Proc Indian natn Sci Acad **78** No. 4 December 2012 pp. 671-691

Review Article

The Role of Lectins and HD-ZIP Transcription Factors in Isoprenoid Based Plant Stress Responses**

SANGITA KUMARI[1], SMRITI SHRIDHAR[1], DALJIT SINGH[1], PIYUSH PRIYA[1], ROHIT FARMER[1], JASREET HUNDAL[1], PRIYANKA SHARMA[1], KRUTIKA BAVISHI[1], KATHRIN SCHRICK[2] and GITANJALI YADAV[1]*

[1]*Computational Biology Laboratory, National Institute of Plant Genome Research, New Delhi 110 067, India*
[2]*Division of Biology, Kansas State University, Manhattan, Kansas, USA*

(Received 26 July 2011; Revised 15 March 2012; Accepted 29 March 2012)

It was over half a century ago when the overwhelming array of chemicals found in plants was postulated to be more than just by-products of primary metabolism. Ever since, extensive research has been conducted on plant secondary metabolites which are now known to be the end points of sophisticated survival mechanisms that plants have developed as a response to various kinds of stresses. Stress, defined by its negative effect on the growth and development of an individual, can be internal (metabolic or genetic), external (biotic or abiotic), permanent or acute. To cope, organisms must develop tolerance, resistance or avoidance mechanisms. Isoprenoids, often released as volatiles from plants, constitute the most diverse groups of natural products and play an essential part in plant defense systems, both directly (as emitted volatiles) and indirectly (the principle of inviting friends to feast on foes). Research over the last decade has resulted in a significant improvement in our understanding of the isoprenoid biosynthesis but there remains much to learn about the complex regulatory network controlling the various steps of these pathways and their dynamic co-ordination. Here we identify novel plant proteins and provide a putative role for them in isoprenoid based stress responses, along with insights into future perspectives for research.

Key Words : Isoprenoid Biosynthetic Pathway; Systems Biology; Network Analysis; Structural Modelling; Lectins; HD-ZIP Transcription Factors

Introduction

For thousands of years natural products have been harvested for their medicinal properties in an effort that is now called Bioprospecting. The astonishing chemical diversity of biologically active substances in various organisms reflects an equally staggering diversity in function. Isoprenoids represent one of the most varied class of natural compounds in all organisms but their largest structural diversity occurs in plants where their roles range from primary metabolism with functions in photosynthesis, respiration, regulation of growth and development to secondary metabolism such as role in ecological interaction of plants with the environment [1-3]. The classification of isoprenoids is based on the number of five carbon units they have. The important classes of isoprenoids are hemiterpenes (C_5), monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), triterpene (C_{30}) and tetraterpenes (C_{40}). Isoprene, the simplest isoprenoids having a single C_5 unit is a volatile product significantly released by some plants. The emission of isoprene by plants is an important component of atmospheric biochemistry. Isoprene-emitting plants are better able to resist thermal and oxidative stress as compared to non-emitting plants

**Author for Correspondence : E-mail: gy@nipgr.ac.in; Tel: +91-11-26735103; Mob. 9911126926*
**Based on the talk given by Dr. Gitanjali Yadav for INSA Medal for Young Scientists in 2011

[4]. Monoterpenes are the major constituent of essential oil responsible for plant originated flavors and fragrances. In addition to being component of essential oils, many sesquiterpenes are reported to act as phytoalexin, antibiotic and feeding deterrent that may help plants in defense and communication. The diterpenes are pharmacologically most important class of isoprenoids that include anticancerous taxol and forskolin (glaucoma treatment). The triterpenes and tetraterpenes include some vital primary metabolites such as brassinosteroids, phytosterols and carotenoids that are essential and usually evident [3].

Despite the diversity in function and structure, all isoprenoids derive from the universal C_5 precursors, isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). The fundamental pathway of terpenoid biosynthesis has three distinct stages. The first stage in the pathway results in formation of the basic C_5 isoprene units. This stage proceeds via two alternative, spatially separated pathways, the classical cytosolic mevalonic acid (MVA) pathway and the recently discovered, mevalonate independent, plastidic, methyl-erythritol phosphate (MEP) pathway as shown in Fig. 1. Interestingly, it was through bioinformatics and computational genomics efforts that the enzymes of the MEP pathway were first elucidated [5]. This

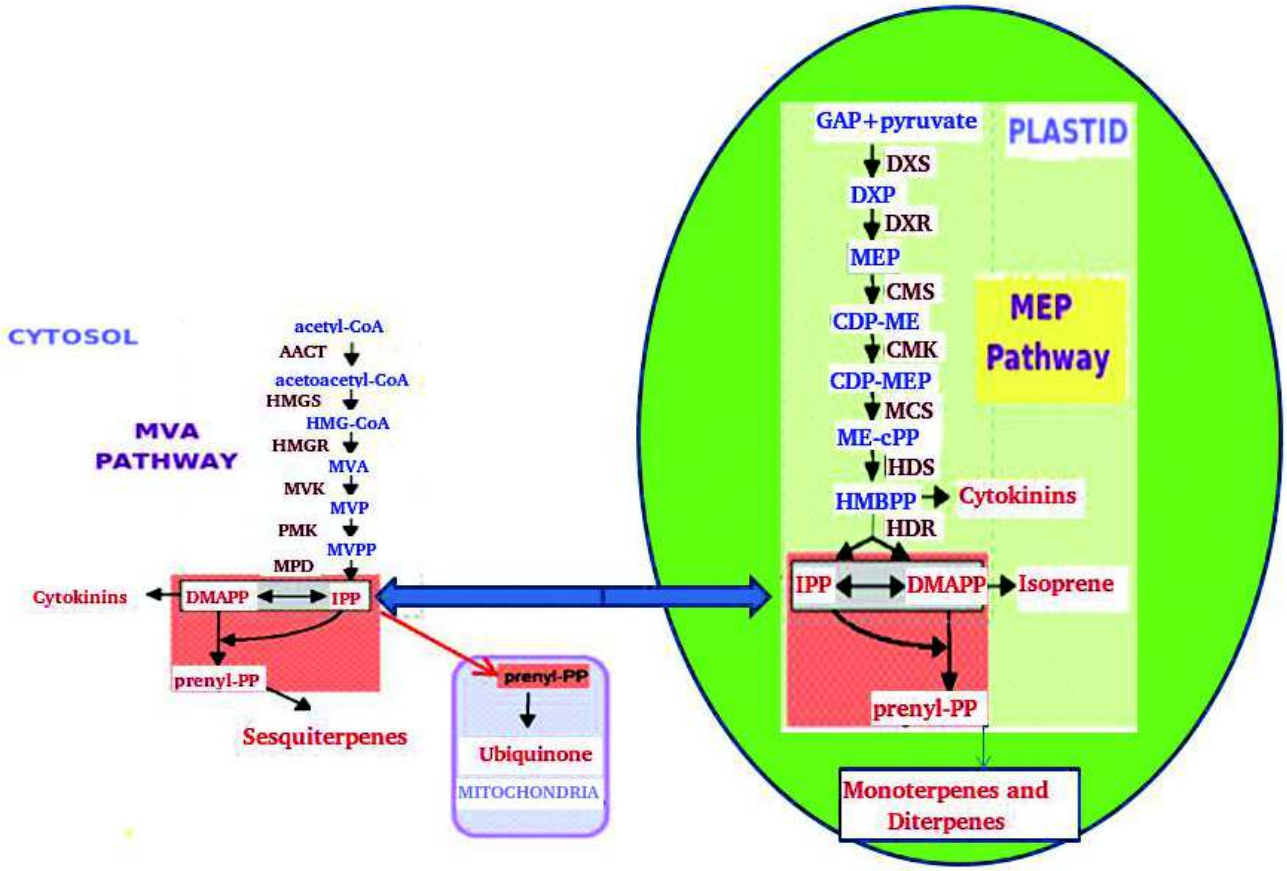

Fig. 1: Diagrammatic depiction of Isoprenoid Biosynthesis representing MVA and MEP pathway. (Left) cytosolic mevalonate (MVA) pathway: AACT, Acetoacetyl-CoA thiolase; HMGS, 3-hydroxy-3-methylglutaryl coenzyme A synthase; HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A; HMGR, 3-hydroxy-3-methylglutaryl coenzyme A reductase; MVA, mevalonate; MVK, mevalonate kinase; MVP, mevalonate phosphate; PMK, phosphomevalonate kinase; MVPP, mevalonate diphosphate; MPD, mevalonate diphosphate decarboxylase; DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; Prenyl-PP, prenyl diphosphate; (Right) The plastidic MEP pathway: GAP, glyceraldehyde 3-phosphate: DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXP, 1-deoxy-D-xylulose-5-phosphate; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MEP, methyl-erythritol phosphate; CMS, 2C-methyl-D-erythritol-4-phosphate cytidylyltransferase; CDP-ME, 4-Diphosphocytidyl-2-C-methyl-D-erythritol; CMK, 4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol kinase; CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate; MCS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; ME-CPP, methyl- erythritol 2,4-cyclodiphosphate; HDS, hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase; HMBPP, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate; HDR, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase.

