Transcriptional regulatory networks in Arabidopsis thaliana during single and combined stresses

Pankaj Barah¹*, Mahantesha Naika B N², Naresh Doni Jayavelu³, Ramanathan Sowdhamini², Khader Shameer and Atle M. Bones¹,*

¹Cell, Molecular Biology and Genomics Group, Department of Biology, Norwegian University of Science and Technology, Trondheim N-7491, Norway, ²National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKV campus, Bangalore 560 065, India and ³Department of Chemical Engineering, Norwegian University of Science and Technology, Trondheim N-7491, Norway

ABSTRACT
Differentially evolved responses to various stress conditions in plants are controlled by complex regulatory circuits of transcriptional activators, and repressors, such as transcription factors (TFs). To understand the general and condition-specific activities of the TFs and their regulatory relationships with the target genes (TGs), we have used a homogeneous stress gene expression dataset generated on ten natural ecotypes of the model plant Arabidopsis thaliana, during five single and six combined stress conditions. Knowledge-based profiles of binding sites for 25 stress-responsive TF families (187 TFs) were generated and tested for their enrichment in the regulatory regions of the associated TGs. Condition-dependent regulatory sub-networks have shed light on the differential utilization of the underlying network topology, by stress-specific regulators and multifunctional regulators. The multifunctional regulators maintain the core stress response processes while the transient regulators confer the specificity to certain conditions. Clustering patterns of transcription factor binding sites (TFBS) have reflected the combinatorial nature of transcriptional regulation, and suggested the putative role of the homotypic clusters of TFBS towards maintaining transcriptional robustness against cis-regulatory mutations to facilitate the preservation of stress response processes. The Gene Ontology enrichment analysis of the TGs reflected sequential regulation of stress response mechanisms in plants.

INTRODUCTION
Plants are sessile organisms and continuously face different harsh environmental conditions as well as biotic invasions in their natural habitat (1). The activation of the plant stress responses is a complex, multidimensional process, acquired through large-scale transcriptional and metabolic reprogramming, which needs a sophisticated regulatory system (2-4). Gene expression variation responding to the surrounding environment confers phenotypic diversity within a species that can lead to consecutive adaptation and genome evolution (5-7). However, the underlying differentiations of transcriptional regulatory networks among different ecotypes of a given plant species remains largely unexplored (8,9). During the process of continuous evolution, the greater part of the plant stress regulatory networks is believed to have evolved through rigorous duplication of transcription factors (TFs) and their target genes (TGs) (10,11). Some core part of the regulatory interactions are conserved to variable degrees among genomes (12). A significant amount of research has been conducted (13) to understand plant responses to different single stress conditions by using various plant models (14,15). Plants have stress-specific adaptive responses as well as global responses, which protect the plants from more than one environmental stress. There are multiple stress perception and signalling pathways, some of which are stress-specific, but others may cross-talk at various steps of stress response processes (16). In a natural environment, plants are exposed to multiple stresses simultaneously, and interaction effect of multiple stresses at a time is more severe to plants (17-19).
Very few genome-scale experiments have been conducted till date, to study the molecular level responses of plants to a combination of different stresses in multiple ecotypes (17). All of these experimental studies have reported that response of plants to combination stresses were unique and could not be inferred directly from the response of plants to each of the individual stresses.

_Arabidopsis thaliana_ genome has ∼2000 TFs (20,21), however experimentally confirmed regulatory interactions were available for <100 TFs only (22,23). Significant efforts have been made to develop robust computational algorithms to extract regulatory information hidden in gene expression data to infer gene regulatory networks, to identify regulatory modules and condition-specific regulators (23–25). Such computational approaches to identify regulators of plant stress response, to infer gene regulatory networks and regulatory activities from experimental data has been reviewed in (26,27). Network-based algorithms are proven to be successful in dealing with underlying complexities of biological systems (28–30). From network perspective, the regulatory interactions between TFs and their TGs can be presented as a directed, bipartite graph depicting transcriptional activation or repression (31). In general, such regulatory networks exhibit a scale-free topology, by the presence of few highly connected or highly central regulatory hubs (28). The topology of such transcriptional stress regulatory networks in plants can be used for predicting global and stress-specific transcriptional regulators (11,32). Network Component Analysis (NCA) is an approach, which has been successfully employed in several species including _A. thaliana_, to determine both activities and regulatory influences for a set of TFs on their TGs in various perspectives (33).

Most of the existing computationally inferred regulatory network models in _A. thaliana_ were based on single stress treatment and from independent experiments (34–36). Integrative analysis of data from independent microarray experiments is challenging in most cases, mainly due to the lack of common standards regarding how to grow plants, conduct expression profile experiments, and finally, how to evaluate the resulting gene expression data (37). As a part of ERA-NET Plant Genomics MultiStress project, 10 different accesses of the model plant _A. thaliana_ were subjected to a set of five individual stresses (Cold, Heat, High-light, Salt and Flagellin—FLG) and six combinations of these stresses (Salt + Heat, Cold + high-light, Cold + FLG, Salt + high-light, Heat + FLG and Heat + high-light) under same experimental and growth conditions. Flagellin (FLG) mimicked biotic stress, and rest of the four single stresses were abiotic stresses. In the present study, a regulatory network in _A. thaliana_ has been constructed based on the assumption that high-dimensional mRNA expression profiles could be decomposed into low-dimensional regulatory signals driven through an interacting bipartite network between the regulating TFs and the regulated TGs (38). NCA was used to analyse this unique homogenous microarray dataset to predict the regulatory relationships between differentially stress-regulated TFs and their corresponding TGs during eleven stress conditions (five single and six combined stress conditions). Condition-dependent regulatory sub-networks were identified, and connectivity among the nodes (genes) were analysed for identifying highly connected hubs. Predicted interactions were compared with interactions derived from independent studies and manually retrieved database information. Hidden Markov Model (HMM) of known TF binding sites were identified and generated for 25 prominent stress-specific TF families from published experimental evidences (14,39). The predicted TF–TG regulatory interactions from gene expression data were tested using STIF algorithm (39) against these knowledge-based HMM profiles of known stress associated _cis_-regulatory elements (CREs). Clustering patterns of transcription factor binding sites (TFBS) on each of the TGs were analysed to identify clusters of TFBS (24). Gene set enrichment analysis (GSEA) was carried out to uncover functional association of the regulated TGs and overlapping interactions.

**MATERIALS AND METHODS**

**Microarray data**

The microarray data for this analysis were obtained from the ERA-NET Plant Genomics, ‘MultiStress’ project (http://www.ericaps.org/joint-calls/era-pg-funded-projects/2006-sub-call/multiple-stress-respondes-and-adaptations). The comprehensive description of the ‘MultiStress’ experiments (207 arrays and 59 unique experiments) has been presented in (17). The microarray data were deposited at the Gene Expression Omnibus (GEO) with the record GSE41935. As a part of the MultiStress project, ten natural ecotypes of _A. thaliana_ plants (Col-0, An-1, Cvi, Eri, Kas-1, Kon, Kyo-2, Ler, C24 and Sha) were subjected to the following stress treatments: Salt, Cold, Heat, High-light (HL), FLG, Salt + Heat, Cold + high-light, Cold + FLG, Salt + high-light, Heat + FLG and Heat + high-light. Combinations of high-light (800 μm photons m⁻² s⁻¹), cold (10°C), heat (38°C), high salinity (100 mM NaCl) and foliar spray application of the bacterial elicitor (20 μM flagellin peptide flg22) were set up in environmentally controlled rooms (17). _Arabidopsis_ NimbleGen 12-plex arrays were used for transcriptome profiling.

