

STIFDB2: An Updated Version of Plant Stress-Responsive Transcription Factor DataBase with Additional Stress Signals, Stress-Responsive Transcription Factor Binding Sites and Stress-Responsive Genes in Arabidopsis and Rice

Mahantesha Naika^{1,2}, Khader Shameer^{1,3}, Oommen K. Mathew¹, Ramanjini Gowda² and Ramanathan Sowdhamini^{1,*}

¹National Centre for Biological Sciences (TIFR), GKVK Campus, Bellary Road, Bangalore 560 065, India

²Department of Plant Biotechnology, University of Agricultural Sciences, GKVK Campus, Bellary Road, Bangalore 560 065, India

³Present address: Division of Biomedical Statistics and Informatics, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA

*Corresponding author: Email, mini@ncbs.res.in

(Received November 4, 2012; Accepted December 18, 2012)

Understanding the principles of abiotic and biotic stress responses, tolerance and adaptation remains important in plant physiology research to develop better varieties of crop plants. Better understanding of plant stress response mechanisms and application of knowledge derived from integrated experimental and bioinformatics approaches are gaining importance. Earlier, we showed that compiling a database of stress-responsive transcription factors and their corresponding target binding sites in the form of Hidden Markov models at promoter, untranslated and upstream regions of stress-up-regulated genes from expression analysis can help in elucidating various aspects of the stress response in Arabidopsis. In addition to the extensive content in the first version, STIFDB2 is now updated with 15 stress signals, 31 transcription factors and 5,984 stress-responsive genes from three species (*Arabidopsis thaliana*, *Oryza sativa* subsp. *japonica* and *Oryza sativa* subsp. *indica*). We have employed an integrated biocuration and genomic data mining approach to characterize the data set of transcription factors and consensus binding sites from literature mining and stress-responsive genes from the Gene Expression Omnibus. STIFDB2 currently has 38,798 associations of stress signals, stress-responsive genes and transcription factor binding sites predicted using the Stress-responsive Transcription Factor (STIF) algorithm, along with various functional annotation data. As a unique plant stress regulatory genomics data platform, STIFDB2 can be utilized for targeted as well as high-throughput experimental and computational studies to unravel principles of the stress regulome in dicots and gramineae. STIFDB2 is available from the URL: <http://caps.ncbs.res.in/stifdb2>

Keywords: Abiotic and biotic stress • Biocuration • Biological data mining • Stress response • Transcriptional regulatory cascade.

Abbreviations: GEO, Gene Expression Omnibus; GO, gene ontology; HMM, Hidden Markov model; STIF, Stress-responsive Transcription Factor; UTR, untranslated region.

Introduction

Among the world population, around 3.1 billion people from developing countries live in rural areas. For a large subset of this population (~2.5 billion people), agriculture is the primary source for their livelihood and it also contributes to economic growth as 30% of the gross domestic product (GDP) (FAO 2012). By the middle of the 21st century, the expected world population will be about 10 billion and we will witness serious food shortages (Smith et al. 2010). The increasing pressure on global food productivity due to climate change, combined with the drastic increase in population, results in a demand for crop varieties that are adaptive and resistant to a variety of stresses. Considering these various socio-economic and agro-economic factors, sustainable agricultural production is an urgent issue to meet these challenges (Takeda and Matsuoka 2008, Turner et al. 2009, Newton et al. 2011).

Plants are sessile, and they are often exposed to a wide range of both biotic and abiotic stresses, and they have developed intricate mechanisms to detect precise environmental changes, allowing optimal responses to adverse conditions (Atkinson and Urwin 2012). Biotic stress factors including bacteria,

Plant Cell Physiol. 54(2): e8(1–15) (2013) doi:10.1093/pcp/pcs185, available online at www.pcp.oxfordjournals.org

© The Author 2013. Published by Oxford University Press on behalf of Japanese Society of Plant Physiologists.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

fungi, viruses, nematodes and herbivorous insects, and abiotic stress factors such as drought (Dubouzet et al. 2003, Mahajan and Tuteja 2005, Chaves et al. 2009, Fleury et al. 2010, Harb and Perreira 2011), cold (Warren 1998), heat, (Kang et al. 2011) salinity (Ulm et al. 2002, Mahajan and Tuteja 2005, Wallia et al. 2005), dehydration (Urao et al. 1993, Shinozaki and Yamaguchi-Shinozaki 2000, Tran et al. 2004), UV-B (Kilian et al. 2007), wounding (Cheong et al. 2002) and heavy metals (Jonak et al. 2004) cause 30–60% yield losses every year globally (Mantri et al. 2010). Stress response and tolerance towards these stresses were propagated by complex biological pathways and regulatory events involving multiple molecular components (Shinozaki and Yamaguchi-Shinozaki 2000, Zhu 2001, Xiong et al. 2002, Abe et al. 2003, Jayasekaran et al. 2006, Liu et al. 2007, Agrarwal and Jha 2010, Baena-Gonzalez 2010, Tran and Mochida 2010, Walley and Dehesh 2010, Sinha et al. 2011). The effects of abiotic or biotic stress may occur singularly or in combination and induce cellular damage at multiple stages of plant growth and development, and may induce varying degrees of phenotypic lethality including reduction in growth, wilting, loss of leaves, etc. (Chinnusamy et al. 2004). However, this is achieved by activating distinct signal transduction cascades, which in turn activate stress-responsive genes, ultimately leading to survival by transcriptional reprogramming at multiple stages of plant growth and development (Xiong and Zhu 2001). Understanding the basic biological mechanisms by which plants respond to multiple stresses is a prerequisite for the development of stress tolerance in crop plants.

A wide variety of protein-coding genes, non-coding genes including microRNAs (Liu et al. 2008), transcription factors, epigenetic mechanisms and various biological pathways have been attributed to stress response (Singh et al. 2002, Hirayama and Shinozaki 2007, Shinozaki and Yamaguchi-Shinozaki 2007, Kilian et al. 2012), tolerance (Sreenivasulu et al. 2007, Roy et al. 2011) and adaptation (Mirouze and Paszkowski 2011). Some genes have been reported to respond to both biotic and abiotic stress signals (Mantri et al. 2010). Transcription factors are the master regulatory elements that directly bind to their distinct *cis*-regulatory elements and activate expression of many downstream genes, resulting in various mechanisms including stress tolerance (Agrarwal and Jha 2010). Various transcription factors such as AREB/ABF, MYB, AP2/EREBP, bZIP, MYC, HSF, DREB1/CBF, NAC, HB and WRKY were shown to influence stress response in plants (Singh et al. 2002, Shameer et al. 2009). Traditional methods to characterize stress-responsive transcription factors include nitrocellulose binding assays, DNA footprinting methods, gel-shift analysis, Southwestern blotting of both DNA and protein and high-throughput techniques such as chromatin immunoprecipitation chip (ChIP-chip) or ChIP-sequence, etc. While experimental methods are highly accurate, identification and characterization of the role of a given gene in a given stress response event will often be laborious and time consuming (Bulyk 2003, Zhou et al. 2010, Krasensky and Jonak 2012). To overcome this, computational

approaches offer a platform to gather information by integrating various public data sets and sensitive prediction algorithms. Concurrent mining of public databases and analysis using robust algorithms would provide a novel platform to understand the major molecular activities involved in stress response, adaptation and tolerance (Hu et al. 2003, Fernandez-Suarez and Birney 2008, Perez-Rodriguez et al. 2010, Guberman et al. 2011, Kinsella et al. 2011, Sucaet and Deva 2011, Spooner et al. 2012).

Genomic technologies are revolutionizing 21st century plant biology with the advent of high-throughput experimental platforms, optimized assay systems and advanced bioinformatics approaches. 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. Transcriptomic studies were employed for quantitative analysis of thousand of genes expressed at germination, growth and development, fertilization, flowering, and under biotic and abiotic stress conditions, thereby providing a convenient medium to characterize the stress regulome in plants (Oktem et al. 2008). An extensive collection of the expression data is currently available from public microarray databases such as Gene Expression Omnibus (GEO) (Barrett et al. 2001), ArrayExpress (Parkinson et al. 2007), Gene Expression Atlas (Kapusheky et al. 2010), etc. Mining such large-scale expression databases and integrating them with diverse data categories, using data interpretation algorithms, enables robust biological discoveries. The completion of the *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative 2000) and *Oryza sativa* L. genomes (Yu et al. 2002, Zhao et al. 2004, International Rice Genome Sequencing Project 2005) further enhances the application of sensitive sequence search algorithms to predict putative transcription factor binding sites at the whole-genome level and to understand stress regulation via multiple transcription factors in plants.

The stress-responsive mechanism in plants involves complex regulation of multiple genes and transcription factors. In the response to abiotic stresses such as ABA, drought, dehydration, cold, salinity, high light, heat, heavy metals, 10 specific families of transcription factors were known to be involved in *A. thaliana* and six specific families of transcription factors were known to be involved in *O. sativa* L. In plant biology, a detailed knowledge of the mechanisms of the transcriptional regulation of genes in response to biotic and abiotic stresses is an important paradigm. We hypothesize that identifying putative transcription factor binding sites for these stress-responsive transcription factors, upstream regions of genes differentially up-regulated in microarray studies due to multiple stress signals, could enable better understanding of genes and regulatory events mediating stress response mechanisms. Earlier, we described the Stress-responsive Transcription factor database, STIFDB (Shameer et al. 2009), a database that catalogued information of stress-responsive genes and transcription factor binding sites for abiotic stress-responsive genes in *A. thaliana*.

