Transcription Factors-microRNAs regulatory network in response to multiple stresses in Arabidopsis thaliana

Fereshteh Izadi1*, Roohollah Nikfekr2, Jahad Soorni1

1Department of Plant Breeding and Genetics, Sari Agricultural sciences and Natural Resources University (SANRU), Mazandaran, IRAN
2Department of Plant Breeding and Genetics, Payame Noor University, PO BOX 19395-3697 Tehran, IRAN

*Corresponding author: izadi1991@yahoo.com

Abstract

Transcription factors (TFs) by interacting with the DNA binding sites of stress responsive gene promoters, control a cascade of reactions in plant cells also on the other hand microRNAs (miRNAs) act on dozen of genes involving in response to a particular stress condition via many of biological processes. Less than 6% of the responding transcripts showed antagonistic responses to stress combinations that require delineation of regulatory interactions underlying multiple concurrent stresses instead of a single biotic or abiotic stress. Therefore, inferring transcriptional networks including both of TFs and miRNAs as two classes of regulators likely will reveal more aspects of complex interlayers especially in a combination of stresses. Utilizing 207 microarrays, we performed data mining approaches following by network construction. As a result, 17 TFs and 12 miRNAs were identified as the most potential regulators in response to a combination of biotic and abiotic stresses. Seven stress responsive TFs targeted 1697 transcripts in ARACNE based constructed network which classified in a cohort of common metabolisms under investigated treatments. The generated network was validated via comparison with a reference network and further motif enrichment. The study identified TFs like ERF/AP2 family also miRNAs including miR5658 and miR172 their crosstalk may undertake transcriptome changes in response to stresses.

Keywords: gene regulatory network, miRNA, multiple stresses, target gene, transcription factor.

Abbreviations: ARACNE_Algorithm for the Reconstruction of Accurate Cellular Networks; AUPR_Area Under the PR curve; AUROC_Area Under the ROC Curve; bHLH_basic HELIX-LOOP-HELIX; bZIP_basic LEUCINE-ZIPPER; CLR_Context Likelihood of Relatedness; DoF_DNA-binding one zinc finger; ERF_Ethylene-responsive transcription factor; GEO_Gene Expression Omnibus; GO_Gene Ontology; GRN_Gene Regulatory Network; limma_Linear Models for Microarray Data; MI_Mutual Information; miRNAs_MicroRNAs; NCBI_National Center for Biotechnology Information; ROC_Receiver Operating Characteristic; TFs_Transcription Factors

Introduction

Plants are sessile organisms and continuously face to external stimuli including various biotic and abiotic stresses thereby gene expression play crucial role in adaptation to changing environments. Biotic and abiotic stresses cause substantial loss to agricultural production globally. Therefore, enormous researches have been conducted on understanding of plants responses to heat, salt, droughts, cold, high light intensity and pathogen infection stresses (Kreps et al., 2002; Seki et al., 2002; Rizhsky et al., 2004; Oono et al., 2006; Kleine et al., 2007; Hannah et al., 2010; González-Pérez et al., 2011; Navarro et al., 2004; Nielsen et al., 2007). A Combination of different stress factors including heat and drought or cold and high light intensity resulted in a severe penalty on plant performance and yields. Moreover this have been exhibited that major fraction of responses to a combined of stresses can’t be delineated from transcriptome changes in response to one or two stresses (Rasmussen et al., 2013). Thus, depiction of the master regulators that respond to combinations of abiotic and biotic stresses is required for breeding and engineering more stress tolerant crop plants. miRNAs are ~22 nt small, single-stranded and noncoding RNAs that regulate gene expression by repressing mRNA translation or decreasing stability of mRNAs and have proven vital roles in controlling gene regulation (Yang et al., 2013; Shan et al., 2013). Accumulating evidences indicate the role of miRNAs in a wide range of biological processes in plants including organ polarity, leaf growth and male or female sterility (Bowman, 2004; Chuck et al., 2007; Millar and Gubler, 2005). Also it has been observed that several of plant’s miRNAs respond to biotic and abiotic stress conditions and several of miRNA targets are stress-related genes suggesting important roles that miRNAs play in response to stresses (Phillips et al., 2007). A number of stress-specific miRNAs have also been identified in model plants under various biotic and abiotic stress conditions, including high salinity (Sunkar et al., 2008), drought (Zhao et al., 2007), cold (Zhou et al., 2008), oxidative stress (Sunkar et al., 2006), UV-B radiation (Zhou et al., 2007) and bacterial infection (Navarro et al., 2006). To date, a myriad of TFs and miRNAs have been associated with biotic and abiotic stresses in Arabidopsis. Although the modules of miRNAs-TFs including miR156–SPL, miR319-MYB and miR172–AP2 have been illustrated to be involved in biological pathways in plants (Spanudakis and Jackson, 2014), previous studies mostly ignored the
reciprocal regulation between TFs and miRNAs in multiple stresses. Nonetheless, apart from noteworthy efforts on the identification of canonical gene sets induced by each of biotic and abiotic stresses, the underlying cross talk of a triangle of miRNAs, TFs and targets in a combination of different stresses have been little studied. Microarray technology facilitates the simultaneous expression profiling of tremendous of genes and one can exploit fundamental aspects of different biological process by computational analysis on data came from this technology. Meta-analysis of microarray studies involving samples from a wide range of tissues, developmental stages and different levels of multiple stress conditions would unravel interesting features related to the stress response (Altmann et al., 2004; Schmid et al., 2005). In the present study, we performed a meta-analysis of gene expression data to elucidate crosstalk between biotic and abiotic stresses and involved miRNAs and TFs. We identified a small number of stress-related genes that could be future targets for plant researches toward improving agricultural yields.