**Statistical analysis**

The published microarray dataset (GEO record GSE41935) was re-processed using the Robust Multi-array Average method implemented in the oligo package (40) in R programming language (41). For the current analysis, data from all of the 10 ecotypes were merged stress wise in an order to increase sample size and to minimize multiple testing errors (42). Stress-specific differentially regulated genes were identified by the Students _t_-test between control genotype and treatment effect (_P_ < 0.01) using the normalized expression values (43). All the normalized stress gene expression data were compared to their respective controls. The resulting differential expression values were used to generate eleven lists of stress-regulated genes by considering the top 500 significant entries from each of the stress conditions. The unified genelist comprised of 3429 genes were used for further network-based analysis (Supplementary Table S1).
**Finding Arabidopsis transcription factors (TFs)**

A list of 1926 Arabidopsis TFs was compiled from DATF: a database of Arabidopsis TFs (20), PlantTFDB: the plant TF database (44) and ‘The Arabidopsis Information Resource’ (TAIR) (45). Eleven stress-wise lists of TFs differentially expressed during the stress conditions were generated by comparing the individual list of eleven differentially stress-regulated gene lists from each of the stress conditions against the compiled list of 1926 TFs (Supplementary Table S2). The overlap of the differentially regulated TFs among eleven different stress conditions were determined and visualized using Sungear tool (46).

**TF binding motifs and sequence logos**

The list of differentially regulated gene-list contained 294 TFs, out of which 187 were grouped into 39 TF families based on manual searches in the DATF database (20). The TFBS motifs for these families were generated from STIFDB2 (47), AGRIS (48) and published literature (49,50). We considered only the experimentally confirmed TFBS motifs. The relevant references have been included with respective PubMed IDs in (Supplementary Table S3). The sequence logos for the consensus TFBS were prepared using WebLogo tool (51) and provided in (Supplementary Figure S1).

**Initial bipartite network reconstruction**

Top 500 stress-regulated genes were chosen from each of the 11 stress experiments and performed on 10 A. thaliana ecotypes were united to a list of 3429 differentially stress-regulated genes. Among them, 294 were TFs and 3135 were TGs. A bipartite gene network was constructed using Pearson correlation coefficient as a measure (PCC threshold ≥ 0.70). This network contained 10,534 bipartite relations. A bipartite gene network was constructed using Pear-son correlation coefficient as a measure (PCC threshold ≥ 0.70). This network contained 10,534 bipartite relations between 240 TFs and 1638 TGs. The constructed network is bipartite because they consist of two layers of nodes: TFs and TGs, with the links directed from the TFs to the TGs. This network was later fed into the ‘NCA’ algorithm to calculate differential activity profiles of the stress-regulated TFs.

**Network component analysis (NCA)**

NCA is a computational method for reconstructing the hidden regulatory signals or the transcription factor activities (TFAs) from gene expression data with known connectivity information regarding matrix decomposition (33).

NCA decomposition can be represented as:

\[
[E] = [C][T]
\]

(1)

Where, \([E]\) is an expression matrix, \([C]\) represents connectivity matrix and \([T]\) corresponds to TFA matrix. Based on the formula (1), the decomposition of \([E]\) into \([C]\) and \([T]\) can be achieved by minimizing the following objective function:

\[
\text{min} \parallel [E] - [C][T] \parallel^2
\]

(2)

s.t. \(C \in Z_0\)

In order to assure uniqueness of the solution of the Equation (2) up to a scaling factor, NCA criteria must be satisfied which includes:

(i) The connectivity matrix \([C]\) must have full-column rank.

(ii) When a node in the regulatory layer is removed along with all of the output nodes connected to it, the resulting network must be characterized by a connectivity matrix that still has full column rank.

(iii) \(T\) matrix must have full row rank.

The algorithm for NCA analysis is implemented in MATLAB by Liao et al. (33), and available online for download (http://www.seas.ucla.edu/~liaoj/downloads.html). With NCA as reconstruction method, we predicted significant TFs and their connectivity strength on TGs and TFAs.

**Simulated random network as control**

The topological properties of the inferred regulatory network was compared with a random network created using Erdős-Rényi (ER) model (52). We used the \(G(n, M)\) formulation of ER model, where \(G\) was the random network, \(n\) was the total number of nodes and \(M\) was the number of connection among the nodes. In the simulated random network \(n\) and \(M\) were kept constant as in our NCA inferred regulatory network. The degree of a node \((k)\) was calculated based on the total number of nearest neighbors. Random network generation and topological calculations were performed using igraph package in R (53).

**Cross-validation of the predicted interactions**

To evaluate the predicted TF–TG interactions, we have compiled a set of TFs, which had reported regulatory relationships in the protein interactome database (54) and the A. thaliana TF–TG interaction database (AtTFDB) (48). We have performed manual searches for predicted interactions in the AthaMap (55) and AraNet (56) databases. AthaMap provides a genome-wide map of potential TF and small RNA binding sites in A. thaliana. Data in AthaMap is based on published TF binding specificities, available as alignment matrices or experimentally determined single binding sites. AraNet is a probabilistic functional gene network of A. thaliana, constructed by the integration of 24 types of ‘omics’ data from multiple organisms, with each data type weighted according to how well it links genes that are known to function together in A. thaliana. As an independent comparison, we have also used the gene regulatory TF–TG network, reported in Vermeirssen et al., to compare with our predicted network (57).

**Identification and cross-validation of transcription factor binding sites using HMM-based profile search using STIF algorithm**

TFBSs of consensus nucleotide regions were taken from earlier studies (14,39,47). In addition, we have collected data regarding additional TFBSs, as described in a section...
above. Detailed methodology regarding searching for TFBS and generating HMM libraries are as described earlier (39). It has been reported in a recent study that overall, 89% of all upstream binding sites in A. thaliana were within 2 kb of the transcription start site (TSS), and highly significant matches of 73% were in 1-kb promoter (58). Therefore, TG sequences for 1-kb upstream from the TSS with their 5′ UTR were extracted by the data mining tool Biomart-Ensembl Plants (59) using TAIR version 10 (45). The STIF algorithm performed a sensitive sequence search that took the TFBS data as an HMM model and searched in both the forward and reverse direction of a target sequence to find potential binding sites. STIF program generated the output data with the start and end of the predicted binding site positions, chromosomal location, the orientation of the strand on forward or reverse, Z-score and normalization score. We have considered predicted TFBS at the Z-score threshold of 1.5 and above for further analysis.

