robust data platform of plant stress regulome. New features such as: additional stress signals, new transcription factors and their binding sites, additional stress-responsive genes from microarray experiments and orthologs recorded from other important crop plants, such as maize, sorghum and soybean have been added in this data base (Naika *et al.* 2013). STIFDB2 provides information on stress-responsive genes from *A. thaliana* and two rice subspecies (*O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica*). A total of 31 stress-responsive transcription factors such as ABRE_ABI3_VP, AuxRE_ARF, C_ABRE_bZIP, GCC_box_AP2_EREBP, G_box2_bZIP, HSE1_HSF, Myb_box1_MYB to Myb_box5_MYB, Nac_box_NAC, W_box_WRKY etc. identified for different stress signals like cold, cold–drought–salt, drought, heat, ABA, aluminum, bacterial blight, high light, iron, NaCl, osmotic stress, oxidative stress, UV-B and wounding was also compiled (Shameer *et al.* 2009, Naika *et al.* 2013). Six families of transcription factors were curated for subsp. *japonica* whereas three families such as AP2/EREBP, NAC, and bZIP for salinity stress, ABA, cold and drought stress were catalogued in subsp. *indica*. Transcription factor binding site was predicted by transcription factor family/subfamily, cis-element, consensus binding site data and corresponding references (PubMed identifiers) using stress-responsive genes in rice species. Using the genomic data mining approach, a data set of 3,150 unique stress responsive genes in *A.*

thaliana, 1,118 genes in *O. sativa* subsp. *japonica* and 1,716 in *O. sativa* subsp. *indica* were identified and annotated with stress signals. Chromosome location of genes in three taxa was also provided, and can be browsed on the basis of chromosomal location. While ABA, bacterial blight, cold, drought, heat, iron and NaCl stress signals were captured for subsp. *japonica*, cold, drought and NaCl-stress was signaled in subsp. *indica* (Naika *et al.* 2013). STIFDB2 has 38,798 associations of stress signal, stress-responsive gene, transcription factor binding site, orientation of binding site and z-scores predicted using the STIF (Stress-responsive Transcription Factor) algorithm ('Hidden –Markov' model). Thus, there are three categories of data compiled in STIFDB2: A. predicted data, B. curated data which can be extracted, and C. annotation data compiled from primary data base. Data can be browsed using one of the four key data sets: gene, transcription factors, stress signals and chromosome, or using a variety of keywords including gene descriptions, stress signals, transcription factors, Gene Ontology annotations, and nucleotide sequences (BLAST). STIFDB2 can also be used to investigate the transcriptional regulatory cascade network underlying abiotic stress responses in *A. thaliana* (Naika *et al.* 2013).

iii) *The Arabidopsis thaliana salt responsive protein data base*

Uniprot (Release 2011_11, <http://www.uniprot.org>) was used to select 292 *A. thaliana* salt-response proteins. The UniProtKB GO annotation program, an integral part of UniProt biocuration, is powered by high-quality Gene Ontology (GO) annotations to proteins in the UniProt Knowledgebase (UniProtKB). Its subsection of GOTERM_BP_FAT provides information on the proteins response to various stresses. All expressed proteins were extracted from Uniprot database to analyze GOTERM_BP_FAT. Subsequently, these proteins with key word “response to salt stress” in the description of GOTERM_BP_FAT were referred as salt-response proteins and their molecular function was classified following GOTERM_MF_FAT (Meili *et al.* 2012). The data base harbors the largest number of proteins related to catalytic activity (18.5%), followed by proteins related to signal transduction (17.5%), related to binding activity (12.0%), and related to ROS scavenging and defense (11.6%) in relation to salt stress. Furthermore, out of 292 salt-response proteins, 146 were reportedly involved in the response of cold, drought, and heavy metal stress in *A. thaliana*. Among them, 67 proteins for cold stress; 71 proteins for drought stress; and 54 proteins responded to heavy metal stress (Meili *et al.* 2012). Largest number of proteins was involved in signal transduction, followed by catalytic activity (16.4%), and then ROS scavenging and defense. Besides Uniprot, the cellular localizations and functions of different salt-response proteins were further analyzed by Cytoscape v2.8.2, which is an open-source software package, widely used to integrate and visualize diverse data sets in biology (Shannon *et al.* 2003). Available information on known compound-target interactions in the context of a biological network of interest can be used by users to rapidly identify novel avenues to perturb the system and potentially identify therapeutically relevant targets (Meili *et al.* 2012).