Multiple experimental studies that profiled stress-responsive analysis in plants (Kang et al. 2011, Sanghera et al. 2011, Babitha et al. 2012) and computational studies (Mishra et al. 2009, Georgii et al. 2012) on stress-specific gene regulation have utilized the data compiled in STIFDB. In this paper, we report a new version of the database, called 'STIFDB2', as a data platform for the investigation of the plant stress regulome. To enable the utility of such a database for a wider plant research community, we have now updated the database with new features such as: the addition of agriculturally important crop species, new stress signals, curated transcription factors and their binding sites, additional stress-responsive genes from microarray experiments and orthologs recorded from other important crop plants. We have also integrated predicted orthologs for other agriculturally important crop species, i.e. maize, sorghum and soybean.

Results

The data compiled using biocuration and genomic data mining, stress-responsive transcription factor prediction and data integration were compiled as a web-based database called STIFDB2.

STIFDB2: a database for analysis of stress regulome in plants

STIFDB2 provides information on stress-responsive genes from a dicot (*A. thaliana*) and two gramineae species (*O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica*). A total of 31 transcription factors (Table 1) were identified using the biocuration approach (see Table 1). A library of 15 stress signals that affect plants (ABA, aluminum, bacterial blight, cold, cold-drought-salt, dehydration, drought, heat, high light, iron, NaCl, osmotic stress, oxidative stress, UV-B and wounding) was also compiled (Table 2; details about transcription factors and consensus binding site data for *A. thaliana* were provided in Shameer et al. 2009). Transcription factor family/subfamily, *cis*-element, consensus binding site data and corresponding references (PubMed identifiers) used to perform transcription factor binding site prediction using stress-responsive genes in rice species are provided. Using the genomic data mining approach, a data set of 5,984 unique genes were identified and annotated with stress signals. 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.

Table 1 Summary of species, stress-responsive genes, stress-responsive transcription factors and stress signals in STIFDB2

Species	Stress-responsive genes in STIFDB2	Stress-responsive transcription factors	Stress signals
<i>Arabidopsis thaliana</i>	3,150	ABRE_ABI3_VP1 AuxRE_ARF C_ABRE_bZIP DREB_AP2_EREBP GCC_box_AP2_EREBP G_ABRE_bZIP G_box1_bZIP G_box2_bZIP G_box_bHLH HBE_HB HSE1_HSF Myb_box1_MYB Myb_box2_MYB Myb_box3_MYB Myb_box4_MYB Myb_box5_MYB N_box_bHLH Nac_box_NAC W_box_WRKY	ABA, aluminum, cold, cold-drought-salt, dehydration, drought, high light, iron, NaCl, osmotic stress, oxidative stress, UV-B and wounding
<i>Oryza sativa</i> subsp. <i>japonica</i>	1,118	ABRE_ABI3_VP1 ABRE_bZIP DREB_AP2_EREBP G_box1_bZIP Nac_box_NAC OsIRO2_bHLH PRE2_WRKY PRE4_WRKY	ABA, bacterial blight, cold, drought, heat, iron and NaCl
<i>Oryza sativa</i> subsp. <i>indica</i>	1,716	DREB_AP2_EREBP ABRE_bZIP G_box1_bZIP Nac_box_NAC	Cold, drought and NaCl

Transcription factors are named using the name of the *cis*-element followed by the transcription family name or subfamily name.

Table 2 Transcription factor family/subfamily, cis-element, consensus binding site data and corresponding references (PubMed identifiers) used to perform transcription factor binding site prediction using stress-responsive genes curated from the literature (PubMed identifiers of studies are provided) in *Oryza sativa* subsp. *japonica* and *Oryza sativa* subsp. *indica*

Transcription factor family name	Stress signal	Reference (Stress signal)	Name of the cis-element	Cis-element	Reference (Cis-element)
<i>Oryza sativa</i> subsp. <i>japonica</i>					
ABI3/VP1	Iron toxicity	19737364	ABRE	CATGC	19737364
AP2/EREBP	Cold, drought, NaCl	12609047 15834008 18470484	CRT/DRE	RCCGAC (ACCGAC GCCGAC)	12609047
bHLH	Iron toxicity	16887895	OsIRO2	CACGTGG	16887895
bZIP	ABA, NaCl, drought, heat	18236009 11828032	G-box1 ABRE	CCACGTGTC (C/T)ACGTGGC	18236009 11828032
NAC	Cold, drought, NaCl, ABA	20632034 19135985 16924117		CATGTG	17587305
WRKY	Bacterial blight	17986178	PRE2 PRE4	ACGCTGCCG TGCGCTT	17986178
<i>Oryza sativa</i> subsp. <i>Indica</i>					
AP2/EREBP	Drought, cold, salinity	12609047 18470484	CRT/DRE	(G/A)CCGAC	12609047
bZIP	ABA, NaCl, drought	10636868 19048288	G-box1 ABRE	CCACGTGTC (C/T)ACGTGGC	18236009 11828032
NAC	Cold, drought, NaCl	17587305		CATGTG	17587305

Table 3 Stress-specific distribution of *Arabidopsis thaliana* genes in STIFDB2 with predicted binding sites using the STIF algorithm

Transcription factors	ABA	Alu	Col	CDS	Dehy	Dro	Heat	Light	Iron	NaCl	OS	OxS	UV-B	Wounding
ABRE_ABI3_VP1	1	0	2	0	0	2	0	1	0	3	0	0	0	0
AuxRE_ARF	299	13	520	17	48	372	32	313	70	473	31	15	54	23
C_ABRE_bZIP	52	1	105	3	5	86	6	48	12	80	5	4	10	6
DREB_AP2_EREBP	230	8	375	23	29	344	12	198	34	362	13	14	29	17
GCC_box_AP2_EREBP	70	2	112	5	12	98	2	63	11	88	1	6	9	5
G_ABRE_bZIP	54	0	63	3	1	79	8	39	1	90	7	2	7	3
G_box1_bZIP	33	0	46	3	1	51	6	19	6	61	5	1	8	4
G_box2_bZIP	234	7	346	9	29	294	26	223	50	356	24	17	38	12
G_box_bHLH	241	4	361	20	17	369	33	201	33	405	31	10	40	20
HBE_HB	48	2	71	2	6	65	5	34	9	81	5	3	11	2
HSE1_HSF	563	19	892	30	73	728	51	558	106	876	47	34	94	35
Myb_box1_MYB	502	16	832	29	56	691	49	502	114	822	45	33	95	38
Myb_box2_MYB	247	10	396	15	28	315	13	224	51	381	15	14	47	24
Myb_box3_MYB	307	4	470	15	25	395	24	280	53	472	24	21	55	20
Myb_box4_MYB	142	5	242	4	19	172	8	150	24	248	7	9	29	10
Myb_box5_MYB	703	24	1,133	41	93	922	63	674	144	1,117	58	40	120	47
N_box_bHLH	163	6	228	13	13	224	12	164	25	261	10	10	26	15
Nac_box_NAC	225	6	352	14	25	330	30	227	50	382	24	12	42	16
W_box_WRKY	464	20	741	24	64	612	47	463	102	716	45	27	84	30

Alu, aluminum; Col, cold; CDS, cold-drought-salt; Dehy, dehydration; Dro, drought; OS, osmotic stress; OxS, oxidative stress.

Tables 3–5 provide a stress signal- and transcription factor-based summary of stress-responsive genes in STIFDB2. A summary of orthologs retrieved from sorghum, maize and soybean using stress-responsive genes in STIFDB2 is provided in **Table 6**.

The percentage of multiple stress-responsive transcription factors responsive to various stress signals in STIFDB2 is provided in **Fig. 1**. Chromosomal distributions of genes in STIFDB2 for three species are provided in **Fig. 2**.