Gene set enrichment analysis (GSEA)

The Biological Networks Gene Ontology (BiNGO) tool (60), an open-source Java based tool was used to determine significantly over-represented Gene Ontology (GO) terms associated with differentially regulated gene lists. The calculated P-values were Bonferroni corrected for minimizing multiple testing errors. Gene annotations were acquired from TAIR database (v10) (45) using biomaRt (59) data mining tool.

RESULTS

Our analytics approach followed the schematic presented in Figure 1. The NCA method requires two inputs: a set of gene expression profiles and an initial network of TFs and TGs. In this study, we utilized a unique A. thaliana gene expression dataset comprising of 207 arrays during eleven different stress conditions performed on ten ecotypes originated from different geographical locations. For maintaining the homogeneity, we have selected top 500 significantly stress-regulated genes from each stress condition and the final united list contained 3429 transcripts from all stress conditions (Supplementary Table S1).

Arabidopsis ecotypes have different transcriptome signatures while responding to eleven stress conditions

After the statistical analysis of the microarray data, we have observed significant variation among the 10 ecotypes while responding to 11 different stresses (here we consider the combination of two stresses as an independent stress). For example, cold treatment differentially regulated (P ≤ 0.01) 1950 genes in the southern ecotypes Cvi, while it regulated only 138 transcripts in the Col ecotype, known to be cold tolerant (61). Similar results were obtained for all other single and double stresses (summarized in Table 1). Intraspecific natural variation or within-species phenotypic variation, caused by spontaneously arising favourable mutations facilitate the evolutionary process, which contribute towards the local adaptation of the plant (62,63). Environmental stress is a key factor in determining the genome regulation, adaptive evolution and geographical distribution of any living organism including plants (64). A few speculations have been made elsewhere, regarding the contribution of cis-regulatory evolution in genetic variations during stress-responsive gene expression signatures that facilitate the adaptation to local climate (65–67). We wanted to explore the diverse activities of stress-regulated TFs and their connected TGs in A. thaliana during 11 stress conditions.

Stress-regulated transcription factors

The differentially stress-regulated gene lists contained many Arabidopsis TFs. Initially, we have created a list of total 1926 Arabidopsis TFs, as reported in TAIR (45), GO (68), DATF (20) and PhTFDB (69). By comparing the list of differentially regulated genes in our data and the compiled reference list of Arabidopsis TFs, we identified 294 Arabidopsis TFs to be present among the differentially stress-regulated transcripts in our dataset (see Supplementary Table S2). Most of the TFs were differentially regulated in single conditions while very few TFs were regulated in multiple stress conditions. It was primarily observed from manual biocuration that the multi functional stress-regulated TFs were mostly related to circadian pathways controlling various physiological and developmental processes coupled with environmental stimuli (70,71). For example, At5g57600 (CONSTANS-like 5), known to be a core regulator involved photoperiodic flowering pathway was regulated in seven stress conditions (Figure 2) (72). A recent study has reported that flowering time variation in A. thaliana was associated with cis-regulatory variation in CONSTANS (73). This work establishes a link between regulatory mutations and evolutionary changes in plants at the species level (74). Two other TFs, namely At5g24470 (PSEUDO-RESPONSE REGULATOR, PRR5) and At5g47640 (NUCLEAR FACTOR Y SUBUNIT B2, NF-YB2) (75) were regulated in six stress conditions. Out of them, PRR5 is known to form a negative feedback loop within the Arabidopsis circadian clock (76). Five TFs, namely At1g25440 (B-BOX DOMAIN PROTEIN 15, BBX15), At1g71030 (MYB-LIKE 2, MYBL2) (77), At3g15540 (INDOLE-3-ACETIC ACID INDUCIBLE 19, IAA19) (78), At5g25190 (ETHYLENE AND SALT INDUCIBLE 3, ESE3) and At5g44260 (TANDEM CCCCH ZINC FINGER PROTEIN 5, TZF5) were overlapped among five stress conditions. Eight TFs, At1g13260 (RELATED TO ABI3/VPI, RAV1) (79), At2g46790 (TOC1-LIKE PROTEIN 1, TL1) (80), At3g47500 (CYCLING DOF FACTOR 3, CDF3) (81), At3g54810 (GATA TRANSCRIPTION FACTOR 8, GATA8) (82), At3g59060 (PHYTOCHROME-INTERACTING FACTOR, PIL6) (83), At4g01250 (WRKY22) (84), At4g25480 (DEHYDRATION RESPONSE ELEMENT B1A, DREB1A) (85) and At5g02810 (PSEUDO-RESPONSE REGULATOR 7, PRR7) (86) were regulated in four stress conditions. Among other TFs, 19 were overlapped in three stress conditions and 58 were overlapped in two stress conditions. In total, 147 TFs were regulated only in single stress conditions (Figure 2 and Supplementary Table S2). Functional associations of some of these TFs in stress-related processes have been discussed in later sections.
Figure 1. Methodology flowchart. (A) Top 500 stress-regulated genes were chosen from each of the eleven stress experiments performed on 10 Arabidopsis thaliana natural ecotypes. The united lists of stress-regulated genes contained 3429 genes, out of which 294 were TFs. A gene co-expression network was constructed using Pearson correlation coefficient (PCC $\geq 0.70$), which contained 10,534 bipartite relations between 240 TFs and 1638 target genes (TGs). This network was later fed into the Network Component Analysis (NCA) algorithm, which then calculated the activity profiles of 194 stress-regulated TFs. Regulatory signal strength based filtering criteria ($r \geq 0.75$) was applied, which retained 5380 interactions between 182 TFs and 1199 TGs. Eleven regulatory sub-networks were extracted. Predicted connections were compared with independent interactions obtained from STIF analysis, AGRIS, AthaMap, AraNet and independent literature. (B) Flowchart of STIF algorithm, which was used for predicting transcription factor (TF) binding sites using HMM-based profile search. Total 187 stress-regulated TFs were classified into 39 TF Families and corresponding HMM profiles were generated. STIF algorithm predicted 10,474 interactions between 25 TF families and 1,021 TGs at the Z-score threshold of 1.5. (adapted and modified with permission from (39))
Table 1. Number of differentially regulated genes ($P$-value $\leq 0.01$) in 10 ecotypes during 11 different stress conditions