iv) *QlicRice data base*

Developed during the year 2011, QlicRice is an online data base and search engine for abiotic stress responsive QTLs and annotated gene loci (MSU/TIGR) in rice. It is an interactive platform for collecting, managing and searching particular rice QTLs along with user-friendly web-enabled data mining of genomic information in rice, particularly in *O. sativa* subsp. *japonica* cv. Nipponbare. Rice QTLs can be searched for abiotic stress, QTL identifier or locus identifier. About 974 abiotic stress related QTLs with 460 overlapping TIGR/MSU loci were collected in the data base. Data mining for QTLs were performed by using Gene Ontology and KEGG Orthology and the positions of QTLs have been identified relative to their overlapping TIGR loci in comprehensively constructed physical and genetic map (Smita *et al.* 2011). A relevant glossary has also been provided.

v) *Rice Stress Gene Catalog tool*

This is a global classification of genes from rice plant genome. This software was developed under the adobe flex 3 platform while the central database was established by SQLight, and has been used only in the portal. The MXML (www.adobe.com/) and Action Script 3.0 (www.actionscript.org/) was the major programming languages. While MXML is used for mainly front end interface designing, Action Script is used for data base establishment, online resource communication and in all programming logics. The system was tested on computation node(s) with Linux, Windows XP and MAC platform (Ali *et al.* 2011). Facilities are available for analysis of genomic and proteomic data, to find the homology between the sequences, profile

searching, sequence comparison etc. Well-known packages like Clustal W and T-Coffee for multiple sequence alignment, and FASTA and BLAST for sequence comparison have also been provided. There are four main options for using the tool like browsing stress gene entries, search against tool by gene names or protein names, and insert data and manipulations on FASTA sequence. Details about each stress responsive protein family, for example, the list of Uniprot IDs, protein names, Gene Ontology terms and key publications for each protein can be accessed through searching particular protein name against the Uniprot database (Mi *et al.* 2005). Search can be facilitated by key words while users can modify it by inserting new data on stress gene/protein families. Separate windows have been created for sequence alignment and comparison packages which can be run independently (Ali *et al.* 2011).

BIOCOMPUTATION IN PLANT STRESS RESPONSIVENESS: The miRNAs

MicroRNAs (miRNAs) are endogenous, single-stranded, small (~21–23nt), non-coding, regulatory RNA molecules that manipulate messenger RNA degradation and translation repression by binding complementary sites in the protein-encoding, or 3'-untranslated regions, of target mRNAs (Bartel 2004, Zhang *et al.* 2009). Accumulating evidences suggest that plant miRNAs play critical roles in plant physiology, such as floral organ identity, leaf morphogenesis, cellular signaling, and stress responses (Lu *et al.* 2008, Chiou & SI 2011). Over 200 published studies representing 1038 regulatory relationships between 682 miRNAs and 35 different abiotic stresses in 33 plant species have been reported. However, data from these individual reports has not been collected into a single database. The lack of a curated database of stress-related miRNAs limits research in this field, and thus a cohesive database system should necessarily be constructed for data deposit and further application.

The PASmiR data base

PASmiR is a literature-curated and web-accessible database, developed to provide detail searchable descriptions of miRNA molecular regulation in different plant abiotic stresses (Zhang *et al.* 2013). The database was constructed using freely available and open source frameworks, such as Apache, Java Server Pages (JSP), MySQL, Struts2, Spring, Hibernate and so on. The MySQL Server stored the curated miRNA-stress regulatory relationships which are accessible through Structured Query Language (SQL) and web-hosted on Apache Tomcat server and Java language and JSP scripts were combined to generate web page. Additionally, Hibernate modules were applied to manipulate data and convert data formats (Zhang *et al.* 2013). The data base allows users to browse, search, download, and update the data relevant to miRNA regulation in diverse abiotic stresses such as heavy metal, salinity, drought, radiation, etc. through key words and/or through search fields like plant species, abiotic stress, and miRNA. *PASmiR* also offers a fuzzy search engine, allowing querying of miRNA-stress regulatory information, and feedback can be submitted. For each of regulatory relationship, detail information regarding plant species name, type of abiotic stress, a miRNA identifier, change in miRNA expression pattern under stress, detection method for miRNA expression (*e.g.* miRNA microarray, qRT-PCR, northern blot, and deep sequencing, etc.), a reference literature, the target gene(s) of miRNA extracted from the corresponding reference, and validated (predicted) target gene(s) derived from miRBase were collected (Zhang *et al.* 2013). Data provided by *PASmiR* will facilitate improvements in the understanding of stress-focused miRNA functional evolutionarily relationship in plant species. It has great potential in applications for experimental biology aiming to improve crop performance and breeding characteristics. High connectivity for different stresses in rice can be studied by studying 162 regulatory relationships between 113 miRNAs and 13 abiotic stresses in the crop. The *PASmiR* database is freely accessible at: <http://hi.ustc.edu.cn:8080/PASmiR>, and <http://pcsb.ahau.edu.cn:8080/PASmiR>