Table 4 Stress-specific distribution of genes of *Oryza sativa* subsp. *japonica* genes in STIFDB2 with predicted binding sites using the STIF algorithm

Transcription factors	ABA	Bacterial blight	Cold	Drought	Heat	Iron	NaCl
ABRE_ABI3_VP1	30	424	11	87	531	157	33
ABRE_bZIP	3	63	0	14	71	15	6
DREB_AP2_EREBP	20	260	9	64	289	102	20
G_box1_bZIP	1	18	0	2	13	5	1
Nac_box_NAC	16	208	5	54	330	97	21
OslRO2_bHLH	6	69	2	17	82	21	3
PRE2_WRKY	0	1	2	1	7	2	0
PRE4_WRKY	2	26	1	2	30	11	0

Table 5 Stress-specific distribution of *Oryza sativa* subsp. *indica* genes in STIFDB2 with predicted binding sites using the STIF algorithm

Transcription factors	Cold	Drought	NaCl
ABRE_bZIP	26	98	109
DREB_AP2_EREBP	181	473	785
G_box1_bZIP	3	29	22
Nac_box_NAC	132	321	576

Table 6 Orthologs of stress-responsive genes compiled in STIFDB2 from Ensembl Plant BioMart using the Compara pipeline

Species	Stress-responsive genes in STIFDB2	Orthologs retrieved using BioMart		
		Sorghum (<i>Sorghum bicolor</i>)	Maize (<i>Zea mays</i>)	Soybean (<i>Glycine max</i>)
<i>Arabidopsis thaliana</i>	3,150	2,515	2,506	2,519
<i>Oryza sativa</i> subsp. <i>japonica</i>	1,118	909	838	727
<i>Oryza sativa</i> subsp. <i>indica</i>	1,716	1,155	1,055	926

Features of STIFDB2

Users can browse and search STIFDB2 using TAIR, TIGR or RAP database identifiers. Genes were also segregated into different chromosomes across three species and can be browsed on the basis of chromosomal location. The data can be browsed using one of the four key data sets: gene, transcription factors, stress signals and chromosome (Fig. 3). Users can also search the database using a variety of keywords including gene descriptions, stress signals, transcription factors and Gene Ontology (GO) annotations. A Basic Local Alignment Search Tool (BLAST) interface is provided to search STIFDB2 using nucleotide sequences. Sequence searches can be performed against the completed set of sequences in STIFDB2 or can be selectively searched against the genome of interest. STIFDB2 also provides

'TFMap' that enables users to track the transcription factor of interest visually using a 2D map that consists of the upstream region + untranslated region (UTR) and predicted transcription factors. The database also provides additional information including information about predicted binding sites, chromosomes, various database identifiers and cross-references to other plant databases.

Applications of STIFDB2

The data compiled in the previous version of STIFDB were utilized for a variety of experimental and computational studies related to abiotic stress response and transcription factor binding site predictions. Briefly, the database can be utilized for identifying putative transcription factor binding sites in the upstream regions/UTRs of stress-upregulated genes curated from gene expression studies. The data can be utilized to study various network features associated with genes up-regulated in the setting of various stresses. These data can also be extrapolated to identify protein–protein interactions amongst transcription factors (manuscript in preparation). STIFDB2 enables the identification of master transcription factors (transcription factors that bind in the up-regulated region of genes perturbed due to diverse stress signals). The database can also be used to study the functional role of highly up-regulated genes using function annotation data integrated in the database (manuscript in preparation). With the inclusion of orthologs, the database can be employed to study evolutionary conservation and cross-genome comparative analyses of abiotic stress-responsive genes across different plant species. Data compiled in STIFDB2 can also be used for the analysis of stress-responsive transcriptional regulatory networks in *Arabidopsis* or rice genomes.

Analysis of the transcriptional regulatory cascade of the abiotic stress response in *A. thaliana* using STIFDB2

To illustrate the application of STIFDB2, we performed an in-depth analysis of the transcriptional regulatory cascades of 19 stress-responsive transcription factors in *A. thaliana*.

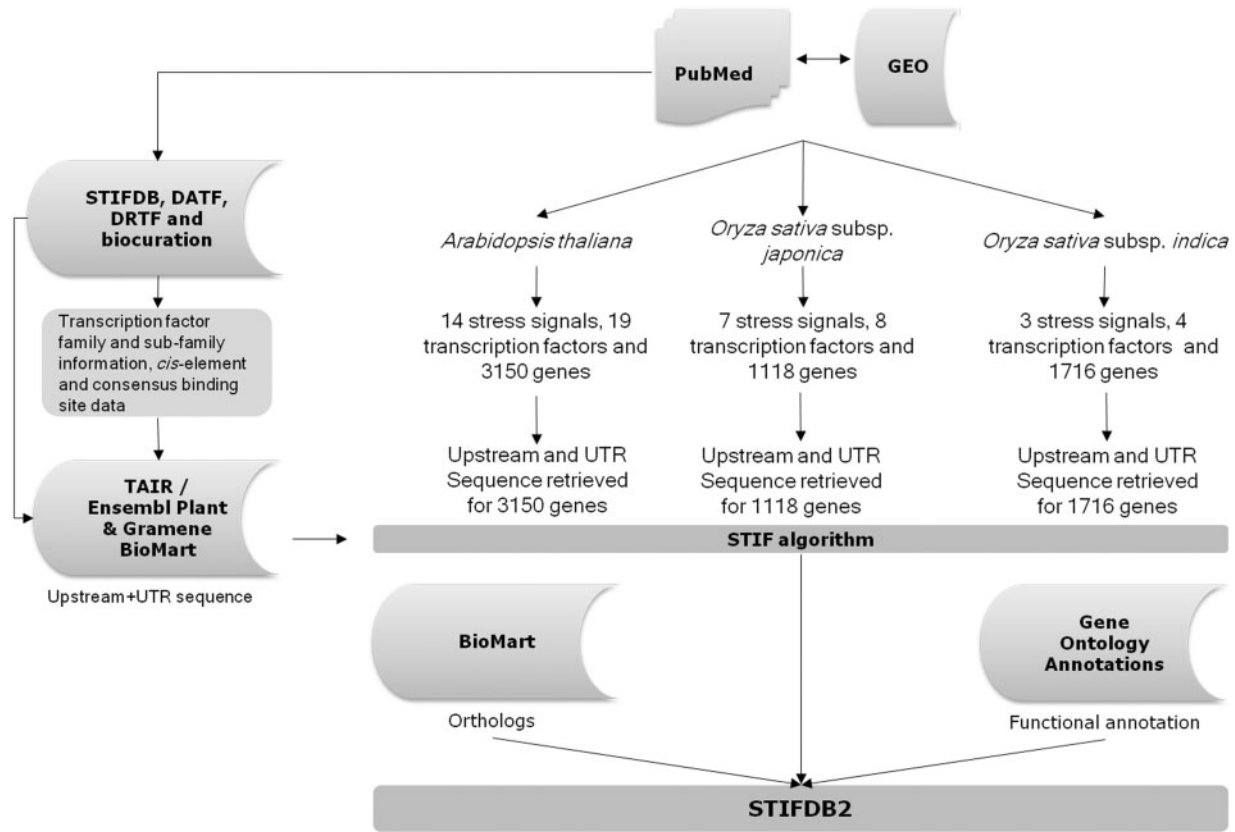


Fig. 1 Bioinformatics pipeline used to develop STIFDB2.

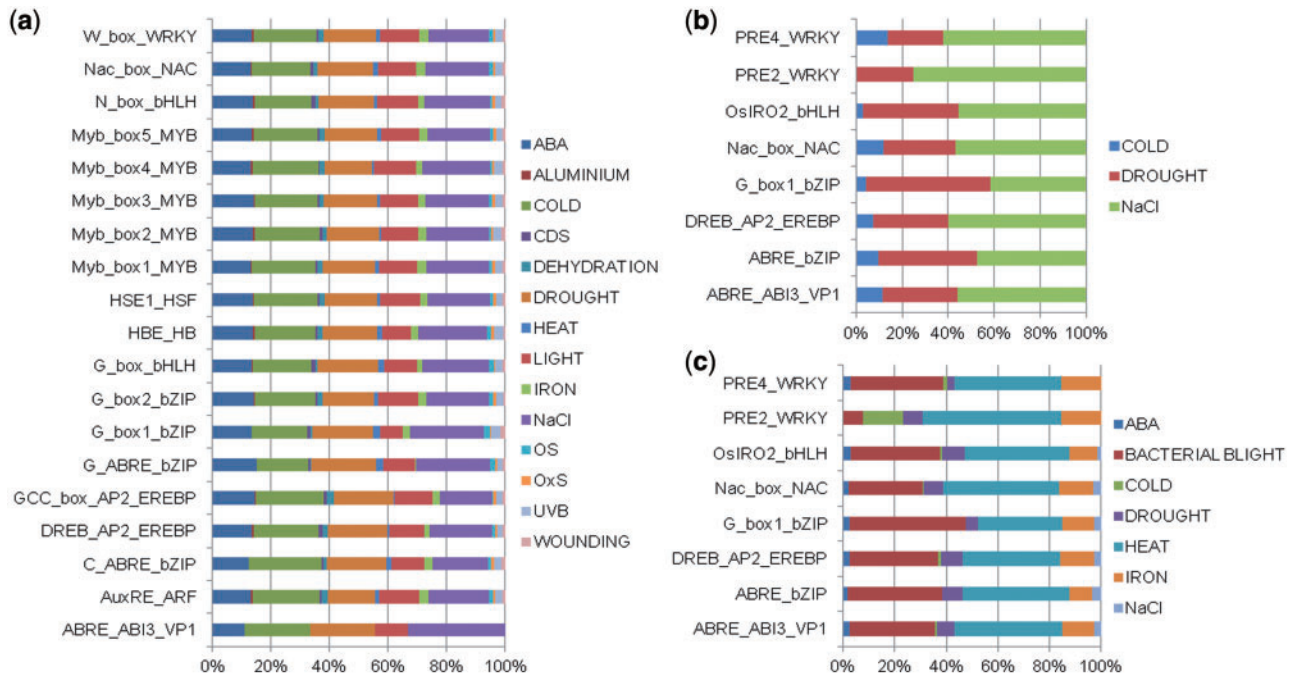


Fig. 2 Distribution of transcription factor binding sites predicted in the up-regulated stress-responsive genes due to the perturbation of various stress signals in (a) *Arabidopsis thaliana* (b) *Oryza sativa* subsp. *japonica* and (c) *Oryza sativa* subsp. *indica*.