<table>
<thead>
<tr>
<th>Stress/Ecotypes→</th>
<th>Ler</th>
<th>Col</th>
<th>C24</th>
<th>Eri</th>
<th>Kyo-2</th>
<th>Cvi</th>
<th>Sha</th>
<th>An-1</th>
<th>Kond</th>
<th>Kas-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat</td>
<td>321</td>
<td>95</td>
<td>107</td>
<td>384</td>
<td>469</td>
<td>775</td>
<td>237</td>
<td>659</td>
<td>280</td>
<td>790</td>
</tr>
<tr>
<td>High-light</td>
<td>648</td>
<td>276</td>
<td>535</td>
<td>243</td>
<td>446</td>
<td>266</td>
<td>636</td>
<td>679</td>
<td>312</td>
<td>854</td>
</tr>
<tr>
<td>Salt</td>
<td>139</td>
<td>97</td>
<td>523</td>
<td>126</td>
<td>399</td>
<td>486</td>
<td>230</td>
<td>983</td>
<td>828</td>
<td>1396</td>
</tr>
<tr>
<td>Cold</td>
<td>470</td>
<td>138</td>
<td>1332</td>
<td>1025</td>
<td>702</td>
<td>1950</td>
<td>510</td>
<td>550</td>
<td>784</td>
<td>901</td>
</tr>
<tr>
<td>FLG</td>
<td>2863</td>
<td>586</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat + FLG + Silwet</td>
<td>558</td>
<td>399</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salt + Heat</td>
<td>311</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold + High-light</td>
<td>412</td>
<td>156</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salt + High-light</td>
<td>234</td>
<td>233</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold + FLG + Silwet</td>
<td>249</td>
<td>950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold + Silwet</td>
<td>186</td>
<td>181</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat + High-light</td>
<td>498</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold + FLG + Silwet</td>
<td>249</td>
<td>950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of differentially regulated genes ($P \leq 0.01$) in 10 ecotypes during eleven different stress exposures. The transcriptomic profiling was conducted at single time point. Combined stress experiments and flagellin (FLG) experiments were performed only on Col and Ler ecotypes, except for the ‘salt + heat’ combined stress experiments, which were carried out in seven ecotypes. Silwet detergent was used for all the FLG treatments. All stress treatments were compared with their respective controls, except the treatments including FLG (Heat + FLG and Cold + FLG), which were compared with control and silwet samples.

Figure 2. Overlapping statistics of stress-regulated TFs. Number of stress specific differentially regulated TFs are shown inside each circle along the periphery of the hendecagon. The small circles in the diagram represent overlapping TFs among different stress conditions. The arrow marks on the inner circles point towards the stress conditions in which they are significantly regulated. The size of the inner circles represents the number of TFs. The filled inner circle represents CONTSTANS-LIKE 5, which was significantly regulated in seven conditions. (Also see Supplementary Table S2).

Initial co-expression based network

To derive the TF activity profiles from gene expression data, an initial input network had to be constructed. Due to the limited availability of experimentally determined large-scale regulatory interactions in the model plant *A. thaliana*, we opted for an ab initio method to compile the initial bipartite network using our microarray data (17). In the list of differentially stress-regulated genes, we identified 294 TFs and 3135 TGs. We predicted the correlation between each pair of TFs and TGs expression levels and retained an interaction if the correlation coefficient was significant ($r > 0.7$) (36, 87–88). The network was represented in matrix form containing the regulatory information of each TF on their TGs. This analysis resulted in a network of 10 534 interactions with 240 TFs and 1638 TGs (Supplementary Table S4 and Figure S3A). There were 8450 positive interactions (activation) and 2084 negative interactions (repression).

Analysis of the connectivity patterns of the nodes in the transcriptional regulatory networks revealed that some TFs are involved in multiple sub-networks, suggesting a hierarchical structure whereby specific TFs might play major synergistic roles in the greater regulatory network, and function as multifunctional hubs. A subset of highly connected TFs in the global network are presented in Table 2. To obtain more information about these genes, we opted for biocuration method (89). Most of these highly connected genes mentioned above were also reported in published literature to be involved in important regulatory functions associated with representative GO categories, such as stress response, defence response, cellular development, regulation of secondary metabolism, etc. Evaluation of the regulatory sub-networks retained several of these regulatory hubs that control many aspects of cellular metabolism associated with plant stress responses. The biological relevance of some of these stress regulators has been described briefly in later sections (stress wise regulatory sub-networks).

Comparison of the inferred co-regulatory network and a random control network

Biological networks are known to be scale-free in nature, which means the very little number of nodes has higher degrees, and maximum nodes have a lower degree (90, 91). The inferred TF–TG network has exhibited a power law degree distribution, showing its scale-free nature (Figure 3B). In a random network, the degrees of the nodes follow a Poisson distribution (28), which indicated that most nodes had approximately the similar number of links, roughly equal to the network’s average degree (Figure 3C).

Transcription factor activities under eleven stress conditions

Considering the PCC as control strength for NCA, we constructed a matrix of regulatory relationships between the stress-regulated TFs and their corresponding TGs (92). Based on the assumption defined by NCA algorithm (33),
Figure 3. Topology of the predicted bipartite regulatory network visualized using Cytoscape software version 3.2.1 (169). (A) A topological overview of the predicted bipartite regulatory network between each pair of TF and TGs using PCC (r > 0.7). TFs are represented as triangles and TGs are represented as ellipses. This network is composed of 10,534 interactions (8,450 positive interactions and 2,084 negative interactions) between 240 TFs and 1,638 TGs. The network topology has been simplified for presentation purpose. Detailed TF–TG interactions are provided in Supplementary Table S4. (B) The degree distribution of the nodes in the predicted regulatory network showed a power-law. (C) The degree distribution of the control network showed a Poisson distribution. (D) The topology of the regulatory sub-network during FLG stress condition. The TFs are represented as triangles and TGs are represented as ellipses. The size of the triangle reflects the number of connecting TGS to the corresponding TF. Previously reported key biotic stress regulators, such as WRKY22, MYB117, GNL, HSFA1A, MYB124, etc. are prominently visible as highly connected hubs in this network topology.
Table 2. List of few highly connected stress regulators in the global TF–TG network