BIOCOMPUTATION IN PLANT STRESS RESPONSIVENESS: Other data bases

i) Plant Stress-Responsive Gene Catalog data base (PSRGC data base)

The Plant Stress-Responsive Gene Catalog or PSRGC is designed and developed by Consultative Group of International Agricultural Research (CGIAR) in the year 2011. This is a data base of orthologous and paralogous relationships between different stress responsive genes for water and drought stress in different crop plants. This data base is housed by NAR (Nucleic Acid Research) Molecular Biology (MB) Database Collection with an entry number 1116. It has two main categories: 1. Nucleotide sequence data base with a sub-category of

‘Transcriptional regulator sites and transcription factors’, and 2. Plant data base with sub-category of ‘General Plant Data Base’.

ii) Plant Stress Gene Database (PSGD)

Designed and developed through LAMP technology (Linux, Apache, MySQL, PHP) this data base (<http://ccbb.jnu.ac.in/stressgenes/>) include 259 stress-related genes of 11 plant species along with all the available information about the individual genes in model plant *Arabidopsis thaliana*, cereals like *Oryza sativa*, *Hordeum vulgare*, *Triticum aestivum*, *Zea mays*, and *Saccharum officinarum*, legumes like *Phaseolus vulgaris*, *Glycine max*, *Arachis hypogaea*, forage cereals like *Pennisetum glaucum* and vegetable crop *Solanum lycopersicum*. Stress related ESTs were also found for *Phaseolus vulgaris*. Database also includes ortholog and paralog of proteins which are coded by stress related genes (Prabha *et al.* 2011). About 33 genes such as ABF3/DPBF5 (ABA responsive elements binding factor 3/DNA binding/protein binding), ABI4 (ABA Insensitive 4), RAB18 (Responsive to ABA 18), CSD1, CSD2 & CSD3 (Cu/Zn Superoxide Dismutase 1, Cu/Zn Superoxide Dismutase 2 & 3), FsD1, FSD2 and FSD3 (Fe-Superoxide Dismutase 1, 2 & 3), MSD1 (Mn Superoxide Dismutase 1), DREB1A & DREB2A (Dehydration Response Element B1A & B2A), HOS10 (High Response to Osmotic Stress 10), HSA32 (Heat Stress Associated 32), LOS4 (Low Expression of osmotically responsive genes), SEP 1 & 2 (stress Enhanced Protein 1, 2), SOS1 (Salt Overly Sensitive 1) and STRS1 & 2 (Stress Responsive Suppressor 1 and 2) have been catalogued in *A. thaliana* for stress like dehydration, salinity, Abscisic acid, osmotic stress, and heat stress. Among cereals and millets, about 30 genes each in rice and wheat, 39 genes in *Zea mays*, 24 in *Hordeum*, and 03 each in *P. glaucum*, and *Saccharum officinarum* have been catalogued in PSGD data base. Prominent genes involved in stress responsiveness documented here are for drought/heat/temperature stress (putative receptor kinase, putative 1,4-benzoquinone reductase, auxin-responsive GH3, Calmodulin-regulated ion channel, heat shock factor type, Superoxide dismutases, temperature stress-induced lipocalin, WRKY transcription factor 20, 28, 36, & 41, Universal Stress Proteins), Oxidative injury (Ascorbate peroxidase, Chitinase 1, 2 & 3), osmotic stress (glutathione S-transferase, NADPH oxidase, LOX11-Lipoxygenase 11, betaine aldehyde dehydrogenase), cold (cold acclimation protein, dehydrin WZY1-2, WCS120, WRKY transcription factor 20, 28, 29, 36), salt stress (Zinc-finger motif, salt-stress responding, adenosine kinase, aldolase1 (preferred), aldolase1, ZmL54, gpm362, umc216, umc216 (ald1), ATP synthase CF1 alpha subunit, beta glucosidase1, GPN1, GAPN, methionine synthase, rbcL, ribosomal protein S4, voltage dependent anion channel protein), aluminium stress (wali2, 3 & 5 protein, PAL), endogenous hormone (LOC-548234-alpha-hordothionin precursor, LOXA, LOX1.1), and in all stresses (sulfate transporter ST1). Among legumes, two stress-related genes coding acetyl-CoA carboxylase carboxyltransferase beta subunit responsive to wounding, *Pseudomonas syringae* infection, H₂O₂, jasmonic acid (JA), ethylene, or auxin treatment, and rps4 (ribosomal protein S4) along with 42 ESTs for drought stress responsiveness have been catalogued for *Phaseolus vulgaris* (<http://ccbb.jnu.ac.in/stressgenes/Phaseolus.html>) while 62 proteins responding to different abiotic stresses have been curated for soybean (<http://ccbb.jnu.ac.in/stressgenes/glycinemax.html>). Two genes coding for DREB2A (DRE-BINDING PROTEIN 2A), DNA binding / transcription activator/ transcription factor and ACC deaminase/D-cysteine desulhydrase family protein have been catalogued for water stress and salinity stress, respectively, in ground nut (http://ccbb.jnu.ac.in/stressgenes/Arachis_hypogaea.html). Among the vegetables, 31 genes coding for proteins having roles in different stress conditions like salt stress and potassium and iron deficiency signalling pathways, water-deficit, heavy metal, chilling tolerance, and other abiotic stress conditions have been documented in tomato (<http://ccbb.jnu.ac.in/stressgenes/solanum.html>). Notables proteins catalogued are abscisic stress ripening protein 1 (water-deficit), APX, TAPX (Ascorbate peroxidase in all stresses), calcium-dependent protein kinase CDPK1, calmodulin-independent protein kinase (salt and wounding), cat1, CAT2 (catalase-chilling tolerance, H₂O₂-mediated antioxidant defense), GPX, phytoene synthetase, and rbcL (Rubisco Large sub-unit) for salt tolerance and other environmental stresses.