pathway is also a strong candidate for the discovery of novel antimicrobial and fungicidal drug candidates since it is absent in humans but essential for viability in several major pathogens, like the malarial parasite *Plasmodium falciparum* [6]. One of the major breakthroughs in the immune response to microbial pathogens came when an intermediate of the MEP pathway, namely (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate (HMB-PP), was identified as the most potent activator known till date of the gammadelta T-cells [7]. The second isoprenoid biosynthetic stage is the condensation of two or three C_5 units to form C_{10} C_{15} or C_{20} prenyl diphosphates (PDPs), catalyzed by short chain prenyl transferases. PDPs such as geranyl diphosphate (GDP), farnesyl diphosphate (FDP) and geranylgeranyl diphosphate (GGDP) are the immediate precursors to mono-, sesqui- and di-terpenes respectively. The second stage involves condensation of two or three C_5 units to form C_{10} C_{15} or C_{20} prenyl diphosphates (PDPs), catalyzed by short chain prenyl transferases. In the final stage of terpene biosynthesis, these PDPs are converted to end products by a large family of unique enzymes called the terpene synthases (TPSs). In the final stage of terpene biosynthesis, these PDPs are converted to end products by a large family of unique enzymes called the terpene synthases (TPSs). One of the most outstanding properties of these enzymes is their ability to make multiple products from a single substrate.

Most of the work in characterization of the MEP pathway has been done with *Escherichia coli* proteins including the crystal structure of many of its component enzymes. By contrast only limited knowledge is available about the catalytic properties of plant enzymes. Several experiments involving terpenoid production have suggested that a significant coordination exists between both the MEP and the carotenoid pathways through the control of expression of key genes [8]. Although enzymes of isoprenoid biosynthesis have been elucidated and extensively studied through enzymology and crystallography [5] and considerable knowledge has been gained over the last two decades, there are several questions on physiological stress responses in plants involving individual proteins and genes that remain unanswered, thereby making the exploration of this acclimatization process and its regulation a complex and potentially rewarding field of study. Such a study is particularly relevant in an era when the complete sequences of several different organisms are now becoming available for the first time. Recent advances in genomics make it possible to identify a large number of genes as candidates for involvement in particular processes. For example, a computational genome wide analysis of functional linkages led to the identification of 27 novel pathways in bacteria [9]. A large number of genes encoding enzymes that synthesize volatile compounds have been reported. This well studied area is ideal for applying systems biology and for investigation of evolution not only of the enzymes individually but also of the pathways as a whole. In this report, we attempt to integrate recently gained knowledge from our ongoing work, with previously unknown aspects in this area, to place newly identified proteins on the isoprenoid based stress response map.

Methods

Data Compilation

This step involved a comprehensive large scale compilation of data, including genes, proteins, regulators, as well as interactions involved in the various events during terpene metabolism. Sequence and Structural data was compiled for both systems under focus, namely the MVA and the MEP pathways. Protein/Genetic Interaction data and mRNA expression values in addition to regulatory data were collected from Yeast (*Saccharomyces cerevisae*) (http://www.ncbi.nlm.nih.gov/genome/15), Arabidopsis (*Arabidopsis thaliana*) (http://www.ncbi.nlm.nih.gov/genome/4) and Rice (*Oryza sativa*) (http://www.ncbi.nlm.nih.gov/genome/10). The most comprehensive interaction data available to date is for *S. cerevisiae* at the BIOGRID website [10]. We also used STRING [11], and PAIR databases [12] for this analysis. Expression profile analyses were carried out using PRIMe [13], MPSS [14], GENEVESTIGATOR [15] and VIRTUAL PLANT [16]. Protein Evolutionary rates and abundance measures for both the systems were extracted and collected from literature through PUBMED. Gene

Ontology Data and Sub cellular localization data was extracted and compiled for the two systems from TAIR [17] and public databases. Homologues of various genes and enzymes involved in the MVA and MEP pathways were identified by a comparative analysis of available genomic and proteomic data. Specific sequence motifs and catalytic residues were identified for functional annotation of uncharacterized proteins. Genomic information was compared with 3D structural data wherever possible to enhance the predictive accuracy.

Computational Analyses

Pairwise sequence alignment were carried out using BLAST [18], and the PHYLIP suite of programs [19], EMBOSS [20] and CLUSTALW [21] were used for phylogenetic analyses. Network analysis was carried out in Cytoscape [22], Pajek [23], R packages, in-house FORTRAN and perl programs and UNIX scripting. For structural and cavity architectural analyses, we have made use of Discovery Studio version 3.0, VOIDOO [24], SCWRL [25], R packages, VMD [26] and PyMOL [27].

Results

Pathway Evolution

An attempt to understand the evolution of the two pathways revealed that the mevalonate pathway (MVA) enzymes all have archaebacterial ancestors, and show very frequent horizontal gene transfer. This pathway is widespread in archaea. In contrast, the MEP or non-mevalonate pathway is completely absent in archaea and the pathway enzymes show varied ancestries. For example, the enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) appears to

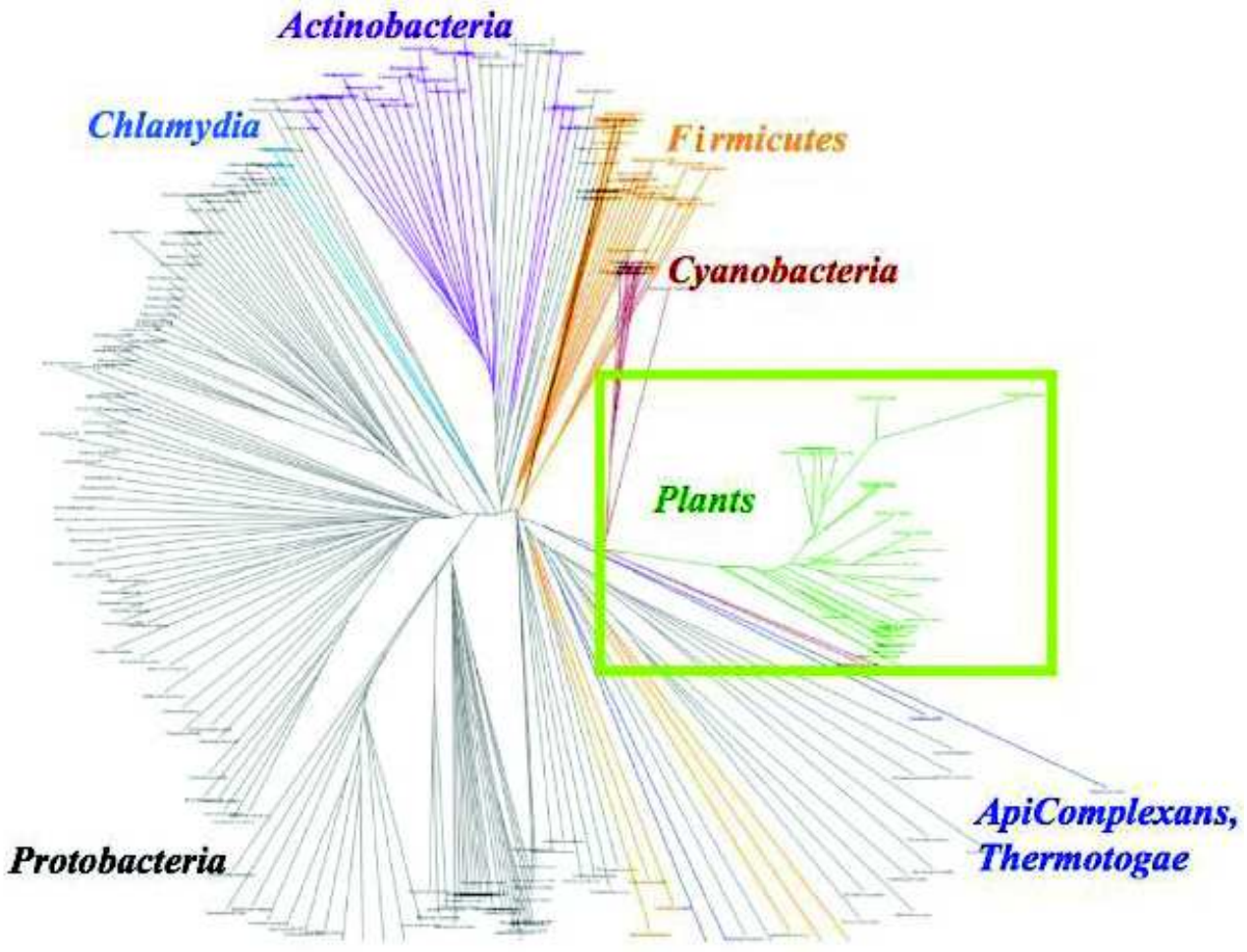

Fig. 2: Phylogenetic tree for the (a) 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)