<table>
<thead>
<tr>
<th>At-IDs</th>
<th>Alias</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>At4g01250</td>
<td>WRKY22</td>
<td>WRKY FAMILY TRANSCRIPTION FACTOR 22 (84)</td>
</tr>
<tr>
<td>At3g24500</td>
<td>MBF1C</td>
<td>MULTIPROTEIN BRIDGING FACTOR 1C (153)</td>
</tr>
<tr>
<td>At5g66620</td>
<td>DAR6</td>
<td>DAI-RELATED PROTEIN 6</td>
</tr>
<tr>
<td>At1g72210</td>
<td>bHLH096</td>
<td>BASIC HELIX-LOOP-HELIX (BHLH) DNA-BINDING SUPERFAMILY PROTEIN 96</td>
</tr>
<tr>
<td>At3g12130</td>
<td></td>
<td>KI DOMAIN-CONTAINING PROTEIN/ ZINC FINGER (CCCH TYPE) FAMILY PROTEIN</td>
</tr>
<tr>
<td>At4g18170</td>
<td>WRKY28</td>
<td>WRKY DNA-BINDING PROTEIN 28 (154)</td>
</tr>
<tr>
<td>At1g7380</td>
<td>JAZ5</td>
<td>JASMONATE-ZIM-DOMAIN PROTEIN 5 (155)</td>
</tr>
<tr>
<td>At3g32640</td>
<td>ZBF1</td>
<td>BASIC HELIX-LOOP-HELIX (BHLH) DNA-BINDING FAMILY PROTEIN (156,157)</td>
</tr>
<tr>
<td>At1g26780</td>
<td>MYB117</td>
<td>MYB DOMAIN PROTEIN 117</td>
</tr>
<tr>
<td>At1g71030</td>
<td>MYBL2</td>
<td>MYB-LIKE 2 (77)</td>
</tr>
<tr>
<td>At1g52150</td>
<td>HB-15</td>
<td>HOMEBOX-LEUCINE ZIPPER FAMILY PROTEIN (159)</td>
</tr>
<tr>
<td>At4g17490</td>
<td>ERF6</td>
<td>ETHYLENE RESPONSIVE ELEMENT BINDING FACTOR 6 (160)</td>
</tr>
<tr>
<td>At4g26150</td>
<td>CGA1</td>
<td>CYTOKININ-RESPONSIVE GATA FACTOR 1 (161)</td>
</tr>
<tr>
<td>At4g23550</td>
<td>WRKY29</td>
<td>WRKY FAMILY TRANSCRIPTION FACTOR 29 (162)</td>
</tr>
<tr>
<td>At2g28550</td>
<td>RAP2.7</td>
<td>RELATED TO AP2.7 (163)</td>
</tr>
<tr>
<td>At3g48160</td>
<td>DEL1</td>
<td>DP-E2F-LIKE 1 (164)</td>
</tr>
<tr>
<td>At5g16600</td>
<td>MYB43</td>
<td>MYB DOMAIN PROTEIN 43</td>
</tr>
<tr>
<td>At1g80840</td>
<td>WRKY40</td>
<td>WRKY DNA-BINDING PROTEIN 40</td>
</tr>
<tr>
<td>At2g33810</td>
<td>SPL3</td>
<td>SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3</td>
</tr>
<tr>
<td>At5g47640</td>
<td>NF-YB2</td>
<td>NUCLEAR FACTOR Y, SUBUNIT B2</td>
</tr>
<tr>
<td>At1g54160</td>
<td>NFA5</td>
<td>NUCLEAR FACTOR Y, SUBUNIT A5</td>
</tr>
<tr>
<td>At3g50060</td>
<td>MYB77</td>
<td>MYB DOMAIN PROTEIN 77</td>
</tr>
<tr>
<td>At5g19360</td>
<td></td>
<td>ZINC FINGER (CCCH-TYPE) FAMILY PROTEIN</td>
</tr>
<tr>
<td>At3g50600</td>
<td>PIL6</td>
<td>PHYTOCHROME INTERACTING FACTOR 3-LIKE 6 (83)</td>
</tr>
<tr>
<td>At3g03750</td>
<td>SDG20</td>
<td>SET DOMAIN PROTEIN 20</td>
</tr>
<tr>
<td>At5g0820</td>
<td>NAC097</td>
<td>NAC DOMAIN CONTAINING PROTEIN 97</td>
</tr>
<tr>
<td>At3g07670</td>
<td></td>
<td>RUBISCO METHYLTRANSFERASE FAMILY PROTEIN</td>
</tr>
<tr>
<td>At5g67030</td>
<td>ABA1/ZEP</td>
<td>ABA DEFICIENT 1/ZEAXANTHIN EPoxidASE (167)</td>
</tr>
<tr>
<td>At4g25990</td>
<td>CIL</td>
<td>CHLOROPLAST IMPORT APPARATUS CILIA-LIK</td>
</tr>
<tr>
<td>At4g17750</td>
<td>HSF1</td>
<td>HEAT SHOCK FACTOR 1</td>
</tr>
<tr>
<td>At5g63160</td>
<td>BT1</td>
<td>BTB AND TAZ DOMAIN PROTEIN 1 (168)</td>
</tr>
</tbody>
</table>

We predicted the differential activity profiles for 194 TFs and their regulation strength to TGs (Supplementary Table S5). All the 194 TFs exhibited differential activity profiles across 11 stress conditions (Figure 4). Global TF activity profiles show differential behaviour across different stresses and ecotypes. NCA algorithm predicted 10119 connections between 194 TFs and 1410 TGs (data not shown). Based on the calculated correlation values, 4011 positive correlations were considered as positive regulatory signals (activation) and 6078 negative correlations were considered as negative regulatory signals (repression). We have applied filtering criteria based on the strength of the calculated regulatory signals ($r \geq 0.75$). The resulting NCA predicted network contained 5380 pairs of interactions between 182 TFs and 1199 TGs (Supplementary Table S5).

Comparison with interactions retrieved from literature, AthaMap and AraNet

Comprehensive experimental data for validating all the predicted TF–TG interactions are still lacking. We have cross-verified some of our TF–TG regulatory predictions with already available interactions from the few available resources. For this purpose, we have downloaded all the TF–TG interaction data from the AthaMap (55) and CCSB plant Interactome Database (AraNet) (54). With this comparison, we found 52 known interactions reported in AthaMap for six TFs and four interactions in AraNet for three TFs overlapped with our results (see Supplementary Table S7). The retrieved overlapping interactions were for the following TFs—EDF4 (ETHYLENE RESPONSE DNA BINDING FACTOR 4: At1g13260), PAP3 (PHYTOCHROME-ASSOCIATED PROTEIN 3: At1g09530), WRKY40 (WRKY DNA-BINDING PROTEIN 40: At1g80840), WRKY18 (WRKY DNA-BINDING PROTEIN 18: At4g31800), ABF1 (ABSCISIC ACID RESPONSIVE ELEMENT-BINDING FACTOR 1: At1g49720), GT2 (At1g76890), EFR11 (ETHYLENE RESPONSE FACTOR 11, At1g28370), NF-YA4 (NUCLEAR FACTOR Y, SUBUNIT A4: At2g34720) and RAV2 (RELATED TO ABI3/VP1 2: At1g68840).

To compare with an independent gene regulatory network, we have downloaded the TF–TG interactions from Vermeirssen et al. (57). The downloaded network contained 200 014 interactions between 1290 TFs and 11 938 TGs, generated from a heterogenous gene expression compendium of 199 abiotic conditions. Our network contained 10 534 TF–TG interactions between 240 TFs and 1638 TGs, generated from a homogeneous gene expression compendium of eleven biotic, abiotic and combined stress conditions. Only three stress conditions were overlapped (cold, heat and salt) between the two gene expression compendia. Despite the differences between both the studies, regarding the conditions, datasets used, number of TFs and TGs; in total, 649 TF–TG interactions were common between the two independent networks, representing 101 common TFs and 340 common TGs (see Supplementary Table S7).

Cross-validation using STIF algorithm

As an another independent cross-validation strategy, we have used a Hidden Markov Model based profile search
Figure 4. Predicted activity profiles of 194 stress-regulated TFs as a heat map. Activity profiles of 194 stress-regulated TFs (x-axis) during eleven stress conditions (y-axis), in 10 Arabidopsis ecotypes, predicted using NCA. The activity values were scaled to +1 to −1. Hierarchical clustering of the TFs performed using Pearson correlation with average linkage method. The predicted profiles show the differential activity of the TFs in different stress conditions, as well as in various ecotypes.

between 25 unique TF families and 1021 individual TGs, based on HMM profile search in 1000 upstream with its 5′ UTR of the TGs for finding enrichment of TF binding sites (see Supplementary Table S8). The Z-score threshold of 1.5 was chosen because it was demonstrated to be appropriate to 90% sensitivity during the validation of STIF algorithm (39). We merged all the redundant TF–TG pairs and created a list of 9311 unique pairs of TF–TGs. A pairwise comparison between STIF predicted unique pairs of TF–TGs versus NCA-predicted 5380 pairs of TF–TG interactions identified 982 common interactions between the NCA predicted regulatory network and STIF based on HMM profile search. This overlap included 20 TF families and 529 TGs (see Supplementary Table S10).