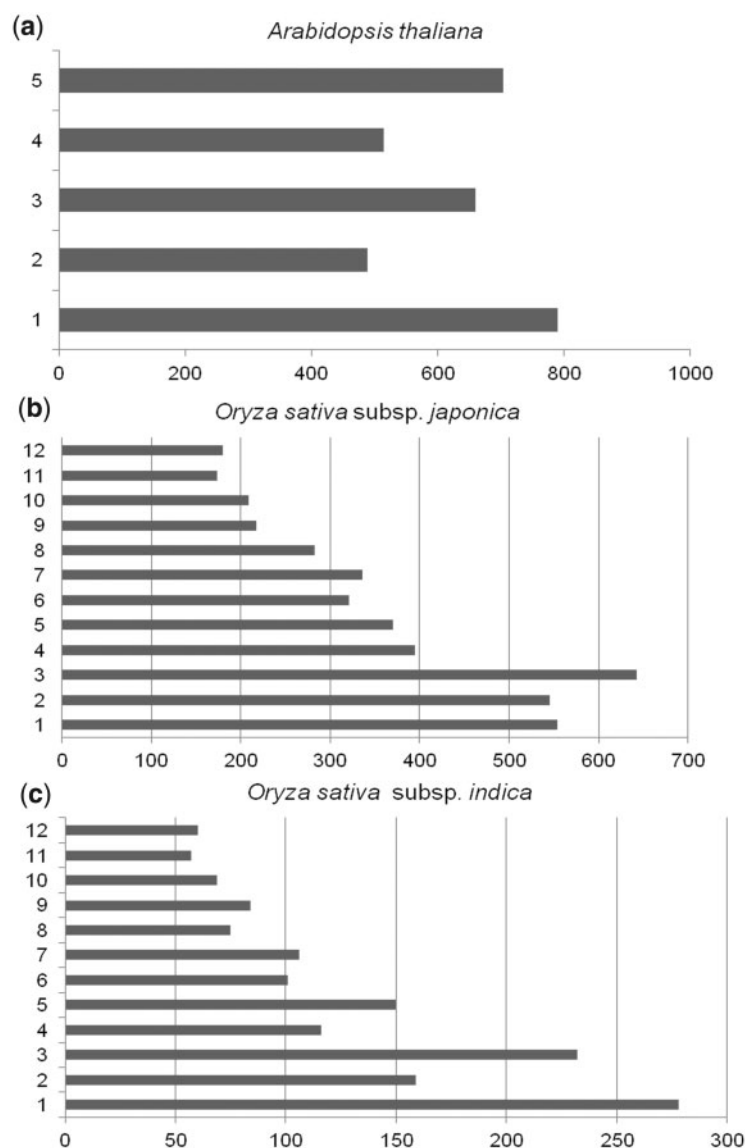


Fig. 3 Chromosomal distribution of stress-responsive genes in STIFDB2. The x-axis shows the number of stress-up-regulated genes curated from publicly available microarray data and the y-axis shows chromosome numbers.

Each transcription factor, inducing expression of another transcription factor due to a stress signal, may act as an activator or repressor of several other genes including those that encode transcription factors. A schematic diagram to depict primary differences between normal transcription factor activity and a transcriptional regulatory cascade event is provided in **Fig. 3**. The initial list of stress-responsive transcription factors in STIFDB was identified from biocuration followed by genomic data mining. These transcription factors may, in turn, activate other families of transcription factors via a transcriptional regulatory cascade mechanism. The transcriptional regulatory cascade mechanism was studied using the kinetics of regulatory cascades of regulatory networks including developmental gene networks (Davidson et al. 2002, Bolouri and Davidson 2003).

Transcriptional cascades were reported to play a mechanistic role in the abiotic stress response in *Arabidopsis* (Nover et al. 2001, Guo et al. 2008, Vandepoele et al. 2009). Several high-throughput experimental and computational studies (Chen and Zhu 2004, Shinozaki and Yamaguchi-Shinozaki 2007, Vandepoele et al. 2009, Moreno-Risueno et al. 2010, Cramer et al. 2011, Less et al. 2011, Krasensky and Jonak 2012, Walley and Dehesh 2012) were focused on the identification of transcriptional regulatory networks perturbed due to individual stress signals. A global survey of transcriptional regulatory cascades driven by 19 different stress-responsive transcription factors and 14 stresses were not reported elsewhere. Deconvoluting the role of a primary transcription factor in the regulation of one or more secondary transcription factors is challenging in *A. thaliana* due to various factors including

paucity of the data. STIFDB2 offers a solution to this problem by providing a large compendium of stress signals, stress-responsive transcription factors, putative target binding sites predicted using the STIF algorithm and upstream regions of stress-responsive genes curated from GEO. We utilized the data from STIFDB2 to investigate the transcriptional regulatory cascade network underlying abiotic stress responses in *A. thaliana*.

To understand the transcriptional regulatory cascades of transcription factors, associated with various stress conditions, we have used an analytical pipeline that performs text mining of annotation data (such as GO terms, protein domain information and Pfam2GO annotations). First, we grouped the STIF prediction results by transcription factors and retrieved the annotation data [gene description from TAIR, GO terms (molecular function subset) and protein domains for genes annotated with various stress signals]. For each transcription factor, genes that were predicted with their corresponding target binding sites were retrieved from STIFDB2. We performed targeted text mining in the annotation for GO terms in the 'molecular function' category pertaining to transcription factors or transcription factor activity. We used the term 'sequence-specific DNA binding transcription factor activity' (GO:0003700) as a primary filter. Terms from the neighborhood of GO:0003700 consisting of 'DNA binding'- or 'transcription'-related terms (see Table 7) were also used for filtering. Once the genes were identified, we further scanned them for protein domains annotated with transcription factors using Pfam2GO annotations and Pfam domain associations, leading to 90 PFAM domains and GO term association data from Pfam2GO (Supplementary Table S1).

A summary of transcription factors perturbed by stress-responsive transcription factors is provided in Table 8. We noted that ABRE_ABI3_VP1 has predicted binding sites in

Table 8 Summary of the transcription factors identified using the targeted annotation mining approach

<i>Arabidopsis</i> stress-responsive transcription factors in STIFDB2	Target genes in STIFDB	GO term hits ^a with transcription factor annotation	Pfam domain hits ^b with transcription factor annotation
ABRE_ABI3_VP1	5	1	1
AuxRE_ARF	1,405	117	40
C_ABRE_bZIP	248	27	8
DREB_AP2_EREBP	952	82	31
GCC_box_AP2_EREBP	299	12	4
G_ABRE_bZIP	173	16	4
G_box1_bZIP	111	6	2
G_box2_bZIP	1,020	60	18
G_box_bHLH	931	87	28
HBE_HB	188	23	8
HSE1_HSF	2,472	197	74
Myb_box1_MYB	2,318	176	68
Myb_box2_MYB	1,050	103	34
Myb_box3_MYB	1,269	108	38
Myb_box4_MYB	659	58	21
Myb_box5_MYB	3,113	250	92
N_box_bHLH	683	47	15
Nac_box_NAC	1,020	81	30
W_box_WRKY	2,077	167	64

^a GO terms used to filter genes associated with transcription factor/transcription factor activities are provided in Table 7.

^b Pfam domains used to define the functional role pertaining to transcription factor/transcription factor activity are filtered from Pfam2GO annotation (see Supplementary Table S1).

Table 7 GO terms used for target text mining of stress-responsive genes to find transcriptional regulatory cascades

GO identifier	GO term
GO:0003700	Sequence-specific DNA binding transcription factor activity
GO:0001073	DNA binding transcription antitermination factor activity
GO:0001199	Metal ion-regulated sequence-specific DNA binding transcription factor activity
GO:0001130	Sequence-specific DNA binding bacterial-type RNA polymerase transcription factor activity
GO:0001142	Sequence-specific DNA binding mitochondrial RNA polymerase transcription factor activity
GO:0001167	Sequence-specific DNA binding RNA polymerase I transcription factor activity
GO:0000981	Sequence-specific DNA binding RNA polymerase II transcription factor activity
GO:0001034	Sequence-specific DNA binding RNA polymerase III transcription factor activity
GO:0001011	Sequence-specific DNA binding RNA polymerase recruiting transcription factor activity
GO:0001010	Sequence-specific DNA binding transcription factor recruiting transcription factor activity
GO:0000975	Regulatory region DNA binding
GO:0043565	Sequence-specific DNA binding
GO:0000975	Regulatory region DNA binding
GO:0043565	Sequence-specific DNA binding
GO:0043566	Structure-specific DNA binding

five different genes in Arabidopsis. One of these genes (AT3G16770; ethylene-responsive transcription factor RAP2-3; AtEBP) was annotated with GO terms (DNA binding molecular_function, sequence-specific DNA binding transcription factor activity) and encodes a Pfam domain Apetala 2 (PF00847; AP2 that belongs to a large family of transcription factors). In addition to ABRE_ABI3_VP1, consensus binding regions for six other *cis*-elements were also predicted in the upstream region + UTR of AtEBP, confirming the importance of this gene in the stress response. ABRE_ABI3_VP1 was primarily considered as a transcription factor responsive to ABA-related stress. Several independent studies have reported that ABA and other plant hormones could participate in the same signaling cascades (Beaudoin et al. 2000, Tuteja 2007). Functional analysis of AtEBP conferred its role in abiotic stress response pathways in *A. thaliana* (Buttner and Singh 1997, Ogawa et al. 2005). Several interesting stress-responsive transcription-based regulatory cascades were observed and may reveal more interesting biological connections that can be studied using functional genomics and regulatory network analyses.