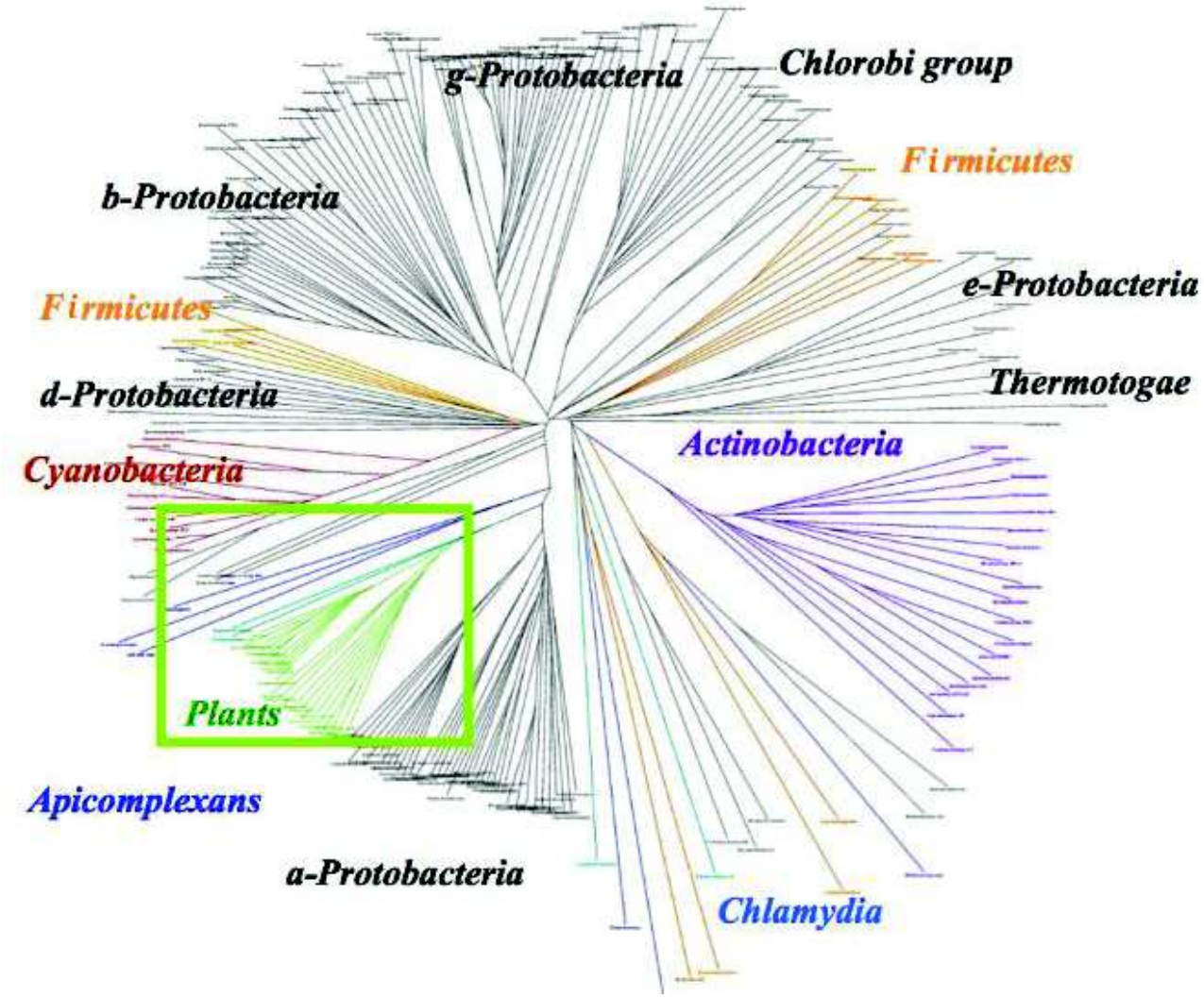

Fig. 2: Phylogenetic tree for the (b) 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) enzymes of the Non-mevalonate pathway showing their varied ancestral decent. Green branches represent plants, other represent lower taxa such as proteobacteria (black), Chlamydia (cyan), Actinobacteria (purple), Cyanobacteria (red), and Firmicutes (orange). Note that plant DXR forms a sub-cluster within the cyanobacterial branch in this lineage, while the DXPS enzymes in plants appears to have evolved from a-proteobacteria (Gram negative organisms having different shapes)

have evolved from a-proteobacteria while 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) from cyanobacteria as shown in Fig. 2 (a & b). 2-C-methylerythritol-4-phosphate cytidyltransferase (MCT) and 4-(cytidine-5'-diphospho)-2-C-methylerythritol kinase (CMK) have chlamydian ancestors, whereas 2-C-methylerythritol-2,4-cyclodiphosphate synthase (MECPS) has reveals and aquifex ancestry. Several complexities are associated with these two pathways, which make them very difficult to analyse, the most important being the nomadic behaviour of the enzymes involved. The major controversy regarding the MEP pathway is its much debated cyanobacterial origin. The alternative hypothesis advocates the post-plastidial origin of this pathway via horizontal gene transfer. The last comprehensive study for these pathways was performed over a decade earlier and it included just about 14 organisms [28]. The current analysis, for the first time, includes information from more than 300 taxa, enabling much better sampling of sequence space and refinement of evolutionary inferences, than previously reported.

Network Analysis and Interpretations

Using an integrated systems approach that enables assessment of multiple datasets together, rather than one source of data at a time, the information compiled for the isoprenoid biosynthetic pathways was superimposed in order to reduce data dimensionality, enabling a systems-level analysis of the various aspects and processes involved in plant stress response. The network analysis is based upon graph theory and our superimposition included data from sequence annotation, functional characterization, gene architecture, protein-protein interaction, genetic interactions, mRNA expression, regulatory interactions of transcription factors, gene ontology information and sub-cellular localization data. In this