iii) Plant Stress Protein Database (PSPDB)

Plant Stress Protein Database (PSPDB) is a web-accessible (<http://www.bioclues.org/pspdb/>) resource which covers 2,064 manually curated plant stress proteins catalogued from 134 plant species with 30 different types of biotic and abiotic stresses. Functional and experimental validation of proteins responsive in biotic and abiotic stresses has been employed as the sole criterion for inclusion in the database (Kumar *et al.* 2014). Proteins were retrieved from UniProt through GO term search for various plant stress-related proteins. The data base is driven

through MySQL v. 5.0, PHP v. 5.2.4, and Perl v. 5.8.8. Seven abiotic and four biotic stress factors have been curated in this data base which has facilities like multiple catalogue search, integrated tools like NCBI, BLAST, ClustalW, NJPLOT, etc, tutorials for search and updated version of proteins. Among abiotic stresses, drought, salt, flooding, temperature, light, wounding, and oxidative stress are being curated while phytohormone, antifungal, antibacterial, and antiviral protein factors have been catalogued for biotic stresses. Notable plant taxa covered in the PSPDB data base are *Arabidopsis thaliana*, *Avicennia marina* (Grey mangrove), *Mesembryanthemum crystallinum* (common ice plant), cereals/millet like rice, wheat, maize, rye, sorghum, *Avena sativa* (oat), and *Hordeum vulgare*, vegetables such as tomato, potato, carrot, *Solanum cheesmanii* (Galapagos island tomato), *Spinacea oleracea*, *Allium cepa*, *A. sativum*, *Brassica napus*, *Benincasa hispida*, *Sinapis alba* (white mustard), *Basella alba*, *Beta vulgaris*, *Cucurbita maxima*, *Cucurbita pepo*, *Capsicum annuum*, *Brassica oleracea*, *Ipomoea batatas* (sweet potato), *Momordica charantia*, *Trichosanthes anguina*, and *Cucumis sativus*, legumes like *Pisum sativum*, *Vigna radiata* var. *radiata*, *Vigna unguiculata* var. *sesquipedalis*, *Vicia faba*, *Lens culinaris* subsp. *culinaris*, ground nut, chickpea, *Medicago sativa* (alfalfa), soybean, *Phaseolus vulgaris*, and *Lotus japonicas*, economically/commercially important plant such as *Sesamum indicum*, *Helianthus annuus*, *Hevea brasiliensis*, *Gossypium hirsutum*, *Vitis vinifera*, *Nigella sativa*, *Citrus sinensis* (sweet orange), *Ginkgo biloba*, *Cycas revoluta*, *Pinus sylvestris*, *Dianthus caryophyllus*, *Passiflora* sp, *Malva parviflora*, *Dahlia merckii*, *Zinnia violacea*, *Catharanthus roseus*, *Petunia hybrida*, *Artemisia annua* (sweet wormwood), *Bryonia dioica*, *Asparagus officinalis*, *Calotropis procera*, and *Nicotiana tabacum*.