In summary, our analysis indicates that varying number of transcription factors could be modulated by the main set of

19 stress-responsive factors. Extensive experimental validation will be required to elucidate the complex patterns of gene regulation via the regulatory network cascade. The data summarized for each stress signal [see [Supplementary File \(.xls\)](#)] and data for 19 different transcription factors (labeled using the transcription factor name and given in [Table 8](#)) could be used for designing such experiments including ChIP-seq experiments to understand stress regulation pathways including mechanisms of response, adaptation and tolerance.

Discussion

Plants are commercially important crops, and it is important to understand the basic mechanism of the natural stress response and improve upon their stress response for better productivity. It will also be crucial to know the transcription factors that control and help with combating stress using stress tolerance pathways and adaptive mechanisms. STIFDB2 is a large compendium of curated stress signals, stress-responsive genes and stress-responsive transcription factors, along with information on putative binding sites where the stress-responsive transcription factors are predicted to be bound to the upstream regions

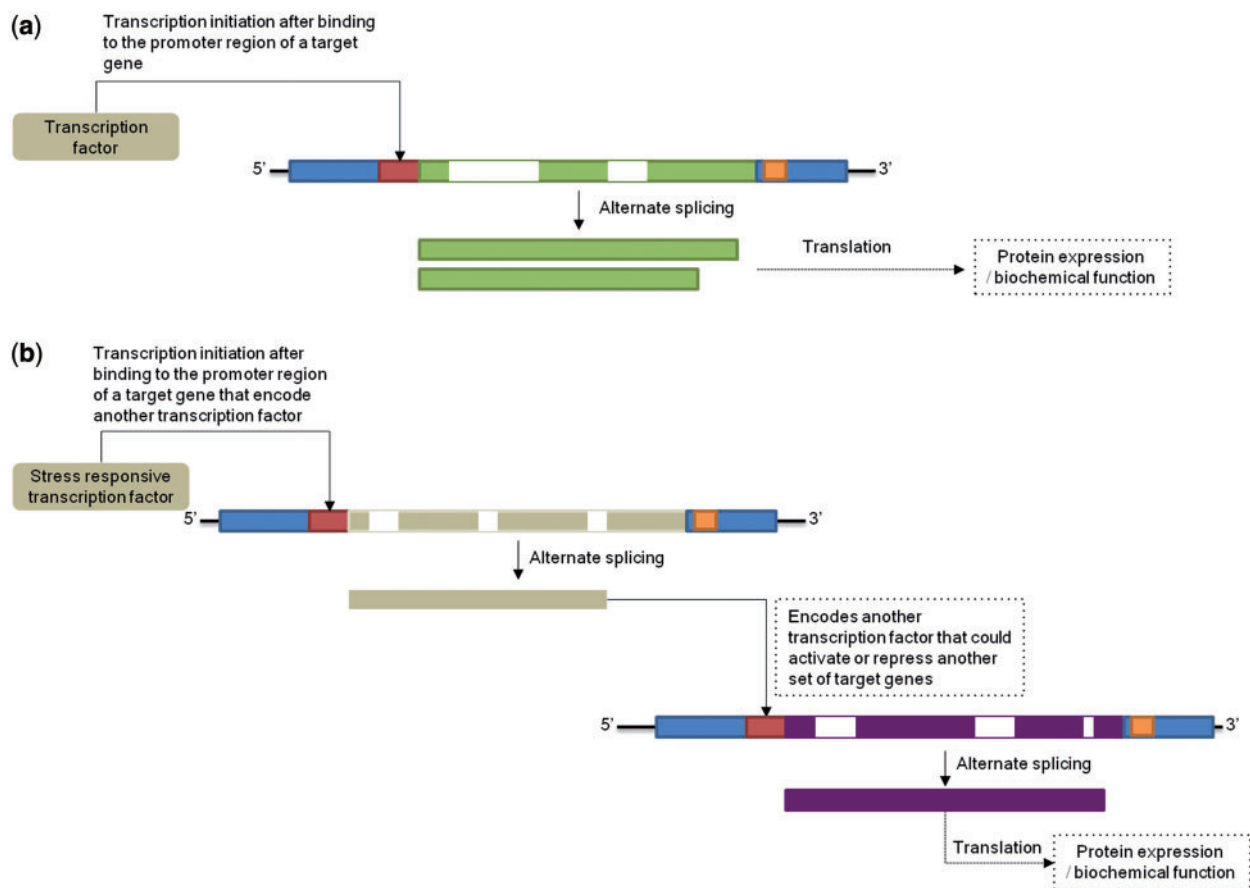


Fig. 4 Schematic diagram of (a) normal transcription factor activity and (b) transcriptional regulatory cascade network. Blue rectangles indicate untranslated regions, red rectangles indicate promoter regions, exon regions are highlighted using green, creme and violet color. Introns are defined using a white region interspersed between exons, and polyadenylation sites in the 3' untranslated regions are colored in orange.

and UTRs of stress-responsive genes. STIFDB2 will be a data-centric platform for performing analyses pertaining to stress response, tolerance and adaptation.

The data compiled in STIFDB can be categorized into three classes: (i) data extracted from biocuration; (ii) prediction results; and (iii) annotation data compiled from primary databases. Here, the first class of data refers to a list of Arabidopsis and rice stresses and stress-associated transcription factors curated from the literature. These data points were used as a query to find stress-responsive gene expression studies in Arabidopsis and rice transcriptomes. For each study, we curated corresponding up-regulated gene lists from the literature. After defining the stress-centric gene list, primary databases were mined to retrieve sequences, upstream regions, UTRs and annotations. Thus, STIFDB2 has multiple layers of biological data types integrated for stress gene analyses. The limitation of such knowledge-based approaches is the availability and reliability of validated binding sites and annotation of all genomes, including plant genomes. In this specific case, the underlying algorithm 'STIF' uses a Hidden Markov model (HMM). However, the models were generated using published consensus sites curated from the literature. The predictive approach in STIFDB2 is limited to pattern searching, and each binding site is provided with a z-score to indicate the strength of prediction. This enables the users to filter out different sites using the z-score threshold. The data, compiled in STIFDB2, can be used to answer biologically relevant questions on the stress regulome. For example, the data can be examined to identify

the functional repertoire and molecular pathways associated with stress-responsive genes and various stress signals in *A. thaliana* (manuscript submitted) and rice (manuscript in preparation). The data can also be utilized to find whether the stress-responsive genes are GC-rich or GC-poor classes of plant genes (Carels and Bernardi 2000). Such analyses can also be performed on the type of stress to understand the role of GC content and stress responses. Transcriptomic diversity of stress-responsive genes has been reported (Carels and Bernardi 2000, Duque 2011, Mastrangelo et al 2012, Syed et al. 2012), and detailed analysis of the transcriptomic plasticity, due to alternative splicing (Reddy et al. 2012) of genes up-regulated due to various stress signals, can be performed using the data integrated in STIFDB2. Recent studies have indicated that RNA-binding proteins may play a crucial role in the stress response in plants (Lorkovic 2009, Duque 2011, Nakaminami et al. 2012). The annotation data integrated in the database can be utilized to identify genes involved in RNA-binding proteins. Another interesting avenue where data in STIFDB2 can be utilized is in the analysis of transcriptional regulatory cascades of abiotic and biotic stress responses in plants. We performed an extensive analysis with the curated transcription factor information, a rich set of annotation data in STIFDB2, and the results were discussed in the application section. In a similar manner, the database can be used for both experimental and computational studies pertaining to plant stress response, tolerance and adaptation.