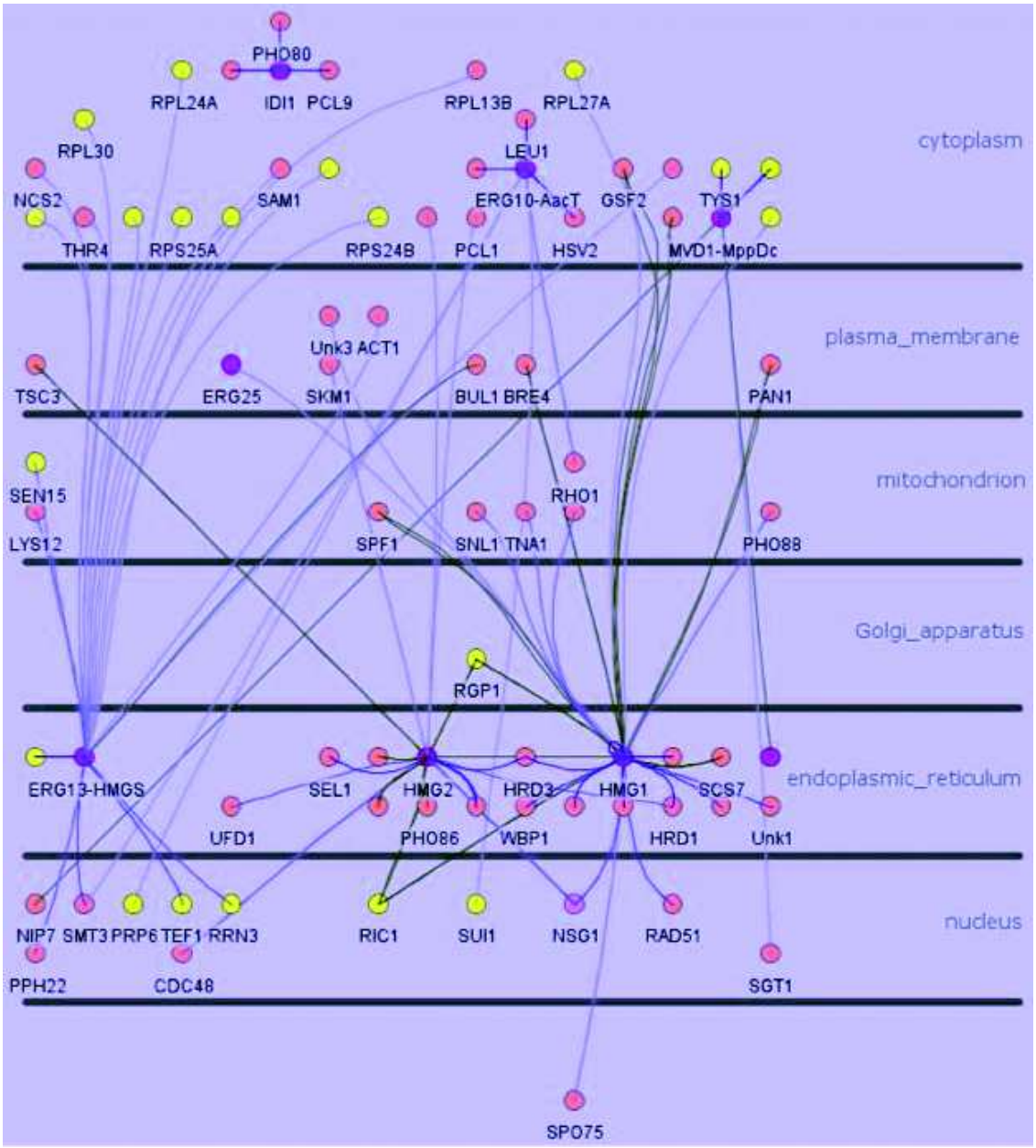

Fig. 3: Network Analysis of the mevalonic acid pathway showing integration of data from various sources as described in text. This model network of interactions was used to identify putative transcription factors which may be involved in the regulation of the isoprenoid pathways. Each circle in this figure represents a gene and each edge represents an interaction. Both physical (blue edges) and genetic (mint edges) interactions have been depicted. Yellow circles denote transcription factors while magenta circles represent pathway enzymes. Sub-cellular locations are also depicted, wherever known.

manner, a systems level understanding of the enzymes involved in the pathways was facilitated as an isoprenoid biosynthetic network shown in Fig. 3. Orthologs of the widely characterized yeast (*S. cerevisiae*) MVA genes and their reported interactors were searched in the available plant genomic data. In case of *A. thaliana*, this search resulted in a set of putatively interacting 'Interlog pairs'. Preliminary verification of predicted Interlog pairs was done by sub-cellular location data as well as mRNA expression levels, wherever available. In case of *S. cerevisiae*, 156 interactions have been reported for the eight genes of MVA pathway. For each of these 156, a study of the predicted *A. thaliana* interlogs was carried out to identify novel molecular players in the isoprenoid network that may have a role to play in regulation or cross-talk between the two pathways.

This analysis revealed the involvement and existence of novel genes, proteins and putative transcription factors. For example a total of twelve transcription factors were identified among the hydroxy-methyl glutaryl CoA synthase (HMGS) interlogs in Arabidopsis, of which two belonged to the type-III HD-ZIP family and contain a unique domain hitherto functionally uncharacterized in plants. Since HMGS is one of the key enzymes in the pathway, we studied this domain further in order to understand its possible implications, and the details are presented in the next section.

In addition to the interaction data, the isoprenoid network described above contained expression data also. A preliminary scan of the Arabidopsis genes was carried out to identify gene families enriched for induction during isoprenoid biosynthetic pathway expression. Among these, it was interesting to note that genes of enzymes in both the mevalonate and non-mevalonate pathways were found to co-express with several members of the lectin superfamily. In order to better understand these lectins and their roles in stress, we studied this superfamily in greater detail, as explained in the penultimate section.

Interaction Data and the Start Domain

The candidate plant domain identified from interaction data was classified as the steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domain, a protein domain spanning ~210 residues that has been known to bind and transport lipid molecules to the inner mitochondrial membrane in mammals [29]. This domain functions in a variety of distinct mammalian physiological processes, such as lipid transfer between intracellular compartments, lipid metabolism and modulation of signaling events. Mutation or misexpression of START proteins is linked to pathological processes, including genetic disorders, autoimmune disease and cancer [30]. However, the precise function of this domain in plants has remained obscure despite the recognition that it is highly amplified in plant genomes [31]. Sequence data from recently completed Plant Genome Projects showed that *Vitis vinifera* has 26 such domains, while *Populus* and *Physcomitrella patens* subsp. *Patens* have eight and 20 domains respectively. Arabidopsis has 35 homologs of this domain in its genome and several of these domains are present in homeodomain transcription factors (HD-START), suggesting a role in regulation of key cellular processes [32]. Fig. 4 shows an alignment of these domains with the closest structural counterparts depicting the secondary structural segments as well as the consensus regions. The identification of two of these domains in the transcription factors associated with the isoprenoid network of Arabidopsis implies their role in recognition of terpene based signals, possibly terpene based activation and downstream regulation of stress responses in plants. The domain may be hypothesized as a ligand-binding module that regulates transcription factor activity, since a few HD-START transcription factors have been reported to play key roles in plant cell development of disease resistance [33]. Other domains that are frequently found to be associated with this domain are bZip_1, MEKHLA, PH, DUF1336 and RPN2 (26S proteasome regulatory complex component) [31].

The mammalian START domains have a conserved 'helix-grip' fold that forms an inner tunnel wide enough to accommodate hydrophobic lipids such as cholesterol or ceramides [34]. Although the identity of the lipids that bind each START domain is known for only a few members of the family, it is

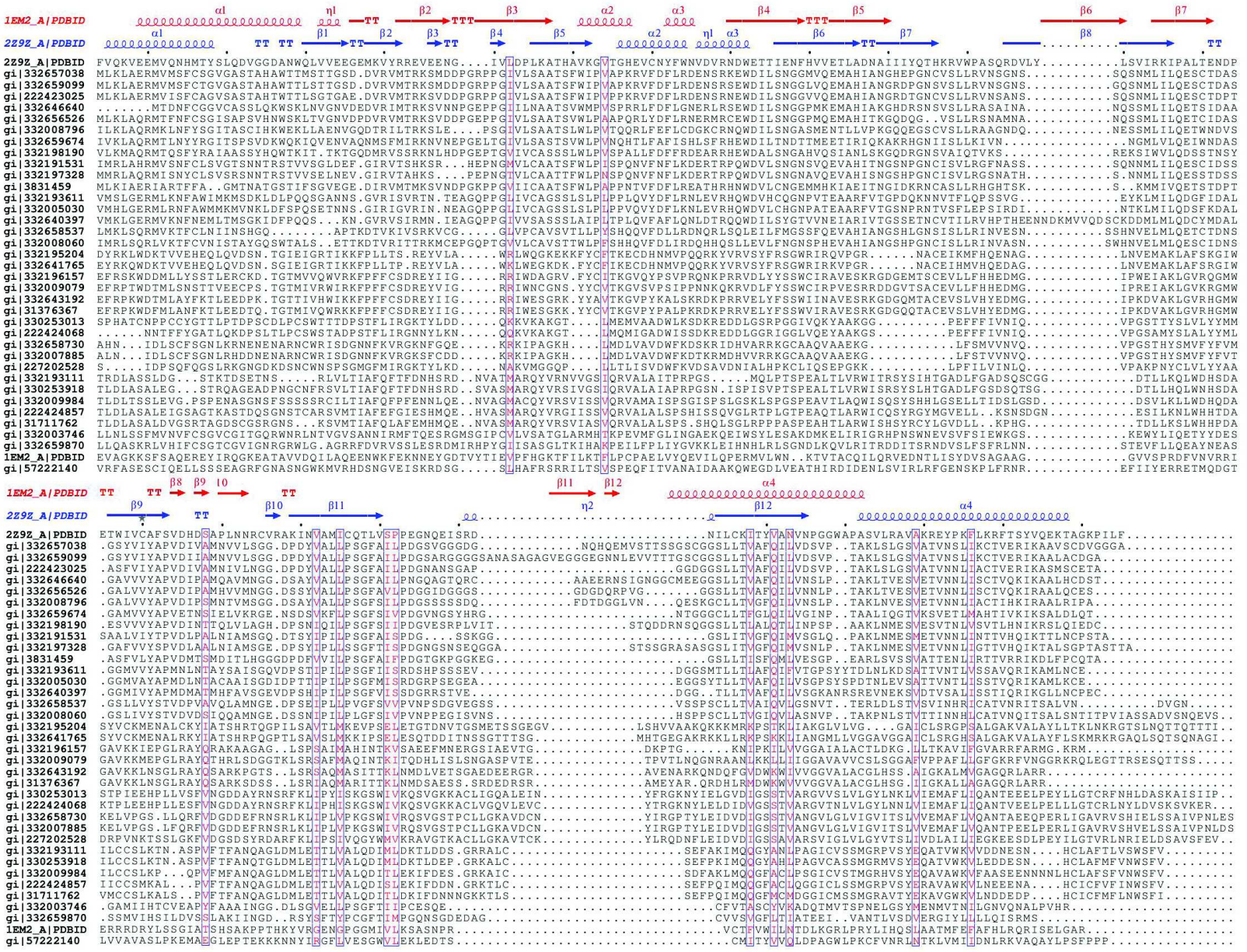

Fig. 4: Alignment of the 35 Arabidopsis START domain proteins with their mammalian counterparts, depicting the placement of secondary structural elements and the high conservation in sequence. This conservation in sequence enabled reliable structural models to be built using the mammalian structures as templates for plant sequences