Stress wise regulatory sub-networks

To extract the 11 stress specific sub-networks from the global TF–TG network, we have considered the stress-wise TF lists and extracted all the connections to the associated TGs from the global network. Highly connected regulatory hubs have been identified, and corresponding biological relevance were retrieved from published literature. Detailed information of the sub-networks is provided in Supplementary Table S6 and as network topology diagrams in Supplementary Figures S2–11. Brief descriptions of the eleven stress-wise sub-networks and functional relevance of few of the highly connected regulatory hubs are described below.

Heat stress sub-network

HSFA2 (HEAT SHOCK TRANSCRIPTION FACTOR A2: At2g26150) is one of the key regulatory components of the cytoplasmic protein response during heat shock in Arabidopsis (93). Arabidopsis HSFB1 (HEAT SHOCK TRANSCRIPTION FACTOR B1: At4g11660) and HSFB2b (HEAT SHOCK TRANSCRIPTION FACTOR B2: At3g54660) act as repressors of many heat-inducible HSFs, but positively regulate the acquired thermotolerance (94). All members of the DREB2 (DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN 2) family (DREB2A, DREB2B and DREB2C) are involved in the regulation of heat-responsive genes. Multiprotein bridging factor 1c (MBF1c) protein functions as a transcriptional regulator which binds DNA and controls the expression of 36 different transcripts during heat stress, including the important transcriptional regulator such as DREB2A, HSFs and several zinc finger proteins (95). Phytochrome interacting factors (PIFs) and pseudo response regulators (PRRs) are known to control growth and flowering in Arabidopsis, by regulating hundreds of TGs via temperature-dependant clock-regulated pathways (96). The putative dehydration-responsive element-binding protein At2g40350 contains one AP2 (APETALA 2) domain, is known to be associated with heat acclimation in Arabidopsis (97). (See Supplementary Figure S2).

Salt stress sub-network

The major hub in the salt stress sub-network is TIFY10A (At1g19180), which plays positive regulatory roles dur-
ing responses to salt-alkaline stress in plants (98). Contrastingly, TIFY 11A (At1g17380), also known as JAZ5 (JASMONATE-ZIM-DOMAIN PROTEIN 5) protein acts as co-repressor, together with the Groucho/Tup1-type TOPOLESS (TPL) for the activation of jasmonate signalling pathways during stress (99). PHYTOCLOCK 1 (PCL1: At3g46640) plays a crucial role in modulating circadian clock associated transcriptional circuitry, mediated by light signals in a synergistic manner (100). Several WRKY TFs such as At4g18170, At4g18170 are an important component in Arabidopsis salt stress pathways, as well as in defence-related regulons (101). (See Supplementary Figure S3).

**Cold stress sub-network**

The CBF (C-repeat binding factor) regulon has a major role in cold stress response in plants. CBF1, CBF2 and CBF3 (also known as DREB1B, DREB1C and DREB1A) genes appear to be present across plant species and are almost always present as a gene family. The CBF TFs regulate a set of about 100 downstream genes (102). There are other CBF-independent pathways and their complex interactions between different components also significantly contribute to cold tolerance. Regulators such as CIL (At4g25990), RAV1 (At1g13260), ZEP (At5g67030), LBD37 (LOB DOMAIN-CONTAINING PROTEIN 37: At5g67420), MYBL2 (At1g71030), IAA29 (At4g32280), OZF2 (OXIDATION-RELATED ZINC FINGER 2: At4g29190) are believed to control cold stress response processes in plants (103). It is also known that light and cold signals integrate and cross talk for cold tolerance, via a CBF and ABA-independent pathways (102). Previous studies have reported that circadian regulators LHY1 (LATE ELONGATED HYPOCOTYL 1: At1g01060), CCA1 (CIRCADIAN CLOCK ASSOCIATED 1: At2g46830), TLI1 (At2g46790), COL1 (CONSTANS-LIKE 1: At5g15850) induced different cold stress response pathways in Arabidopsis plants (104).

**High-light stress sub-network**

Ethylene-responsive element binding factor 6 (ERF6: At4g01250) is one of the major regulatory hubs in biotic stress sub-network (See Figure 3D). As an early defence gene, also known to play a crucial role during signal crosstalk between pathogen-associated molecular patterns (PAMPs)-triggered immunity and photosynthesis (109). Another hub in the network is MYB117 (At1g26780), known to regulate alternative splicing events mediated by various stress signals (110). On the other hand, MYB124 (At1g14330) is associated with stomatal development in Arabidopsis. Entry of pathogens into the host tissue via natural surface openings, such as stomata is the critical first step in causing infection. Hence, MYB124 may play an active role by regulating stomatal development to limit bacterial invasion as part of the plant innate immune system (111). HSFA1A (At4g17750) is a master regulator of the multiple stress conditions and believed to have played an important role during the cross-talking event of biotic and abiotic stresses (112). Cytokinin-responsive gata factor 1 (At4g26150) is involved in gibberellic acid-mediated signalling pathway, negative regulation of flower development, negative regulation of seed germination, during stress conditions (113).

**Heat + FLG stress sub-network**

Global warming and associated severity of biotic invasions and abiotic stresses are the major threat to agricultural production globally. The Heat + FLG (biotic) sub-network represents an ideal regulatory network model for studying biotic and abiotic stress combination. Similar to the FLG sub-network, WRKY22 (At4g01250) is again the top most highly connected regulatory hub in heat + FLG subnetwork, reflecting its cross-talking role during biotic and abiotic combined stress conditions (109). SPOROCYTE-LESS (SPL) TFs (eg. SPL9: At2g42200) are master regulators of developmental transitions and known to be associated with microRNA pathway for the integration of multiple stress signals with developmental processes during the event of recurring multiple stresses (114). The NF-YB2 (At5g47640) is an important stress regulator, also known to required for COL5 (At5g57660) mediated photoperiod-dependent flowering (115). (See Supplementary Figure S6).

**Cold + FLG stress sub-network**

The simultaneous occurrence of cold temperature and pathogen attack results potentially negative impact on...
CBF-COR pathway in responses of cold and abscisic acid signalling to activate the MYB96 or MYBCOV1 (At5g62470) integrated discrete mediated signalling pathway, salicylic acid-mediated signal, ethylene-activated signalling pathway, jasmonic acid-mediated signalling pathway, salicylic acid-mediated signalling pathway, response to abscisic acid, response to auxin etc. (117). A recently published study has shown that the MYB96 or MYBCOV1 (At5g62470) integrated discrete responses of cold and abscisic acid signalling to activate the CBF-COR pathway in *A. thaliana*, ensuring plant adaptation to a diverse set of biotic and abiotic stresses (118). (See Supplementary Figure S7).