iv) PESTD data base

With 175976 entries, this data base was developed by International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India in 2005. This data base is curating transcripts with annotated tentative orthologs from crop abiotic stress transcripts. ESTs from stress cDNA libraries across 16 crop species including 6 important cereal crops (rice, wheat, maize, barley, sorghum, rye), 6 legumes (common bean, soybean, cowpea, groundnut, chickpea, *Medicago*) and other dicots including *Arabidopsis* were systematically collated and subjected to bioinformatics analysis such as clustering, grouping of tentative orthologous sets, identification of protein motifs/patterns in the predicted protein sequence, and annotation with stress conditions, tissue/library source and putative function (Balaji *et al.* 2006). With permission, this data base is available at <http://intranet.icrisat.org/gt1/tog/homepage.htm>

v) GCP- comparative plant stress-responsive gene catalogue

Comparative biology highlights/predicts the evolution of gene families and biological processes, thus providing valuable insights into organismal function and evolution (Koonin 2005, Brown & Sjolander 2006). Accurate predictions of orthologous and paralogous relationships are essential in order to cross-reference genes from one species to other related species. The Generation Challenge Programme (GCP: www.generationcp.org) is a global crop research consortium of CGIAR. In order to enhance efforts in plant breeding for plant stress tolerance, the consortium is relentlessly striving on comparative genomics and molecular analysis to plant genetic resources. Primary goal of the project is to assemble tools for the compilation and visualization of comparative information about stress-responsive genes (Wanchana *et al.* 2008). After Margaret Dayhoff, the famous early pioneer in comparative analysis of sequences, the project was code-named as Dayhoff and hosted on <http://dayhoff.generationcp.org:8080/DayhoffWeb2/>. Dayhoff is a MySQL database, designed to guide the bioinformatics analysis and interpretation of research outputs obtained from comparative genomics experiments with add-inns options to store protein family information such as protein multiple sequence alignments (MSA), phylogenetic trees and supported stress evidence from experiments and the literature (Wanchana *et al.* 2008). The web interface uses GCP- Java-based software technology (<http://pantheon.generationcp.org>) connected to external tools like BLAST software and other GCP-funded comparative gene analysis resource called GreenPhyl (<http://greenphyl.cirad.fr/cgi-bin/greenphyl.cgi>) for analysing and viewing the query's results related to comparative genomics on plant stress responsive genes (Wanchana *et al.* 2008). The core data set in Dayhoff consists of stress-related protein families characterized by a phylogenomic inference approach (Sjolander 2004). Phylogenetic tree in the data base was constructed by 1) retrieving the homologous sequences for each stress protein from Uniprot data base by using the FlowerPower tool (Glanville *et al.* 2007), 2) MSAs of homologous proteins were constructed with the high-accuracy MSA program, MUSCLE v. 3.52 (Edgar 2004), and lastly, (3)

functional subfamilies were identified for each group using the SCI-PHY web server (Glanville *et al.* 2007). The data base has three main options for search: browsing protein families, query database by gene names or protein names and BLAST search against protein families. Users can browse the entire set of stress-related protein families by selecting the database from the main drop-down menu. The front page comprised of a list of protein families, links for phylogenetic trees and MSA. Family ID links have been created for details about each protein family, for example, the list of Uniprot IDs, protein names, Gene Ontology (GO) terms and key publications for each Protein. Additional information is available from the drop-down list. For query, users can retrieve the pertinent information by searching the database by keywords within Family name and Protein name. For BLAST protein families, users can submit a protein or DNA sequence in Fasta or raw format in order to BLAST in the Dayhoff database, interconnected with GreenPhyl database through GCP-compliant BioMOBY (Wilkinson *et al.* 2005) client web service.