believed that identification of the ligand can help in assigning distinct physiological and pathological roles to different START-domain-containing proteins. For this, it would be essential to study the inner tunnels of the respective domains, which aids, or severely restricts the correct substrate by virtue of its shape, size and chemical properties. Accordingly, structural modelling of the ligand binding tunnel/pocket was carried out for a few of the Arabidopsis START domains. This involved mutational analysis of the domains with vertebrate START proteins serving as a template to probe the function of the plant START domain as a regulatory molecule. Detailed structural analysis of full models as well as chimeras generated from *in silico* mutations of residues known to be lining the active site cavities (as per crystal structure data) revealed that in contrast to the large cholesterol binding tunnels of mammalian domains, most plants domains may have small and/or medium sized cavities (Fig. 5). A detailed study of steric constraints in the plant protein pockets revealed that plant START domains may be a novel class of previously unknown lipid translocators with regulatory functions (GY unpublished data). This prediction requires experimental verification that is being presently addressed and has shown highly promising results (KS, personal communication to GY). The next step is to identify the possible terpene based ligands that would fit the small plant START cavities, by employing various strategies like expression patterns, sub-cellular localization patterns, roles of other domains attached to these domains, and computational efforts like molecular docking approaches. Further progress requires molecular docking efforts for analysis of enzyme substrate complexes with the goal of explaining and ultimately predicting the stereo-specificity and substrate specificity of these enzymes.

Expression Data and the Lectin Connection

Lectins have been known to be synthesized and accumulated after microbial attack and related situations and their roles in the natural defense mechanism of plants have been well documented [35]. Upon binding to carbohydrate or carbohydrate containing proteins, and after being ingested by herbivores, lectins bind to epithelial cell lining of the digestive tracts and interfere with nutrient absorption [36]. Our finding that the isoprenoid biosynthetic pathways are co-induced with specific lectins led us to further investigate this superfamily across known plant genomes. Due to the huge diversity in sequence and carbohydrate specificity of lectin domains across different organisms, the first obstacle in a comprehensive analysis was the lack of methods that could improve glycoprotein detection, and/or

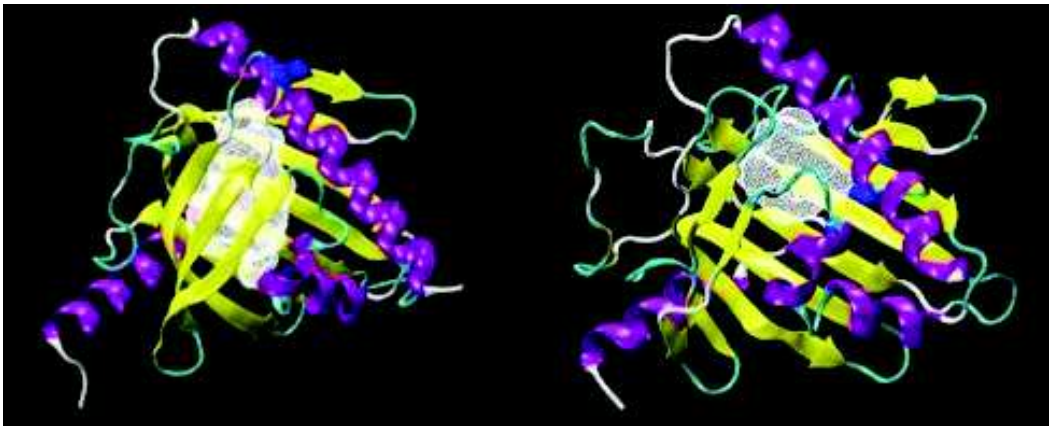

Fig. 5: Structural modelling of START domains from *A. thaliana*. The mammalian template domain (left) and plant model domain (right) are shown, depicting the conspicuous difference in cavity architecture, size and volume despite similarity in overall sequence as observed in multiple alignments in Fig. 4. Note that the mammalian cavity (white mesh; left) is much larger than the plant cavity (white mesh; right), which sterically cannot accommodate a cholesterol molecule (with an approx spatial volume of 700A^3). We believe this is due to specific residues that block the START tunnel extension area in plants. All tunnels are depicted in mesh representation, domains helices are magenta while the beta sheet and turn regions are depicted in yellow and cyan respectively.

prediction of their sugar binding specificity. Our first effort therefore, was to develop an algorithm for identification, annotation and assignation of broad substrate specificity to plant lectin domains, using amino-acid information, and based upon a rigorous analysis of evolutionary relationships between several hundred characterized polypeptide sequences of lectin domains with known specificities, followed by optimization of all parameters for building accurate profile Hidden Markov Models (HMMs) [37]. A comparison of our results with those obtained from general protein annotation databases like PFam, Smart-db and Panther-db, revealed that our algorithm (PLecDom) performed better than any of these programs. Encouraged by the success of our approach, and its ease with handling large scale input, we applied the program to genomic, protein as well as EST data from over fifty plant genome sequencing projects currently underway [37]. In order to make the program accessible to researchers in the area, we converted the algorithm into an online automated server with a very simple user friendly interface, and help pages. This server is available freely to the scientific community at http://www.nipgr.res.in/plecdom.html.

More than 7000 plant lectins were identified in this effort, and the specific domains identified in network analysis were analysed for their possible roles in the isoprenoid biosynthetic pathway regulation and/or induction, by comparing detailed expression profiles across tissues, developmental stages and stress conditions in both *A. thaliana* and *O. sativa*. The results are depicted in Fig. 6. As can be seen in Fig. 6a, the expression profiles of isoprenoid pathway enzymes are positively correlated with several lectin genes in the MVA and MEP pathways of Arabidopsis. In all, we have identified 38 lectins that are upregulated during MVA pathway induction in Arabidopsis, and 58 lectins whose expression is negatively correlated with MVA pathway genes. Further, we have identified 42 lectins whose expression profiles are upregulated during MEP pathway induction and 39 lectins that were downregulated with respect to the MEP pathway in Arabidopsis. Similarly, in case of the rice, expression profiles of 61 and 90 lectins were found to correlate

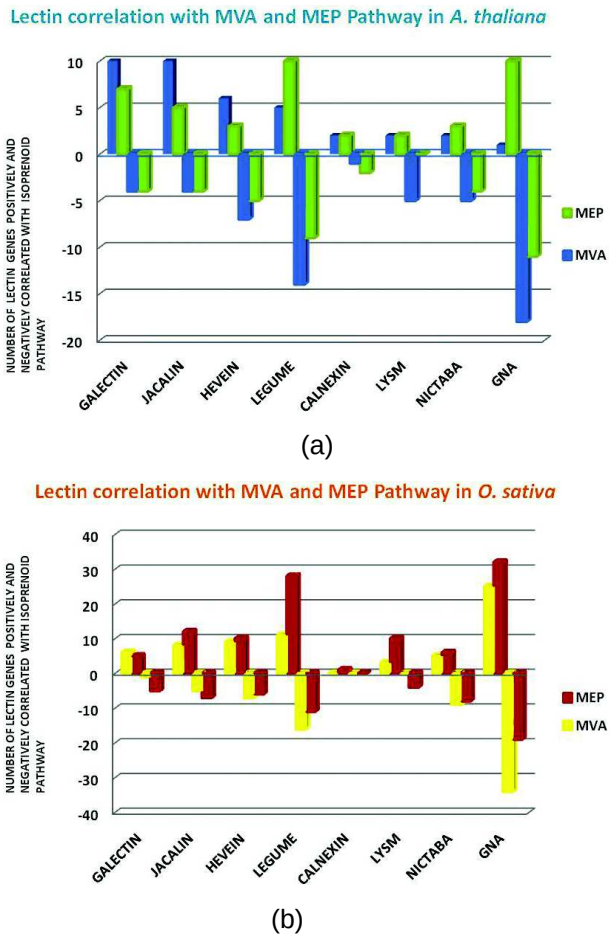

Fig. 6(a&b): Lectin correlation with the enzymes of isoprenoid biosynthetic pathway (MVA and MEP) in (a) *A. thaliana* and (b) *O. sativa.* In this figure each bar represents the number of lectin genes positively and negatively correlated with cytosolic Mevalonic Acid (MVA) and plastidic 2-C-Methyl-D-Erythritol 4-Phosphate (MEP) Pathway. Green bars represent plastidic MEP pathway whereas blue bars correspond to MVA pathway in *A. thaliana.* In case of *O. sativa,* red bars denote plastidic MEP pathway and yellow bars correspond to MVA pathway. In each case, positive Y-axis represents genes that are positively correlated whereas negative Y-axis represents the number of genes that are negatively correlated. Note that GNA-related and Legumes are the major correlates among all lectin families indicating the vital role of these two lectin superfamilies in isoprenoid based stress response. Lectins are classified by their sequence based families, namely, Galectin (galactose-binding lectin); Jacalin [lectin found in *Artocarpus integrifolia* (jackfruit) seed]; Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *Hevea brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in endoplasmic reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba [The tobacco (*Nicotiana tabacum*) agglutinin]; GNA (*Galanthus nivalis* agglutinin related lectin family)