**Cold + high-light stress sub-network**

It has been hypothesized that putative light signalling pathways involved in low-temperature regulation of gene expression (119). TL1 (At2g46790) in the, also known as the pseudo-response regulator (PRR9) is an essential component of a temperature-sensitive circadian system (120). BTB and TAZ domain protein4 (BT4: At5g67480) regulates the expression of various abiotic stress-responsive gene expression by binding to different cis-acting elements in response to various stress signals (121). ZEP (At5g67030) is also a prominent hub in cold and high-light single stress regulatory network. (See Supplementary Figure S8).

**Heat + high-light stress sub-network**

Heat stress is a naturally co-occurring stress with high-light conditions, and both lead to dehydration or osmotic stresses in plants. NFYA5 (NUCLEAR FACTOR Y, SUBUNIT A5: At1g54160) is an important TF for drought resistance in *Arabidopsis*. Drought stress upregulates NFYA5 gene expression not only at the transcriptional level but also at the post-transcriptional level by down-regulating the expression of miR169a that targets NFYA5 transcript for cleavage (122). MYB70 (At2g23290) is one of the putative candidates of *Arabidopsis* repressome, reported to play key roles in diverse biological functions by negatively regulating genes involved in developmental, hormonal and stress signalling pathways (123). Under unfavourable environmental conditions such as an onset of heat, high-light, drought, salt, etc., the demand of the cell for protein folding can exceed its folding capacity and lead to the endoplasmic reticulum stress (ER-stress) by the accumulation of unfolded or misfolded proteins. The membrane-associated TF NAC089 (At5g22290) controls ER-stress induced programmed cell death in plants (124). Modulation of DNA methylation in response to environmental stress represents a potentially robust mechanism to regulate gene expression networks in plants. SDG20 (SET DOMAIN PROTEIN 20: At3g03750) is believed to be involved in the regulation of the *Arabidopsis* methylene through H3K9 methylation (125). (See Supplementary Figure S9).

**Salt + heat stress sub-network**

Salt stress and heat shock are two of the commonly co-occurring environmental conditions particularly for the plants in arid and semi-arid regions (126). The combination of heat and salt stress also generates drought and osmotic stress in the affected plants. The top most highly connected regulatory gene in the salt + heat sub-network is At2g40350, which encodes a member of the DREB sub-family A-2 of ERF/AP2 TF family protein involved in the response to drought (127). Similarly, MYB60 (At1g08810) controls the stomatal aperture, which helps to limit water loss during drought (128). The important regulatory role of WRKY22 (At4g01250) in stress signalling and crosstalk events during multiple conditions was mentioned above. PHYTOCHROME INTERACTING FACTOR 3 (PIF3/POC3) was shown to control transcriptional network in response to initial light exposure, and later, to develop sensitivity to other stress signals like drought and, reactive oxygen species (ROS) upon prolonged exposure (129). ROS-mediated homeostasis and signalling events implicated during drought and salinity stresses has been well demonstrated in earlier studies (130). (See Supplementary Figure S10).

**Salt + high-light stress sub-network**

ERF6 (At4g17490) is a member of the ERF TF family protein involve in response to ROS and light induced stress in plants (131). It has also been reported that modulation of ethylene responses affected plant salt-stress responses (132). Another hub in the salt + high-light network is At5g44260, which encodes a tandem zinc finger C-x8-C-x5-C-x3-H type family protein, involves in early stage response to ethylene in *Arabidopsis*. SOC1 (SUPPRESSOR OF OVEREXPRESSION OF CO : At2g45660) is a key component of *Arabidopsis* flowering time gene regulatory network and was shown to act as a negative regulator of the salinity stress-response via CYCLIN-DEPENDENT KINASE G2 (CDKG2) (133). The important role of plant-specific GATA-type zinc finger TF family protein (GATA8: At3g54810) in light intensity mediated signalling in *A. thaliana* was mentioned in the high-light stress sub-network section above (96). (See Supplementary Figure S11).

**Clustering of transcription factor binding sites (TFBS)**

Regulation of gene expression is executed through the combinatorial action of many CREs, including core promoters, promoter-proximal elements, as well as a various cis-regulatory modules (CRMs) that are situated distally from the TSSs, such as enhancers, silencers, insulators and tethering elements (134–136). We have analysed the clustering patterns of the TFBSs in the regulatory regions of each of the stress-regulated TGs. As many as 852 TF–TG pairs had multiple binding sites for the same TF, also known as homotypic clusters of TFBS (HCTs) (137). In total, 28 TF–TG
pairs had five HCTs, 164 pairs had four HCTs, 266 pairs had three HCTs and 394 pairs had two HCTs. Remaining 8359 pairs of TF–TG interactions were singletons and occurred in heterotypic clusters of TFBSs (see Supplementary Table S9). Heterotypic clusters of TFBSs are CRMs consisting of unique classes of TFs in different combinatorial patterns. The homotypic clusters of multiple TFBSs were prominently enriched for MYB, bHLH (BASIC HELIX-LOOP-HELIX) and bZIP (BASIC LEUCINE-ZIPPER) TF families, which were previously reported as versatile master regulators controlling diverse stress signalling events in plants (138).

Gene set enrichment analysis (GSEA) of the overlapping interactions suggests their involvement in stress response processes

To obtain an overview of the functional categories enriched among the overlapping TGs (529 genes), GSEA was performed using BiNGO tool (60). The network-based visualization of statistically significant over-represented functional categories identified four distinct clusters (Figure 5 and Supplementary Table S11).

Cluster 1: primary processes (photosynthesis, carbohydrate metabolism, lipid metabolism, growth, anatomical development, reproduction, DNA replication, etc.).

Cluster 2: secondary processes (processes related to secondary metabolism, small molecule metabolic processes, production of defence-related secondary metabolite, etc.).

Cluster 3: stress response (response to heat, salt, cold, wounding, radiation, water deprivation, other organisms, etc.).

Cluster 4: regulation of biological processes.

DISCUSSION

Reconstruction of gene regulatory networks using gene expression data from multiple ecotypes, and under multiple conditions is an ideal strategy for identifying important regulators that control the dynamic behaviour of cellular processes related to stress responses in plants. Using a homogeneous dataset from a uniform experiment has the advantage of minimizing the potential influence generated by experimental heterogeneity. To explore the transcriptional regulation in plants during multiple stress conditions, we have constructed an in-silico transcriptional regulatory network model in cellular responses to 11 stresses (five single and six combined) in A. thaliana, using a homogenous gene expression dataset from ten natural ecotypes. Considering the number of differentially regulated genes as a measure, signatures of transcriptomic response variations were observed among the ten Arabidopsis ecotypes while responding to the eleven stress conditions. There were variations within an ecotype while responding to different stresses, as well as variations among different ecotypes while responding to a particular stress condition. There were observable differences between responses to single stress versus response to the combination of stresses (Table 1). Regulation of gene expression is a critical component in development and evolution of living beings along with genome composition, structure and function. Responding to the changing environmental conditions, a functional cell of any living organism would alter the expression of particular genes through transcriptional regulation to make it possible to preserve the robustness of cellular processes (139).