FUTURE CHALLENGES

Following recent advances in technology and the development of ultra-high-throughput technology, the field of plant science is beginning to suffer from data overload. The biggest problems during constructing of genome-scale models of biological systems are acquisition of over-crowded and noisy data which is an inherent property of the high-throughput techniques. False-positive interactions in data sets due to different experimental or *in silico* approaches are one of the main reasons behind this noise. The huge amount of information published daily for any of the model or crop plant makes it almost impossible to store, classify and integrate the massive data in biologically interpretative way. Although bioinformatics tools are extensively used to integrate these data into meaningful data base/data sets, in many cases problems like false homology inference occurs. Besides, there are some technological and computational limitations in calculation of the dynamics and structure visualization of stress responsive pathway and gene to metabolic network study. Furthermore, most of these data are not publicly available, and published paper often only provides limited data which may or may not be freely accessible. Despite rapid advancements of NGS platforms in gene-sequencing methodologies and advantages of metabolomics over transcriptomics and proteomics, the criticality of metabolome study has hindered the progress of this newly introduced branch of ‘omics’ technology in plant stress biology and data base development. Therefore, integration of the stress-responsive gene/protein traits in different model and crop plants will definitely enhance access to gene and protein data in a variety of bioinformatics analysis contexts.

REFERENCES

- Ali M, Khalid RR, Nawaz M & Qamar N (2011) Development of a tool for the analysis of plant stress proteins. *International Journal of Bioautomation* 15: 261–266.
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Atkinson NJ & Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany* 63: 3523–3543.
- Babitha KC, Ramu SV, Pruthvi V, Mahesh P, Nataraja KN & Udayakumar M (2013) Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in Arabidopsis. *Transgenic Research* 22: 327–341.
- Balaji J, Crouch JH, Petite PVNS & Hoisington DA (2006) A database of annotated tentative orthologs from crop abiotic stress transcripts. *Bioinformatics* 1(6): 225–227.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
- Bhardwaj T & Somvanshi P (2015) Plant Systems Biology: Insights and Advancements. In: Barh D *et al.* (eds) *PlantOmics: The Omics of Plant Science*. Springer, India, pp. 791–819.
- Borkotoky S, Saravanan V, Jaiswal A, Das B, Selvaraj S, Murali A & Lakshmi PTV (2013) The Arabidopsis Stress Responsive Gene Database. *International Journal of Plant Genomics* 2013: Article ID 949564. [DOI: 10.1155/2013/949564]
- Brown D & Sjolander K (2006) Functional classification using phylogenomic inference. *PLoS Computational Biology* 2: e77.

- Chakrabarty D, Trivedi PK, Misra P, Tiwari M, Shri M, Shukla D, Kumar S, Rai A, Pandey A, Nigam D, Tripathi RD & Tuli R (2009) Comparative transcriptome analysis of arsenate and arsenite stresses in rice seedlings. *Chemosphere* 74: 688–702.
- Chinnusamy V, Schumaker K & Zhu JK (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *Journal of Experimental Botany* 55: 225–236.
- Chiou TJ & Lin SI (2011) Signaling network in sensing phosphate availability in plants. *Annual Review of Plant Biology* 62: 185–206.
- Cramer GR, Urano K, Delrot S, Pezzotti M & Shinozaki K (2011) Effects of abiotic stress on plants: a systems biology perspective. *BMC Plant Biology* 11: 163.
- Dupl'akov' a N, Re'n' ak D, Hovanec P, Honysov' a B, Twell D & Honys D (2007) *Arabidopsis* gene family profiler (aGFP)-user-oriented transcriptomic database with easy-to-use graphic interface. *BMC Plant Biology* 7: 39.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.
- Fernandez-Suarez XM & Birney E (2008) Advanced genomic data mining. *PLoS Computational Biology* 4: e1000121.
- Gao G, Zhong Y, Guo A, Zhu Q, Tang W, Zheng W, Gu X, Wei L & Luo J (2006) DRTF: a database of rice transcription factors. *Bioinformatics* 22: 1286–1287.
- Georgii E, Salojarvi J, Brosche M, Kangasjarvi J & Kaski S (2012) Targeted retrieval of gene expression measurements using regulatory models. *Bioinformatics* 28: 2349–2356.
- Glanville JG, Kirshner D, Krishnamurthy N & Sjolander K (2007) Berkeley Phylogenomics group web servers: resources for structural phylogenomic analysis. *Nucleic Acids Research* 35: W27–W32.
- Guo A, He K, Liu D, Bai S, Gu X, Wei L & Luo J (2005) DATF: a database of *Arabidopsis* transcription factors. *Bioinformatics* 21: 2568–2569.
- Kang HG, Kim J, Kim B, Jeong H, Choi SH, Kim EK, Lee HY & Lim PO (2011) Overexpression of FTL1/DDF1, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in *Arabidopsis thaliana*. *Plant Science* 180: 634–641.
- Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, Kersey P & Flicek P (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* (Oxford) 2011: bar030.
- Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* 39: 309–338.
- Kumar SA, Hima Kumari P, Sundararajan VS, Suravajhala P, Kanagasabai R & Kavi Kishor PB (2014). PSPDB: Plant Stress Protein Database. *Plant Molecular Biology Reporter* 32: 940–942.
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A & Huala E (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202–D1210.
- Lu S, Sun YH & Chiang VL (2008) Stress-responsive microRNAs in *Populus*. *Plant Journal* 55: 131–151.
- Mahajan S & Tuteja N (2005) Cold, salinity and drought stresses: an overview. *Archives of Biochemistry and Biophysics* 444: 139–158.
- Meili G, Weixi G, Xuejuan Y, Chunxi Z, Fujun L & Xin L (2012) Data mining of *Arabidopsis thaliana* salt-response proteins based on bioinformatics analysis. *Plant Omics Journal* 5: 75–78.
- Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, Kitano H & Thomas PD (2005) The PANTHER Database of Protein Families, Subfamilies, Functions and Pathways. *Nucleic Acids Research* 33: D284–D288.
- Mishra AK, Agarwal S, Jain CK & Rani V (2009) High GC content: critical parameter for predicting stress regulated miRNAs in *Arabidopsis thaliana*. *Bioinformation* 4: 151–154.
- Mochida K & Shinozaki K (2010) Genomics and bioinformatics resources for crop improvement. *Plant & Cell Physiology* 51: 497–523.
- Mueller LA, Zhang P & Rhee SY (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiology* 132: 453–460.