Table 1a: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MVA) in *A. thaliana.* The first column lists accession numbers of lectin genes found to be positively correlated with MVA pathway of *A. thaliana.* Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MVA pathway of *A. thaliana,* respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit) seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *Hevea brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba (The tobacco (*Nicotiana tabacum*) agglutinin); GNA (*G. nivalis* agglutinin related lectin family).

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT1G02360	Hevein domain	AT1G05850	Hevein domain
AT1G05850	Hevein	AT1G08450	Calreticulin/Calnexin
AT1G09210	Calreticulin/Calnexin	AT1G11340	GNA
AT1G11280	GNA	AT1G11350	GNA
AT1G21880	LysM lectin	AT1G21880	LysM
AT1G27120	Galectin	AT1G34300	GNA
AT1G45130	Galectin	AT1G55000	LysM
AT1G52060	Jacalin	AT1G61420	GNA
AT1G52070	Jacalin	AT1G61460	GNA
AT1G53070	Legume	AT1G61610	GNA
AT1G56340	Calreticulin/Calnexin	AT1G63090	Nictaba
AT1G56680	Hevein	AT1G65390	Nictaba
AT1G74800	Galectin	AT1G65790	GNA
AT1G77630	LysM	AT1G65800	GNA
AT2G02230	Nictaba	AT1G67520	GNA
AT2G28470	Galectin	AT1G73040	Jacalin
AT2G33070	Jacalin	AT1G77630	LysM
AT2G39310	Jacalin	AT1G78820	GNA
AT2G43590	Hevein	AT1G78860	GNA
AT2G43610	Hevein	AT2G02360	Nictaba
AT3G13750	Galectin	AT2G19130	GNA
AT3G16390	Jacalin	AT2G23770	LysM
AT3G16430	Jacalin	AT2G33580	LysM
AT3G16440	Jacalin	AT2G37710	Legume
AT3G16450	Jacalin	AT2G43570	Hevein
AT3G16460	Jacalin	AT2G43620	Hevein

(Table 1a contd ...)

(Continuation of Table 1a)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT3G16470	Jacalin	AT2G43730	Jacalin
AT3G52840	Galectin	AT3G04720	Hevein
AT3G53000	Nictaba	AT3G08870	Legume
AT3G53380	Legume	AT3G12500	Hevein
AT3G54080	Legume	AT3G16390	Jacalin
AT4G01700	Hevein	AT3G16460	Jacalin
AT4G21060	Galectin	AT3G16530	Legume
AT4G36360	Galectin	AT3G52840	Galectin
AT5G01090	Legume	AT3G53000	Nictaba
AT5G20710	Galectin	AT3G53380	Legume
AT5G55830	Legume	AT3G53810	Legume
AT5G56870	Galectin	AT3G54420	Hevein
		AT3G59700	Legume
		AT4G01700	Hevein
		AT4G02410	Legume
		AT4G04960	Legume
		AT4G11900	GNA
		AT4G19840	Nictaba
		AT4G27300	GNA
		AT4G28350	Legume
		AT4G29050	Legume
		AT4G32300	GNA
		AT5G01540	Legume
		AT5G01550	Legume
		AT5G03350	Legume
		AT5G18470	GNA
		AT5G35370	GNA
		AT5G56870	Galectin
		AT5G60270	Legume
		AT5G60900	GNA
		AT5G62620	Galectin
		AT5G63800	Galectin

Table 1b: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MEP) in *A. thaliana.* The first column lists accession numbers of lectin genes found to be positively correlated with MEP pathway of *A. thaliana.* Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MEP pathway of *A. thaliana,* respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *H. brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba (The tobacco (*N. tabacum*) agglutinin); GNA (*G. nivalis* agglutinin related lectin family)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT1G05850	Hevein	AT1G11280	GNA
AT1G08450	Calreticulin/Calnexin	AT1G12710	Nictaba
AT1G09210	Calreticulin/Calnexin	AT1G34300	GNA
AT1G11300	GNA	AT1G56340	Calreticulin/Calnexin
AT1G11330	GNA	AT1G56680	Hevein
AT1G11350	GNA	AT1G61370	GNA
AT1G12710	Nictaba	AT1G61550	GNA
AT1G19715	Jacalin	AT1G67520	GNA
AT1G45130	Galectin	AT1G74800	Galectin
AT1G51940	LysM	AT1G78860	GNA
AT1G55000	LysM	AT2G02230	Nictaba
AT1G65790	GNA	AT2G26820	Nictaba
AT1G65800	GNA	AT2G33070	Jacalin
AT1G74800	Galectin	AT2G43580	Hevein
AT1G78820	GNA	AT2G43610	Hevein
AT1G78860	GNA	AT2G43690	Legume
AT2G02360	Nictaba	AT3G04720	Hevein
AT2G32810	Galectin	AT3G21380	Jacalin
AT2G37710	Legume	AT3G45420	Legume
AT2G39310	Jacalin	AT3G51710	GNA
AT2G43620	Hevein	AT3G53000	Nictaba
AT3G04720	Hevein	AT3G53380	Legume
AT3G08870	Legume	AT3G53810	Legume
AT3G13750	Galectin	AT3G54080	Legume
AT3G16390	Jacalin	AT3G54420	Hevein
AT3G16450	Jacalin	AT4G03230	GNA

(Table 1b contd ...)

(Continuation of Table 1b)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT3G16460	Jacalin	AT4G04960	Legume
AT3G16530	Legume	AT4G21060	Galectin
AT3G52840	Galectin	AT4G21390	GNA
AT4G02410	Legume	AT5G03700	GNA
AT4G02420	Legume	AT5G06740	Legume
AT4G11900	GNA	AT5G18470	GNA
AT4G19840	Nictaba	AT5G38540	Jacalin
AT4G27300	GNA	AT5G38550	Jacalin
AT4G29050	Legume	AT5G55830	Legume
AT5G01090	Legume	AT5G59260	Legume
AT5G01540	Legume	AT5G61790	Calreticulin/Calnexin
AT5G01550	Legume	AT5G62620	Galectin
AT5G03350	Legume	AT5G63810	Galectin
AT5G56870	Galectin		
AT5G60900	GNA		
AT5G63800	Galectin		

positively with MVA and MEP pathway enzyme genes respectively, whereas 69 and 54 lectins were negatively correlated with pathway enzyme expression profiles. This is interesting in view of the fact that none of the pathway enzymes in either pathway has any direct interaction with lectins. Overall, the data shows that Arabidopsis has 68 genes from eight family of lectins upregulated with isoprenoid biosynthetic pathway during various expression studies, while in case of rice; 1-deoxy-D-xylulose 5-phosphate synthase (DXPS) enzyme of the MEP pathway has the maximum number of positively correlated lectins among all the enzymes of these pathways. This enzyme has also shown upregulation in various biotic stress conditions as suggested by MPSS data analysis (data not shown). In summary, the data suggests that it is the *G. nivalis* agglutinin (GNA) related lectins and Legumes that play substantial roles in isoprenoid related stress responses. A list of the identified lectins given along with their family classification is provided in Table 1 (for *A. thaliana*) and Table 2 (for *O. sativa*). Efforts to further characterize these roles for each lectin identified are currently underway in our laboratory.