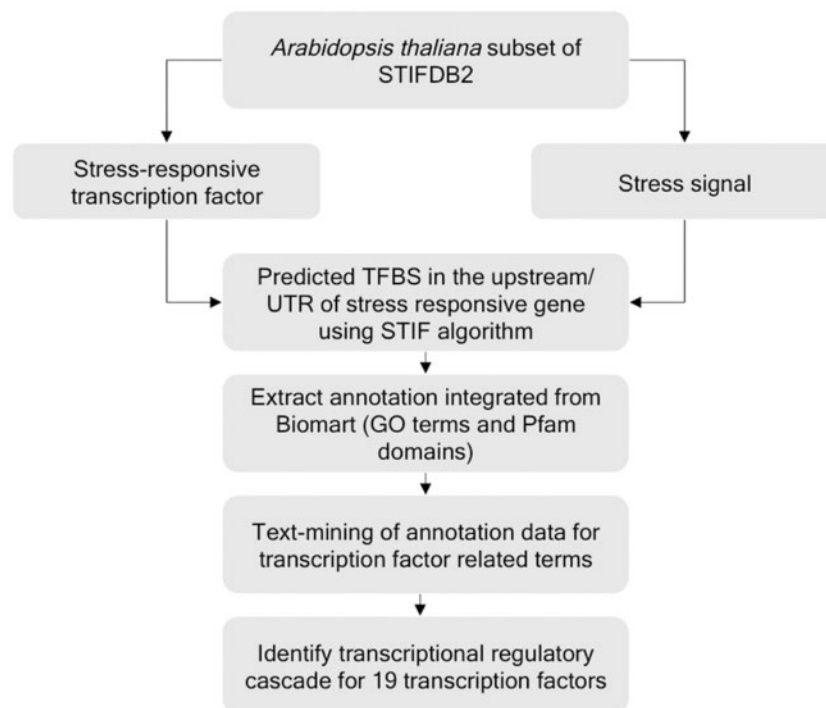


Fig. 5 Schematic diagram of the targeted text mining approach used to find transcriptional regulatory cascades mediated by stress-responsive genes from *Arabidopsis thaliana* in STIFDB2.

Materials and Methods

The content of STIFDB2 was generated using a bioinformatics pipeline consisting of three modules as follows: biocuration and genomic data mining; prediction of stress-responsive transcription factor binding sites; and data integration for developing a database and web-based platform for the analysis of stress-responsive genes and the stress regulome in plants. The analytic pipeline used to develop STIFDB2 is provided below (Fig. 4).

Biocuration and genomic data mining of stress-responsive transcription factors and stress-responsive genes

Information pertaining to stress-responsive transcription factors and transcription factor binding sites of consensus

nucleotide regions was curated from the literature, as explained in the previous version of STIFDB (Sundar et al. 2008, Shameer et al. 2009). The initial list of stress-responsive transcription factors was identified, and the consensus sequences were retrieved and used to generate HMMs for prediction of transcription factor binding sites using the STIF algorithm. The compendium of stress-responsive genes was mined from GEO by consulting the corresponding genes from the literature retrieved from PubMed. A gene was considered as stress responsive in STIFDB2 when it was reported to be differentially up-regulated in a perturbation experiment using one of the stress signals in the stress signal library. The data mining approach was performed to filter stress-responsive genes from three different plant species as follows: *A. thaliana*, *O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica* from public expression data sets. The transcription factors identified

(a) STIFDB V2.0 Stress Responsive Transcription Factor Database

STIFDB V2.0 (Stress Responsive Transcription Factor Database) is a comprehensive collection of basic and abiotic stress responsive genes in *Arabidopsis thaliana* and *Oryza sativa* L., with options to identify precise Transcription Factor Binding Sites in their promoters. In the response to both stress like bacterial and abiotic stresses like ABA, drought, salt, salinity, temperature, UV-B, high light, heat, heavy metals etc., ten specific families of transcription factors in *Arabidopsis thaliana* and six in *Oryza sativa* L. are known to be involved. HMM-based models are used to identify binding sites of transcription factors belonging to these families. We have also summarised literature reports to stress-validate the Transcription Factor Binding Sites predicted by the method.

Various Stress Signals Analyzed: ABA, OSMOTIC STRESS, WOUNDING, HEAT, LIGHT, HEAVY METALS, UV-B, DEHYDRATION, COLD, DROUGHT, SALINITY, OXIDATIVE STRESS.

STIFDB2: Database Statistics as on October 2012

Genome	No. of Stress Resp. Genes
<i>Arabidopsis thaliana</i>	3130
<i>Oryza sativa subsp japonica</i>	1318
<i>Oryza sativa subsp indica</i>	1718

(b) STIFDB V2.0 Stress Responsive Transcription Factor Database

Browse by Gene

<i>Arabidopsis thaliana</i>				<i>Oryza sativa subsp japonica</i>				<i>Oryza sativa subsp indica</i>			
Chromosome	No. of Stress Related Genes (1000bp)	List of Genes	Link	Chromosome	No. of Stress Related Genes (1000bp)	List of Genes	Link	Chromosome	No. of Stress Related Genes (1000bp)	List of Genes	Link
1	787	Link	1	128	Link	1	281	Link	Link	Link	Link
2	487	Link	2	132	Link	2	181	Link	Link	Link	Link
3	658	Link	3	156	Link	3	214	Link	Link	Link	Link
4	514	Link	4	105	Link	4	161	Link	Link	Link	Link
5	794	Link	5	108	Link	5	158	Link	Link	Link	Link
				6	83	Link	6	130	Link	Link	Link
				7	89	Link	7	113	Link	Link	Link
				8	62	Link	8	108	Link	Link	Link
				9	90	Link	9	98	Link	Link	Link
				10	75	Link	10	96	Link	Link	Link
				11	62	Link	11	94	Link	Link	Link
				12	68	Link	12	62	Link	Link	Link

(c) STIFDB V2.0 Stress Responsive Transcription Factor Database

Browse by Transcription Factors

Representative Color	Transcription Factor Frequency in STIFDB	Search Link	
	HSE1_HSF	1278	Link
	HSE1_ZFP	154	Link
	H1_Nuc_ZFP	70	Link
	H1_Nuc_Nuc	42	Link

(d) STIFDB V2.0 Stress Responsive Transcription Factor Database

STIFDB - Entry Name: AT1G01070

STIFDB Provides Data in 2 ways: 1000bp Data 100bp Data

Gene Description: AT1G01070

Accession: Gene: Model: Gene Model: Description: Gene Type

AT1G01070 AT1G01070.1 includes MtN21 /EtnA-like transporter family protein LOCATED IN membrane EXPRESSED IN: 17 plant protein_loading structure EXPRESSED IN: 20 growth stages CDONAS01:000403: Protein of unknown function DUF, transmembrane protein (PF08262) BEST Arabidopsis thaliana protein match to: nodulin MtN21 (Arabidopsis thaliana transporter family protein (CDN AT1G01070.1) has 3212 Blast hits to 2139 proteins in 339 species: Arabid - 23 Bacteria - 1886 Metazoa - 4 Fung - 6 Plants - 1233 Viruses - 0 Other Eukaryotes - 159 Source: NCBI BLAST)

Gene Ontology Data from TAIR: AT1G01070

GO Category: GO Annotation Type: GO ID: GO Evidence Type: References

Cellular Component: membrane: GO:0016020: IEA: AnalysisReference: S0173968

Cellular Component: membrane: GO:0016020: IEA: AnalysisReference: S0173968

(e) STIFDB V2.0 Stress Responsive Transcription Factor Database

STIFDB - 1000bp Information | TAIR Gene ID: AT1G01070

STIFDB - 1000bp TFmap | AT1G01070

Details of 1000bp TFmap | TAIR Gene ID: AT1G01070

No.	Transcription Factor	Z ² -Score	Normalisation Score	Start Residue	End Residue	Chromosome Location	Orientation
1	HSE1_HSF	1.2407	2.289	5046	-179	-172	Upstream region REVERSE
2	H1b_Nuc_HTF	2.4602	0.990	5046	-938	-933	Upstream region REVERSE
3	H1b_Nuc_HTF	2.5322	2.257	5046	-938	-933	Upstream region REVERSE
4	H1_Nuc_SmH	3.8906	9.414	5046	-538	-533	Upstream region REVERSE

Details about the Gene: AT1G01070:

TAIR ID: AT1G01070

Alternate Names: NUC[putative]

Chromosome Position: 38752-40944 (chr1)

Transcription Factors: HSE1_HSF, H1b_Nuc_HTF, H1_Nuc_SmH

Experimental Data From Microarray Database: GEO: DR04028

References: --

Fig. 6 Features of STIFDB2. (a) Front-page of STIFDB2. (b) Browse by chromosome in STIFDB2. (c) Browse by transcription factor page. (d) An intermediate page showing annotations and the link to access the 1,000 or 100 bp prediction results. (e) STIFDB2 page for the gene coding for the nodulin MtN21-like transporter family protein (AT1G01070) in *Arabidopsis thaliana*.

from biocuration were named using a convention that consists of the name of the *cis*-element followed by the transcription family name or subfamily. For example: Myb_box1 consists of the name of the *cis*-element 'Myb_box1' and the name of the transcription factor family 'Myb'. Biotic and abiotic stress-responsive genes up-regulated (in at least two replicates) in microarray experiments with a ≥ 2.5 -fold expression change were considered as candidates for STIFDB2. Gene sequences for 100 and 1,000 bp with their 5' UTR were extracted using Ensembl Plant BioMart (Guberman et al. 2011, Kinsella et al. 2011) using the Arabidopsis Information Resource (TAIR Version 10) database for Arabidopsis (Lamesch et al. 2012), Rice Genome Annotation Project MSU/TIGR (Ouyang et al. 2007) for *O. sativa* subsp. *indica*, and Rice Annotation Project (RAP-DB) (Ohyanagi et al. 2006) for *O. sativa* subsp. *japonica*. Various annotation data (GO annotations, gene descriptions, etc.) and ortholog information were retrieved using Ensembl plant and Gramene BioMart.