- Naika M, Shameer K, Mathew OK, Gowda R & Sowdhamini R (2013) STIFDB2: An updated version of plant stress-responsive transcription factor data base with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. *Plant & Cell Physiology* 54: e8.
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J & Buell CR (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research* 35: D883–D887.
- Perez-Rodriguez P, Riano-Pachon DM, Correa LG, Rensing SA, Kersten B & Mueller-Roeber B (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Research* 38: D822–D827.
- Prabha R, Ghosh I & Singh DP (2011) Plant stress gene database. *ARPJ Journal of Science and Technology* 1: 28–31.
- Priya P & Jain M (2013) RiceSRTFDB: A database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. *Database* 2013: Article ID bat027. [DOI:10.1093/database/bat027]
- Sanghera GS, Wani SH, Hussain W & Singh NB (2011) Engineering cold stress tolerance in crop plants. *Current Genomics* 12: 30–43.
- Shameer K, Ambika S, Varghese SM, Karaba N, Udayakumar M & Sowdhamini R (2009) STIFDB—*Arabidopsis* stress responsive transcription factor database. *International Journal of Plant Genomics* 2009: Article ID 583429.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13: 2498–2504.
- Sjolander K (2004) Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics* 20: 170–179.
- Smita S, Lenka SK, Katiyar A, Jaiswal P, Preece J & Bansal KC (2011) QlicRice: a web interface for abiotic stress responsive QTL and loci interaction channels in rice. *Database (Oxford)* 2011: bar037.
- Spooner W, Youens-Clark K, Staines D & Ware D (2012) GrameneMart: the BioMart data portal for the Gramene project. *Database (Oxford)* 2012: bar056.
- Srivastava A, Somvanshi P & Mishra BN (2013) Reconstruction and visualization of carbohydrate, N-glycosylation pathways in *Pichia pastoris* CBS7435 using computational and system biology approaches. *Systems and Synthetic Biology* 7: 7–22.
- Sucaet Y & Deva T (2011) Evolution and applications of plant pathway resources and databases. *Briefings in Bioinformatics* 12: 530–544.
- Talukdar D & Talukdar T (2013) Superoxide-dismutase deficient mutants in common beans (*Phaseolus vulgaris* L.): genetic control, differential expressions of isozymes, and sensitivity to arsenic. *BioMed Research International* 2013: Article ID 782450. [DOI: 10.1155/2013/782450]
- Talukdar D & Talukdar T (2014) Coordinated response of sulfate transport, cysteine biosynthesis and glutathione-mediated antioxidant defense in lentil (*Lens culinaris* Medik.) genotypes exposed to arsenic. *Protoplasma* 251: 839–855.
- Talukdar D (2011) Morpho-physiological responses of grass pea (*Lathyrus sativus* L.) genotypes to salt stress at germination and seedling stages. *Legume Research* 34: 232–241.
- Talukdar D (2012) Exogenous calcium alleviates the impact of cadmium-induced oxidative stress in *Lens culinaris* Medic. Seedlings through modulation of antioxidant enzyme activities. *Journal of Crop Science and Biotechnology* 15: 325–334.
- Talukdar D (2013a) Arsenic-induced oxidative stress in the common bean legume, *Phaseolus vulgaris* L. seedlings and its amelioration by exogenous nitric oxide. *Physiology and Molecular Biology of Plants* 19: 69–79.
- Talukdar D (2013b) Studies on antioxidant enzymes in *Canna indica* plant under copper stress. *Journal of Environmental Biology* 34: 93–98.
- Talukdar D (2013c) Allelopathic effects of *Lantana camara* L. on *Lathyrus sativus* L.: oxidative imbalance and cytogenetic consequences. *Allelopathy Journal* 31: 71–90.