In this paper, we focused mostly on inferring a global stress regulatory network and 11 conditions dependent regulatory sub-networks in the model plant A. thaliana by uniting the dataset from 10 ecotypes and 11 stress conditions. The united lists of stress-regulated genes from all the stress conditions and 10 ecotypes contained 3429 differentially expressed genes, out of which 294 were Arabidopsis TFs. The initial bi-partite network contained 10 534 interactions between 240 stress regulated TFs and 1638 TGs. NCA algorithm was able to predict activity profiles of 194 TFs. Applying the regulatory signal strength based filtering criteria, 5380 connections were retained between 182 TFs and 1199 TGs. The stress-wise lists of differentially expressed TFs were used to generate eleven condition-dependent regulation regulatory sub-networks. Very few TFs were differentially expressed in more than one stress conditions, and most of the multifunctional regulators were directly or indirectly related to circadian pathways, controlling various physiological and developmental processes coupled with environmental stimuli. Circadian clocks consist of positive and negative elements that comprise a transcriptional feedback loop (140). There have been several reports about circadian clock genes that are not essential for oscillator control under diurnal cycles of light and dark (141). Many such circadian clocks of higher plants were believed to act as the master regulator by orchestrating the various layers of metabolic and physiological processes that coordinate growth and performance under a changing environment (142,143). Our analysis identified few highly connected clock-related genes in multiple stress conditions. Stress-specific lists of TFs were identified for five single and six combined stress treatments. Many of the single stress-regulated TFs were already reported in benchmarked literature elsewhere (14,17,19,144). However, results for combined stress-regulated TFs were not available in multiple ecotypes of A. thaliana. We have created six lists of TFs for each of the combined stress conditions used in our analysis, and extracted their corresponding interactions with TGs. Information regarding regulation of combined stress gene expression is of great interest to study signalling crosstalk events during combined stress conditions (145). It is a well-known fact that TFs alter their interactions to varying degrees in response to diverse stimuli, thereby rewiring the whole network topology. By analysing the stress regulatory sub-networks, we have observed that a few TFs serve as multifunctional regulatory hubs, but most act transiently only responding to certain conditions (146). Finding functional association of the condition-specific regulators in combined stress conditions is difficult due to the non-availability of published information in this direction. The multifunctional regulators maintain the core stress response processes while the transient regulators confer the specificity to certain conditions by regulating a diverse range of downstream TGs. The relationship between network topology and TF activity dynamics allow differential utilization of the same underlying network topology by different regu-
Figure 5. GSEA of the overlapping TGs. Network based visualization of GSEA results identified four clusters of statistically significant over-represented functional categories. Parental GO terms are shown in the centre. The differentially regulated primary processes were present in cluster 1, and secondary processes in cluster 2. The various stress response processes were distinctly clustered together in cluster 3. Processes that regulate the series of complex stress–response events were grouped together in cluster 4. These categories reflect the sequential modulation of stress-response processes. Upon exposure to stressed conditions, plants compromise with primary processes such as primary metabolism, photosynthesis, reproduction, to allocate extra resources to the activation of various defence and stress-response processes and secondary metabolic processes. Nodes were coloured according to their corrected P-values. Node sizes reflected the total number of genes in that category.

Consider how each TF contributes independently to gene expression. However, researchers have now demonstrated the complex regulation of gene expression by combinatorial promoters (147). Identified HCTs in the predicted stress regulatory TF–TG network were mainly comprised of MYB, bHLH and bZIP TF families. The stress-regulated lists of TFs contained 17 MYBs, 16 bHLHs and 7 bZIP members of TFs respectively (see Supplementary Table S3). Literature survey confirmed the role of MYB, bHLH and bZIP TFs as conserved ‘versatile master regulators’ in signalling pathways in responses to multiple stress conditions in plants (138,148–149). Our results suggest that the homotypic clusters of TFBSs may play important evolutionary role for preserving the transcriptional robustness of the key stress regulators by avoiding any mutational effect in the cis-regulatory regions. Clustering of multiple TFBSs for the same TF is a common feature of CRMs in lower organisms (137). Several
reports have indicated the potential role of cis-regulatory evolution on natural variations in stress-responsive gene expression in plants (61,66,73,150). However, the occurrence of such CRMs in the plant genome has remained largely unknown (58). Large scale experimental screening of cis-regulatory regions of stress associated genes in multiple ecotypes, together with sequence polymorphism data would be able to confirm the evolutionary significance of the HCTs in stress response, acclimation and successive evolution. Generating large-scale phenotypic and physiological data and combining predictive models with downstream experimental characterisation would also be an added advantage in this regard.

Network-based GO analysis of the overlapping TGs has shown the significant over-representations of sequentially activated defence response processes. The complex orchestration of plant biotic and abiotic defence responses require an adequate supply of energy (149,151). While modulating the multitude of defence response processes during a physiologically stressed condition, plants have to reorganize some of the resources of central metabolism by allocating resources for the activation of defence-related processes (152). Along with secondary metabolism, primary metabolism of the plant is also differentially regulated during a stressful event and finally lead to smoothly orchestrated defence responses. The resulting GO-network (Figure 5) reflects this observation.

Experimental evidences regarding TF-TG interactions in plant systems are very limited (58). Again, within the available information, condition-specific regulatory interactions in plants are almost rare. It poses a significant challenge in benchmarking computational predictions. By cross-comparing of our predicted network with the limited information retrieved from AraNet and AthaMap, we were able to identify few known interactions. Considering the differences in used data, conditions and methods, the overlap of 649 TF-TG interactions (out of 10 534) with an independently predicted network was relatively good (57).

Our study has suggested many novel interactions between key TFs and differently expressed TGs during the stress response during 11 conditions. This information will help to identify and experimentally validate molecular markers or relevant stress regulators by forward genetics experiments to develop single and multi-stress tolerant crop varieties. In addition, similar integrative approaches could be employed to study networks regulating any biological process in any biological system.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

P.B., N.D.J and A.M.B. acknowledge infrastructure support from the Norwegian University of Science and Technology. M.N.B.N., K.S. and R.S. would like to acknowledge National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India for infrastructure support.

FUNDING

The Multi-Stress transcriptomic experiments were funded by the European Research Area Network Plant Genomics (ERA-NET PG) (http://www.erapog.org/) project; Biotechnology and Functional Genomics (FUGE) and FRIMEDBIO programs of the Norwegian Research Council [RCN184146 and 214329 to A.M.B]; Department of Biotechnology, India for the development of STIF algorithm (to K.S., R.S.); NCBS-Bridge Postdoctoral Fellowship of the National Centre for Biological Sciences, TIFR, Bangalore, India (to M.N.B.N.). Funding for open access charge: Norwegian Research Council. Conflict of interest statement. None declared.

REFERENCES