Prediction of stress-responsive transcription factor binding sites

The STIF algorithm was used to identify putative transcription factor binding sites in the upstream regions and UTRs of stress-responsive genes retrieved from publicly available microarray data. Briefly, the STIF algorithm encodes the consensus binding site data as an HMM model (Supplementary Fig. S1) and performs a sensitive sequence search in both the forward and reverse direction. The program gives the following output: transcription factor, z-score, normalization score, start and end of the predicted binding site, chromosomal location (upstream region or UTR depends on the location of the predicted binding site) and orientation of the strand (forward or reverse). We predicted stress-responsive transcription factors at two levels. The first level surveyed 1,000 bp + UTRs and the second level surveyed 100 bp + UTRs for putative binding sites. The detailed background on the STIF algorithm is available elsewhere [see Sundar et al. (2008) and Shameer et al. (2009) for a detailed description of the STIF algorithm and scoring method].

Data integration

Data retrieved from biocuration (transcription factor, *cis*-elements, binding site information and library of stress signals), genomic data mining (genes differentially expressed due to plant stress) and STIF prediction (transcription factors in the upstream regions + UTRs of stress-up-regulated genes) were integrated in STIFDB2. STIFDB2 has a large number of unique associations of stress signals, stress-responsive genes and transcription factor binding sites predicted using the STIF algorithm and targeted for plant stress regulome studies. These data sets were compiled as a database using a new interface that enables searching, browsing and retrieval of various information via a user-friendly web interface. STIFDB was developed on a MySQL backend. The web interface of STIFDB was developed using HTML and JavaScript. Perl-CGI programs were used for the development of search, query and retrieval system. Programs

for searching putative binding sites and for performing STIF prediction were coded in Perl. Stress profiles have been created for each gene that indicates the associated stress signals. We have also integrated GO associations, gene descriptions from TAIR (Lamesch et al. 2012), Rice Annotation Project (RAP) database (Ouynag et al. 2007), and transcription factor-related information from Database of Arabidopsis Transcription Factors (DATF) (Guo et al. 2005) and Database of Rice Transcription Factor (DRTF) (Gao et al. 2006).

Conclusion

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. Such information will also help to understand the metabolic, physiological and cellular mechanisms implicated in such processes. STIFDB2, as an expanded resource that includes two additional genomes (*O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica*), seven new stress signals (aluminum, bacterial blight, heat, iron, osmotic stress, UV-B and wounding) and 3,355 genes, will be an ideal information resource for plant biologists and computational biologists to perform in-depth analyses. The database provides 40,217 data points, and the data can be used to find putative RNA-binding proteins, transcriptional regulatory cascades, transcriptomic diversity and GC content of stress-responsive genes. We envisage that, analogously to its previous version, STIFDB2 will be widely accepted by the community and aid in unraveling novel aspects of stress response in plants.

Supplementary data

Supplementary data are available at PCP online.

Funding

The authors thank University of Agricultural Sciences (Bangalore) and National Centre for Biological Sciences, Tata Institute of Fundamental Research for infrastructural support. MN acknowledges University Grant Commission, New Delhi, India, for the Senior Research Fellowship during the course of this research work.

References

- Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 15: 63–78.
- Agarwal, P.K and Jha, B. (2010) Transcription factors in plants and ABA dependent and independent abiotic stress signaling. *Biol. Plant.* 54: 201–212.

- Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Atkinson, N.J. and Urwin, P.E. (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot.* 63: 3523–3543.
- Babitha, K.C., Ramu, S.V., Pruthvi, V., Mahesh, P., Nataraja, K.N. and Udayakumar, M. (2012) Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in *Arabidopsis*. *Transgenic Res.* (in press).
- Baena-Gonzalez, E. (2010) Energy signaling in the regulation of gene expression during stress. *Mol. Plant* 3: 300–313.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F. et al. (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.* 39: D1005–D1010.
- Beaudoin, N., Serizet, C., Gosti, F. and Giraudat, J. (2000) Interactions between abscisic acid and ethylene signaling cascades. *Plant Cell* 12: 1103–1115.
- Bolouri, H. and Davidson, E.H. (2003) Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proc. Natl Acad. Sci. USA* 100: 9371–9376.
- Buttner, M. and Singh, K.B. (1997) *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. *Proc. Natl Acad. Sci. USA* 94: 5961–5966.
- Bulyk, M.L. (2003) Computational prediction of transcription-factor binding site locations. *Genome Biol.* 5: 201.
- Carels, N. and Bernardi, G. (2000) Two classes of genes in plants. *Genetics* 154: 1819–1825.
- Chaves, M.M., Flexas, J. and Pinheiro, C. (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Ann. Bot.* 103: 551–560.
- Chen, W.J. and Zhu, T. (2004) Networks of transcription factors with roles in environmental stress response. *Trends Plant Sci.* 9: 591–596.
- Cheong, Y.H., Chang, H.S., Gupta, R., Wang, X., Zhu, T. and Luan, S. (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. *Plant Physiol.* 129: 661–677.
- Chinnusamy, V., Schumaker, K. and Zhu, J.K. (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *J. Exp. Bot.* 55: 225–236.
- Cramer, G.R., Urano, K., Delrot, S., Pezzotti, M. and Shinozaki, K. (2011) Effects of abiotic stress on plants: a systems biology perspective. *BMC Plant Biol.* 11: 163.
- Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caestani, C., Yuh, C.H. et al. (2002) A genomic regulatory network for development. *Science* 295: 1669–1678.
- Dubouzet, J.G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E.G., Miura, S. et al. (2003) OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33: 751–763.
- Duque, P. (2011) A role for SR proteins in plant stress responses. *Plant Signal. Behav.* 6: 49–54.
- FAO. (2012) Statistical Yearbook.
- Fernandez-Suarez, X.M. and Birney, E. (2008) Advanced genomic data mining. *PLoS Comput. Biol.* 4: e1000121.
- Fleury, D., Jefferies, S., Kuchel, H. and Langridge, P. (2010) Genetic and genomic tools to improve drought tolerance in wheat. *J. Exp. Bot.* 61: 3211–3222.
- Gao, G., Zhong, Y., Guo, A., Zhu, Q., Tang, W., Zheng, W. et al. (2006) DRTF: a database of rice transcription factors. *Bioinformatics* 22: 1286–1287.
- Georgii, E., Salojarvi, J., Brosche, M., Kangasjarvi, J. and Kaski, S. (2012) Targeted retrieval of gene expression measurements using regulatory models. *Bioinformatics* 28: 2349–2356.
- Guberman, J.M., Ai, J., Arnaiz, O., Baran, J., Blake, A., Baldock, R. et al. (2011) BioMart Central Portal: an open database network for the biological community. *Database (Oxford)* 2011: bar041.
- Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L. et al. (2005) DATF: a database of *Arabidopsis* transcription factors. *Bioinformatics* 21: 2568–2569.
- Guo, J., Wu, J., Ji, Q., Wang, C., Luo, L., Yuan, Y. et al. (2008) Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. *J. Genet. Genomics* 35: 105–118.
- Harb, A. and Pereira, A. (2011) Screening *Arabidopsis* genotypes for drought stress resistance. *Methods Mol. Biol.* 678: 191–198.
- Hirayama, T. and Shinozaki, K. (2010) Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J.* 61: 1041–1052.
- Hu, Y., Hines, L.M., Weng, H., Zuo, D., Rivera, M., Richardson, A. et al. (2003) Analysis of genomic and proteomic data using advanced literature mining. *J. Proteome Res.* 2: 405–412.
- International Rice Genome Sequencing Project. (2005) The map-based sequence of the rice genome. *Nature* 436: 793–800.
- Jayasekaran, K., Kim, K.N., Vivekanandan, M., Shin, J.S. and Ok, S.H. (2006) Novel calcium-binding GTPase (AtCBG) involved in ABA-mediated salt stress signaling in *Arabidopsis*. *Plant Cell Rep.* 25: 1255–1262.
- Jonak, C., Nakagami, H. and Hirt, H. (2004) Heavy metal stress. Activation of distinct mitogen-activated protein kinase pathways by copper and cadmium. *Plant Physiol.* 136: 3276–3283.
- Kang, H.G., Kim, J., Kim, B., Jeong, H., Choi, S.H., Kim, E.K. et al. (2011) Overexpression of FTL1/DDF1, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in *Arabidopsis thaliana*. *Plant Sci.* 180: 634–641.
- Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., Malone, J. et al. (2010) Gene expression atlas at the European bioinformatics institute. *Nucleic Acids Res.* 38: D690–698.
- Kilian, J., Peschke, F., Berendzen, K.W., Harter, K. and Wanke, D. (2012) Prerequisites, performance and profits of transcriptional profiling the abiotic stress response. *Biochim. Biophys. Acta* 1819: 166–175.
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O. et al. (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* 50: 347–363.
- Kinsella, R.J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G. et al. (2011) Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011: bar030.
- Krasensky, J. and Jonak, C. (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* 63: 1593–1608.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R. et al. (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40: D1202–D1210.
- Less, H., Angelovici, R., Tzin, V. and Galili, G. (2011) Coordinated gene networks regulating *Arabidopsis* plant metabolism in response to various stresses and nutritional cues. *Plant Cell* 23: 1264–1271.