Conclusion

As mentioned earlier, the vast amount of publicly available data offers a unique opportunity to translate genomic information into useful new biological knowledge for understanding terpene biosynthetic diversity, its regulation, and the molecular basis of scent emission, using sequence and structural data, background biological knowledge and novel inference techniques, that could eventually provide experimentally testable predictions of mutations leading to the precise production of specified isoprenoid end products. The molecular mechanisms by which specific isoprenoids influence stress

Table 2a: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MVA) in *Oryza sativa.* The first column lists accession numbers of lectin genes found to be positively correlated with MVA pathway of *O. sativa.* Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MVA pathway of *O. sativa,* respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit) seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *H. brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba (The tobacco (*N. tabacum*) agglutinin); GNA (*G. nivalis* agglutinin related lectin family)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os01g18400	Hevein	LOC_Os01g06500	Nictaba
LOC_Os01g24710	Jacalin	LOC_Os01g18400	Hevein
LOC_Os01g25280	Jacalin	LOC_Os01g25280	Jacalin
LOC_Os01g48040	GNA	LOC_Os01g47900	GNA
LOC_Os01g57560	GNA	LOC_Os01g48040	GNA
LOC_Os01g65030	GNA	LOC_Os01g57560	GNA
LOC_Os01g66610	GNA	LOC_Os01g65030	GNA
LOC_Os01g66680	GNA	LOC_Os01g66610	GNA
LOC_Os01g72810	GNA	LOC_Os01g66680	GNA
LOC_Os02g12730	Galectin	LOC_Os02g19530	Legume
LOC_Os02g48200	Legume	LOC_Os02g42780	Legume
LOC_Os02g48210	Legume	LOC_Os02g48200	Legume
LOC_Os02g52850	GNA	LOC_Os02g48210	Legume
LOC_Os02g56750	Nictaba	LOC_Os02g52850	GNA
LOC_Os02g56820	Nictaba	LOC_Os02g56750	Nictaba
LOC_Os03g04060	Hevein	LOC_Os02g56820	Nictaba
LOC_Os03g06940	Galectin	LOC_Os03g02550	Nictaba
LOC_Os03g12150	GNA	LOC_Os03g04060	Hevein
LOC_Os03g60810	Legume	LOC_Os03g12150	GNA
LOC_Os04g01310	GNA	LOC_Os03g30890	GNA
LOC_Os04g03579	Legume	LOC_Os03g56160	Legume
LOC_Os04g09390	Hevein	LOC_Os03g60810	Legume
LOC_Os04g15580	GNA	LOC_Os04g01310	GNA
LOC_Os04g23760	GNA	LOC_Os04g03579	Legume
LOC_Os04g28780	GNA	LOC_Os04g15580	GNA
LOC_Os04g34390	GNA	LOC_Os04g23760	GNA
LOC_Os04g41620	Hevein	LOC_Os04g28780	GNA
LOC_Os04g41680	Hevein	LOC_Os04g34250	GNA
LOC_Os04g42740	GNA	LOC_Os04g34270	GNA

(Table 2a contd ...)

(Continuation of Table 2a)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os04g44900	Legume	LOC_Os04g34290	GNA
LOC_Os04g44910	Legume	LOC_Os04g34390	GNA
LOC_Os04g49480	Legume	LOC_Os04g41620	Hevein
LOC_Os04g53998	GNA	LOC_Os04g41680	Hevein
LOC_Os04g54180	GNA	LOC_Os04g42740	GNA
LOC_Os05g04690	Hevein	LOC_Os04g44900	Legume
LOC_Os06g17490	Legume	LOC_Os04g44910	Legume
LOC_Os06g22290	Legume	LOC_Os04g53998	GNA
LOC_Os06g37750	GNA	LOC_Os04g54130	GNA
LOC_Os06g40030	GNA	LOC_Os04g54180	GNA
LOC_Os06g51050	Hevein	LOC_Os05g04690	Hevein
LOC_Os06g51060	Hevein	LOC_Os05g07300	GNA
LOC_Os07g09670	Galectin	LOC_Os05g13770	Legume
LOC_Os07g36570	GNA	LOC_Os05g42210	GNA
LOC_Os08g03020	Legume	LOC_Os06g06960	GNA
LOC_Os08g03070	Legume	LOC_Os06g22290	Legume
LOC_Os09g27890	LysM	LOC_Os06g37750	GNA
LOC_Os09g28180	GNA	LOC_Os06g40030	GNA
LOC_Os09g37600	LysM	LOC_Os06g51050	Hevein
LOC_Os09g37834	GNA	LOC_Os07g03860	Legume
LOC_Os10g04270	Jacalin	LOC_Os07g03870	Legume
LOC_Os10g06680	GNA	LOC_Os07g03880	Legume
LOC_Os10g18400	Galectin	LOC_Os07g09670	Galectin
LOC_Os10g39680	Hevein	LOC_Os07g36570	GNA
LOC_Os11g03860	GNA	LOC_Os08g03020	Legume
LOC_Os11g05240	GNA	LOC_Os08g03070	Legume
LOC_Os11g39530	Jacalin	LOC_Os08g05480	Nictaba
LOC_Os12g03594	Nictaba	LOC_Os09g28180	GNA
LOC_Os12g03740	Nictaba	LOC_Os10g04270	Jacalin
LOC_Os12g12720	Jacalin	LOC_Os10g06680	GNA
LOC_Os12g30180	Nictaba	LOC_Os10g39680	Hevein
LOC_Os12g44320	GNA	LOC_Os11g03860	GNA
		LOC_Os11g05240	GNA
		LOC_Os11g39530	Jacalin
		LOC_Os12g03594	Nictaba
		LOC_Os12g03740	Nictaba
		LOC_Os12g12720	Jacalin
		LOC_Os12g14440	Jacalin
		LOC_Os12g30180	Nictaba
		LOC_Os12g44320	GNA

Table 2b: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MEP) in *O. sativa.* The first column lists accession numbers of lectin genes found to be positively correlated with MEP pathway of *O. sativa.* Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MEP pathway of *O. sativa,* respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit) seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *H. brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba (The tobacco (*N. tabacum*) agglutinin); GNA (*G. nivalis* agglutinin related lectin family)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os01g06500	Nictaba	LOC_Os01g06240	GNA
LOC_Os01g18400	Hevein	LOC_Os01g18400	Hevein
LOC_Os01g24710	Jacalin	LOC_Os01g24710	Jacalin
LOC_Os01g25280	Jacalin	LOC_Os01g25280	Jacalin
LOC_Os01g36550	LysM	LOC_Os01g48040	GNA
LOC_Os01g47900	GNA	LOC_Os01g57560	GNA
LOC_Os01g48000	GNA	LOC_Os02g12730	Galectin
LOC_Os01g57480	GNA	LOC_Os02g39330	Hevein
LOC_Os01g57560	GNA	LOC_Os02g48200	Legume
LOC_Os01g65030	GNA	LOC_Os02g48210	Legume
LOC_Os01g66250	GNA	LOC_Os02g52850	GNA
LOC_Os01g66610	GNA	LOC_Os02g56750	Nictaba
LOC_Os01g66680	GNA	LOC_Os02g56820	Nictaba
LOC_Os01g72810	GNA	LOC_Os03g02550	Nictaba
LOC_Os02g12730	Galectin	LOC_Os03g06940	Galectin
LOC_Os02g19530	Legume	LOC_Os04g01310	GNA
LOC_Os02g39330	Hevein	LOC_Os04g03579	Legume
LOC_Os02g45320	Nictaba	LOC_Os04g09390	Hevein
LOC_Os02g48200	Legume	LOC_Os04g21130	Nictaba
LOC_Os02g48210	Legume	LOC_Os04g23760	GNA
LOC_Os02g52850	GNA	LOC_Os04g28780	GNA
LOC_Os02g53000	LysM	LOC_Os04g34390	GNA
LOC_Os03g02685	LysM	LOC_Os04g42740	GNA
LOC_Os03g04060	Hevein	LOC_Os04g44900	Legume
LOC_Os03g04110	LysM	LOC_Os04g44910	Legume
LOC_Os03g06940	Galectin	LOC_Os04g49480	Legume

(Table 2b contd ...)