- Talukdar D (2013d) Balanced hydrogen peroxide metabolism is central in controlling NaCl-induced oxidative stress in medicinal legume, fenugreek (*Trigonella foenum-graecum* L.). *Biochemistry & Molecular Biology* 1: 34–43.
- Talukdar D (2015a) Digitization of regional plant flora: Step towards global biodiversity information service. *Journal of Biotechnology, Bioinformatics and Bioengineering* 2: 7–12.
- Talukdar D (2015b) Functional interplay between glutathione and hydrogen sulfide in regulation of thiol cascade during arsenate tolerance of common beans (*Phaseolus vulgaris* L.) genotypes. *3 Biotech* ‘in press’. [DOI: 10.1007/s13205-015-0285-6]
- Tripathi P, Mishra A, Dwivedi S, Chakrabarty D, Trivedi PK, Singh RP & Tripathi RD (2012a) Differential response of oxidative stress and thiol metabolism in contrasting rice genotypes for arsenic tolerance. *Ecotoxicology and Environmental Safety* 79: 189–198.
- Tripathi RD, Tripathi P, Dwivedi S, Dubey S, Chatterjee S, Chakrabarty D & Trivedi PK (2012b) Arsenomics: omics of arsenic metabolism in plants. *Frontiers in Physiology* 3: 275.
- Wanchana S, Thongjuea S, Ulat VJ, Anacleto M, Mauleon R, Conte M, Rouard M, Manuel Ruiz M, Krishnamurthy N, Sjolander K, van Hintum T & Bruskiwich RM (2008) The Generation Challenge Programme comparative plant stress-responsive gene catalogue. *Nucleic Acids Research* 36: D943–D946.
- Wilkinson M, Schoof H, Ernst R & Haase D (2005) BioMOBY successfully integrates distributed heterogeneous bioinformatics web services. The planet exemplar case. *Plant Physiology* 138: 5–17.
- Xiong L & Zhu JK (2001) Abiotic stress signal transduction in plants: molecular and genetic perspectives. *Physiologia Plantarum* 112: 152–166.
- You L (2004) Toward computational systems biology. *Cell Biochemistry and Biophysics* 40: 167–184.
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L & Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79–92.
- Yuan JS, Galbraith DW, Dai SY, Griffin P & Stewart Jr CN (2008) Plant systems biology comes of age. *Trends in Plant Science* 13: 165–171.
- Zhang B, Stellwag EJ & Pan X (2009) Large-scale genome analysis reveals unique features of microRNAs. *Gene* 443: 100–109.
- Zhang S, Yue Y, Sheng L, Wu Y, Fan G, Li A, Hu X, Shang Guan M & Wei C (2013) PASmiR: a literature-curated database for miRNA molecular regulation in plant response to abiotic stress. *BMC Plant Biology* 13: 33.
- Zhao W, Wang J, He X, Huang X, Jiao Y, Dai M, Wei S, Fu J, Chen Y, Ren X, Zhang Y, Ni P, Zhang J, Li S, Wang J, Wong GK, Zhao H, Yu J, Yang H & Wang J (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Research* 32: D377–D382.
- Zimmermann P, Hirsch-Hoffmann M, Hennig L & Gruissem W (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiology* 136: 2621–2632.