- Liu, H.H., Tian, X., Li, Y.J., Wu, C.A. and Zheng, C.C. (2008) Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA* 14: 836–843.
- Liu, J.X., Srivastava, R., Che, P. and Howell, S.H. (2007) Salt stress responses in *Arabidopsis* utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J.* 51: 897–909.
- Lorkovic, Z.J. (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.* 14: 229–236.
- Mahajan, S. and Tuteja, N. (2005) Cold, salinity and drought stresses: an overview. *Arch. Biochem. Biophys.* 444: 139–158.
- Mantri, N.L., Ford, R., Coram, T.E. and Pang, E.C.K. (2010) Evidence of unique and shared responses to major biotic and abiotic stresses in chickpea. *Environ. Exp. Bot.* 69: 286–292.
- Mastrangelo, A.M., Marone, D., Laido, G., De Leonardi, A.M. and De Vita, P. (2012) Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* 185–186: 40–49.
- Mirouze, M. and Paszkowski, J. (2011) Epigenetic contribution to stress adaptation in plants. *Curr. Opin. Plant Biol.* 14: 267–274.
- Mishra, A.K., Agarwal, S., Jain, C.K. and Rani, V. (2009) High GC content: critical parameter for predicting stress regulated miRNAs in *Arabidopsis thaliana*. *Bioinformatics* 4: 151–154.
- Moreno-Risueno, M.A., Busch, W. and Benfey, P.N. (2010) Omics meet networks—using systems approaches to infer regulatory networks in plants. *Curr. Opin. Plant Biol.* 13: 126–131.
- Nakaminami, K., Matsui, A., Shinozaki, K. and Seki, M. (2012) RNA regulation in plant abiotic stress responses. *Biochim. Biophys. Acta* 1819: 149–153.
- Newton, A.C., Johnson, S. and Gregory, P.J. (2011) Implications of climate change for diseases, crop yields and food security. *Euphytica* 179: 3–18.
- Nover, L., Bharti, K., Doring, P., Mishra, S.K., Ganguli, A. and Scharf, K.D. (2001) *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress Chaperones* 6: 177–189.
- Ogawa, T., Pan, L., Kawai-Yamada, M., Yu, L.H., Yamamura, S., Koyama, T. et al. (2005) Functional analysis of *Arabidopsis* ethylene-responsive element binding protein conferring resistance to Bax and abiotic stress-induced plant cell death. *Plant Physiol.* 138: 1436–1445.
- Ohyang, H., Tanaka, T., Sakai, H., Shigemoto, Y., Yamaguchi, K., Habara, T. et al. (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.* 34: D741–D744.
- Oktem, H.A., Eyidogan, F., Selcuk, F., Oz, M.T., Teixeira da Silva, J.A. and Yucel, M. (2008) Revealing response of plants to biotic and abiotic stresses with microarray technology. *Genes Genomes Genomics* 2: 1–35.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K. et al. (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 35: D883–D887.
- Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farnie, A. et al. (2007) ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 35: D747–D750.
- Perez-Rodriguez, P., Riano-Pachon, D.M., Correa, L.G., Rensing, S.A., Kersten, B. and Mueller-Roeber, B. (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38: D822–D827.
- Reddy, A.S., Rogers, M.F., Richardson, D.N., Hamilton, M. and Ben-Hur, A. (2012) Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. *Front. Plant Sci.* 3: 18.
- Roy, S.J., Tucker, E.J. and Tester, M. (2011) Genetic analysis of abiotic stress tolerance in crops. *Curr. Opin. Plant Biol.* 14: 232–239.
- Sanghera, G.S., Wani, S.H., Hussain, W. and Singh, N.B. (2011) Engineering cold stress tolerance in crop plants. *Curr. Genomics* 12: 30–43.
- Shameer, K., Ambika, S., Varghese, S.M., Karaba, N., Udayakumar, M. and Sowdhamini, R. (2009) STIFDB—*Arabidopsis* Stress-responsive Transcription Factor DataBase. *Int. J. Plant Genomics* 2009: 583429.
- Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Curr. Opin. Plant Biol.* 3: 217–223.
- Shinozaki, K. and Yamaguchi-Shinozaki, K. (2007) Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* 58: 221–227.
- Singh, K., Foley, R.C. and Onate-Sanchez, L. (2002) Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol.* 5: 430–436.
- Sinha, A.K., Jaggi, M., Raghuram, B. and Tuteja, N. (2011) Mitogen-activated protein kinase signaling in plants under abiotic stress. *Plant Signal. Behav.* 6: 196–203.
- Smith, P., Gregory, P.J., van Vuuren, D., Obersteiner, M., Havlik, P., Rounsevell, M. et al. (2010) Competition for land. *Philos. Trans. R. Soc. B: Biol. Sci.* 365: 2941–2957.
- Spooner, W., Youens-Clark, K., Staines, D. and Ware, D. (2012) GrameneMart: the BioMart data portal for the Gramene project. *Database (Oxford)* 2012: bar056.
- Sreenivasulu, N., Sopory, S.K. and Kavi Kishor, P.B. (2007) Deciphering the regulatory mechanisms of abiotic stress tolerance in plants by genomic approaches. *Gene* 388: 1–13.
- Sucaet, Y. and Deva, T. (2011) Evolution and applications of plant pathway resources and databases. *Brief. Bioinform.* 12: 530–544.
- Sundar, A.S., Varghese, S.M., Shameer, K., Karaba, N., Udayakumar, M. and Sowdhamini, R. (2008) STIF: identification of stress-upregulated transcription factor binding sites in *Arabidopsis thaliana*. *Bioinformatics* 2: 431–437.
- Syed, N.H., Kalyna, M., Marquez, Y., Barta, A. and Brown, J.W. (2012) Alternative splicing in plants—coming of age. *Trends Plant Sci.* 17: 616–623.
- Takeda, S. and Matsuoka, M. (2008) Genetic approaches to crop improvement: responding to environmental and population changes. *Nat. Rev. Genet.* 9: 444–457.
- Tran, L.S. and Mochida, K. (2010) Identification and prediction of abiotic stress-responsive transcription factors involved in abiotic stress signaling in soybean. *Plant Signal. Behav.* 5: 255–257.
- Tran, L.S., Nakashima, K., Sakuma, Y., Simpson, S.D., Fujita, Y., Maruyama, K. et al. (2004) Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* 16: 2481–2498.
- Turner, W.R., Oppenheimer, M. and Wilcove, D.S. (2009) A force to fight global warming. *Nature* 462: 278–279.
- Tuteja, N. (2007) Abscisic acid and abiotic stress signaling. *Plant Signal. Behav.* 2: 135–138.

- Ulm, R., Ichimura, K., Mizoguchi, T., Peck, S.C., Zhu, T., Wang, X. et al. (2002) Distinct regulation of salinity and genotoxic stress responses by Arabidopsis MAP kinase phosphatase 1. *EMBO J.* 21: 6483–6493.
- Urao, T., Yamaguchi-Shinozaki, K., Urao, S. and Shinozaki, K. (1993) An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell* 5: 1529–1539.
- Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L. and Van de Peer, Y. (2009) Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol.* 150: 535–546.
- Walia, H., Wilson, C., Condamine, P., Liu, X., Ismail, A.M., Zeng, L. et al. (2005) Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage. *Plant Physiol.* 139: 822–835.
- Walley, J.W. and Dehesh, K. (2010) Molecular mechanisms regulating rapid stress signaling networks in Arabidopsis. *J. Integr. Plant Biol.* 52: 354–359.
- Warren, G.J. (1998) Cold stress: manipulating freezing tolerance in plants. *Curr. Biol.* 8: R514–516.
- Xiong, L., Schumaker, K.S. and Zhu, J.K. (2002) Cell signaling during cold, drought, and salt stress. *Plant Cell* 14(Suppl), S165–S183.
- Xiong, L. and Zhu, J.K. (2001) Abiotic stress signal transduction in plants: molecular and genetic perspectives. *Physiol. Plant.* 112: 152–166.
- Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B. et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296: 79–92.
- Zhao, W., Wang, J., He, X., Huang, X., Jiao, Y., Dai, M. et al. (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res.* 32: D377–D382.
- Zhou, X., Sumazin, P., Rajbhandari, P. and Califano, A. (2010) A systems biology approach to transcription factor binding site prediction. *PLoS One* 5: e9878.
- Zhu, J.K. (2001) Cell signaling under salt, water and cold stresses. *Curr. Opin. Plant Biol.* 4: 401–406.