(Continuation of Table 2b)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os03g12150	GNA	LOC_Os04g54180	GNA
LOC_Os03g30890	GNA	LOC_Os05g04690	Hevein
LOC_Os03g56160	Legume	LOC_Os06g17490	Legume
LOC_Os03g60810	Legume	LOC_Os06g22290	Legume
LOC_Os04g01320	GNA	LOC_Os06g37750	GNA
LOC_Os04g03579	Legume	LOC_Os06g51050	Hevein
LOC_Os04g09390	Hevein	LOC_Os07g09670	Galectin
LOC_Os04g15580	GNA	LOC_Os07g36570	GNA
LOC_Os04g22120	Legume	LOC_Os08g03020	Legume
LOC_Os04g23700	GNA	LOC_Os08g03070	Legume
LOC_Os04g34250	GNA	LOC_Os08g03160	Legume
LOC_Os04g34270	GNA	LOC_Os08g05480	Nictaba
LOC_Os04g34290	GNA	LOC_Os09g27890	LysM
LOC_Os04g41620	Hevein	LOC_Os09g28180	GNA
LOC_Os04g41680	Hevein	LOC_Os09g37600	LysM
LOC_Os04g53998	GNA	LOC_Os09g37780	GNA
LOC_Os04g54002	GNA	LOC_Os09g37834	GNA
LOC_Os05g04690	Hevein	LOC_Os10g04270	Jacalin
LOC_Os05g13770	Legume	LOC_Os10g06680	GNA
LOC_Os05g42210	GNA	LOC_Os10g38040	LysM
LOC_Os05g43170	CAL	LOC_Os10g39680	Hevein
LOC_Os05g43240	Jacalin	LOC_Os11g03860	GNA
LOC_Os06g06940	GNA	LOC_Os11g05240	GNA
LOC_Os06g06960	GNA	LOC_Os11g39530	Jacalin
LOC_Os06g17490	Legume	LOC_Os12g03594	Nictaba
LOC_Os06g22290	Legume	LOC_Os12g03740	Nictaba
LOC_Os06g37560	Galectin	LOC_Os12g30180	Nictaba
LOC_Os06g37750	GNA	LOC_Os12g44320	GNA
LOC_Os06g40030	GNA		
LOC_Os06g51050	Hevein		
LOC_Os06g51060	Hevein		

(Table 2b contd ...)

(Continuation of Table 2b)

LOC_Os06g51360	LysM
LOC_Os07g03790	Legume
LOC_Os07g03830	Legume
LOC_Os07g03840	Legume
LOC_Os07g03860	Legume
LOC_Os07g03870	Legume
LOC_Os07g03880	Legume
LOC_Os07g03900	Legume
LOC_Os07g03930	Legume
LOC_Os07g04040	Legume
LOC_Os07g18230	Legume
LOC_Os07g38250	Legume
LOC_Os07g38800	Legume
LOC_Os07g38810	Legume
LOC_Os08g02996	Legume
LOC_Os08g03020	Legume
LOC_Os08g03070	Legume
LOC_Os08g03160	Legume
LOC_Os08g05480	Nictaba
LOC_Os09g27890	LysM
LOC_Os09g37600	LysM
LOC_Os09g37834	GNA
LOC_Os09g37890	GNA
LOC_Os10g04270	Jacalin
LOC_Os10g39680	Hevein
LOC_Os11g10290	GNA
LOC_Os11g39420	Jacalin
LOC_Os11g39530	Jacalin
LOC_Os12g03594	Nictaba
LOC_Os12g14440	Jacalin
LOC_Os12g30180	Nictaba
LOC_Os12g34320	GNA
LOC_Os12g44320	GNA

responses in plants are not clearly understood. This work was carried out with the aim to integrate information at the level of genes, proteins, their interactions and expression profiles, enabling a reliable systems analysis. Insights from the network analysis have led to the identification of a new class of plant amplified transcription factors, which are hypothesized to be regulating at least partially, the biosynthetic crosstalk between the pathways. Detailed sequence and structural analysis of these transcription factors proteins supports the idea that they may be involved in the crosstalk between the MVA-MEP pathways. The most fascinating aspect of this endeavor has been the discovery of several new directions for potential research, including prediction of substrate specificity of plant lectin domains and START domains, and bioprospecting for new molecular players that may have strong agri-biotechnological, pharmaceutical and economic applications.

Acknowledgements

The authors are thankful to the Director, NIPGR for encouragement and support. The authors (SK) and (PP) are recipients of Senior Research Fellowship from CSIR, India. The author (SS) was funded by grants to National Institute of Plant Genome Research under BTIS project, from the Department of Biotechnology, Government of India. One of the authors (KB) was a recipient of the Summer Research Fellowship Programme (SRFP) of the Indian Science Academies at the time of the work. This work was supported by grants to one of the authors (GY) under the Senior Innovative Young Biotechnologist Award (IYBA) by DBT, India.

References

1. Fraenkel GS *Science* **129** (1959) 1466
2. Kirby J and Keasling J D *Annu Rev Plant Biol* **60** (2009) 335
3. Buchanan B B, Gruissem W and Joneas R L (Eds) *Biochemistry and Molecular Biology of Plants* Rodney Croteau, Toni M Kutchan and Norman G Lewis (2000) pp 1250
4. Sharkey TD, Wiberley AE and Donohue AR *Ann Bot* **101** (2008) 5
5. Cane DE *Science* **287** (2000) 818
6. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Turbachova I, Eberl M, Zeidler J, Lichtenthaler H K, Soldati D and Beck E *Science* **285** (1999) 1573
7. Eberl M, Hintz M, Reichenberg A, Kollas A K and Wiesner J and Jomaa H *FEBS Lett* **544** (2003) 4
8. Laule O, Furholz A, Chang H S, Zhu T and Wang X, Heifetz P B, Gruissem W and Lange M *Proc Natl Acad Sci USA* **100** (2003) 6866
9. Date S V and Marcotte E M *Nat Biotechnol* **21** (2003) 1055
10. Stark C, Breitkreutz B J, Chatr-aryamontri A, Boucher L, Oughtred R, Livstone M S, Nixon J, Auken K V, Wang X, Shi X, Reguly T, Rust J M, Winter A, Dolinski K and Tyers M *Nucleic Acids Res* **39** (2011) D698
11. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen L J and Mering C V *Nucleic Acids Res* **39** (2011) D561
12. Lin M, Shen X and Chen X *Nucleic Acids Res* **39** (2011) D1134
13. Akiyama K, Chikayama E, Yuasa H, Shimada Y and Tohge T, Shinozaki K, Hirai M Y, Sakurai T, Kikuchi J and Saito K *In Silico Biol* **8** (2008) 339
14. Nakano M, Nobuta K, Vemaraju K, Tej S S, Skogen J W and Meyers B C *Nucleic Acids Res* **34** (2006) D731
15. Zimmermann P, Laule O, Schmitz J, Hruz T and Bleuler S and Gruissem W *Mol Plant* **1** (2008) 851
16. Katari M S, Nowicki S D, Aceituno F F, Nero D, Kelfer J, Thompson P, Cabello J M, Davidson R S, Goldberg A P, Shasha D E, Coruzzi G M and Gutierrez R A *Plant Physiol* **152** (2010) 500
17. Rhee S Y, Beavis W, Bernadine T Z, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller L A, Mundodi S, Reiser L, Tacklind J, Weems D C, Wu Y, Xu I, Yoo D, Yoon J and Zhang P *Nucleic Acids Res* **31** (2003) 224
18. Altschul S F, Gish W, Miller W, Myers E W and Lipman D J *J Mol Biol* **215** (1990) 403

19. Felsenstein J *Science* **246** (1989) 941
20. Rice P, Longden I and Bleasby A *Trends Genet* **16** (2000) 276
21. Thompson J D, Higgins D G and Gibson T J *Nucleic Acids Res* **22** (1994) 4673
22. Lopes C T, Franz M, Kazi F, Donaldson S L, Morris Q and Bader G D *Bioinformatics* **26** (2010) 2347
23. Batagelj V and Mrvar A *Connections* **21** (1998) 47
24. Kleywegt G J and Jones T A *Acta Crystallogr D Biol Crystallogr* **50** (1994) 178
25. Canutescu A A, Shelenkov A A and Dunbrack R L J *Protein Sci* **12** (2003) 2001
26. Humphrey W, Dalke A and Schulten K *J Molec Graphics* **14** (1996) 33
27. DeLano W L, The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org" *DeLano Scientific* (2002)
28. Lange B M, Rujan T, Martin W and Croteau R *Proc Natl Acad Sci USA* **97** (2000) 13172
29. Ponting C P and Aravind L *Trends Biochem Sci* **24** (1999) 130
30. Alpy F and Tomasetto C *J Cell Sci* **118** (2005) 2791
31. Schick K, Nguyen D, Karlowski W M and Mayer K F *Genome Biol* **5** (2004) R41
32. Ariel F D, Manavella P A, Dezar CA and Chan R L *Trends Plant Sci* **12** (2007) 419
33. Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, Dubcovsky J, Fahima T, Sela H and Chen X *Science* **323** (2009) 1357
34. Romanowski M J, Soccio R E, Breslow J L and Burley S K *Proc Natl Acad Sci USA* **99** (2002) 6949
35. Loon L C V, Pierpoint W S, Boller T and Conejero V *Plant Molecular Biology Reporter* **12** (1994) 245
36. Peumans W J and Damme E J V *Plant Physiol* **109** (1995) 347
37. Shridhar S, Chattopadhyay D and Yadav G *Nucleic Acids Res* **37** (2009) W452.