Results and Discussion

The identification of potential regulators and network construction

Comprehensive analysis using omics studies have revealed the interactions between a combination of abiotic and biotic stress components with the identification of several genes and TFs involved in stress crosstalk (Liu et al., 2012; Yokotani et al., 2013). While a TF may be post-transcriptionally regulated by miRNAs, there are very less information regarding interactions within multiple stresses. In order to exploring the putative target genes regulated by TFs in crosstalk with miRNAs we built a GRN by mutual information based network learning algorithm ARACNE. To this end, we used 30801 probsets that passed the filter of 207 arrays from a combination of salt, heat, cold, high light intensity and flagellin treatments versus the matched controls. We next extracted 94 stress responsive miRNAs from PMTED database and a study by Xie et al. (2015) (Supplementary Table 1) which their predicted targets were obtained by psRNA Target server. Target prediction of miRNAs led to the identification of total 840 genes regulated by those miRNAs (Supplementary Table 2). Hereby we seek these target genes for identifying TFs focusing on the TFs that expressed at p-value ≤ 0.01 threshold. As a result, 17 TFs (Table 1) were found from which based on STIFDB2 server. seven TFs are strongly known as stress responsive TFs that were applied as the hub genes within ARACNE for GRN inference (Fig 1). As we see in the Table 1, 17 TFs are presumably regulated by miRNAs from 12 families mostly from miR5021 and MiR5658 families. MiR5021 was recently found to target a diverse sets of genes including global transcription factors and disease resistance proteins (Das et al., 2015). Furthermore, miR5658, miR172, miR169, miR166 and miR165 were also reported to act on the TFs (Cheng et al., 2016; for more details see Khraiwesh et al., 2012). Fallowing miR5658 was newly reported to be involved in stress response and observed to target MYB, bZIP, bHLH, AP2 and ERF TFs in absicic acid responsive transcriptome in tomato (Cheng et al., 2016). Among the aforementioned miRNAs in Table 1, miR319, miR167, miR172 and miR165 have been demonstrated to be involved in pathogen defense and a wide range of abiotic stresses (Sunkar et al., 2012; Singh et al., 2012). These findings may imply on possible interactions among the responsive miRNAs and TF’s targets triggering downstream changes in transcriptome to improve stress tolerance in plants. Seven selected TFs as hub genes encoded for TFs from several families like Zinc finger, bHLH, WRKY and AP2. Interestingly AT1G14200 that encodes a zinc finger was observed to be ranked among the 500 top differentially expressed genes in this experiment in response to heat, cold and Flagellin treatment moreover showed notable changes in response to pathogen and combined light and drought stress (Ascencio-Ibáñez et al., 2008; Giraud et al., 2008).

Validation of constructed network

A GRN is a graph representation of biological units in which nodes represent genes and edges are the interactions between nodes. Information-theoretic approaches like ARACNE (Margolin et al., 2006) and CLR (Faith et al., 2007) have been successfully applied for reconstructing GRNs (Emmert-Streib et al., 2014). In these approaches first a pair-wise MI matrix is being calculated between all possible pairs of genes. Afterward, this matrix is being manipulated for identifying regulatory interactions between nodes. In this section in order to identifying downstream events controlled by stress responsive TFs and miRNAs we utilized a relatively large amount of transcriptomics data gathered with microarrays technique. Our reasons for selecting and exploiting these datasets are that: I) these data cover multiple combinations of biotic and abiotic stresses in Arabidopsis and II) the predictions can be compared to a gold standard list consist of 4775 experimentally validated gene regulatory interactions in Arabidopsis obtained from AGRIS database (Supplementary Table 3). This list was used to construct a reference network required for validate function in R package minet to assess the prediction accuracy of built networks ARACNE and CLR. As illustrated in Fig 2 in term of ROC curves, although inferred networks had a high overlap with reference network, ARANCE was the most powerful to predict higher rate of true edges over the CLR. This indicates that the ARACNE based GRN correctly identifies the affected 1697 target genes governed by each of the 7 TFs. AUPR and AUROC values for ARACNE were 0.0003, 0.695053 and for CLR were 0.0001, 0.5760852 indicating a more precise network inferred by ARACNE.

Compared to previous regulatory interaction studies our in silico analysis had two merits; first instead of being merely predicted interactions our inferred networks were validated by comparing with reference network. Furthermore, we investigated reciprocal regulation between TFs and miRNAs that may contribute to the gene-expression variation leading fine-tuning of stress response.

Enrichment analysis of miRNA-TF regulated target genes

We used PlantRegMap server at p-values ≥ 0.01 to find the significantly over-represented biological GO terms and functions of gene products in which predicted target genes involved. We observed that the 1697 potential target genes inferred by ARACNE enriched for common metabolisms under combined stress conditions like potassium ion transport (GO: 0006813, 0.00441), carbohydrate derivative binding (GO: 0097367, 0.00047), toxin catabolic process (GO: 0009407, 0.00509), response to abiotic stimulus (GO: 0050896, 0.00567), absicic acid glucosyltransferase activity (GO: 0010294, 0.00617) and secondary metabolite catabolic process (GO: 0009487, 0.00509) that have reported to be in
Table 1. List of commonly regulated miRNAs under cold, heat, high light intensity, salt and Flagellin stresses and the TFs they targeted. TFs represented in bold were used as hub genes for GRN construction.

<table>
<thead>
<tr>
<th>TFs</th>
<th>Annotation</th>
<th>Responsive miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G02030</td>
<td>Zinc finger</td>
<td>ath-miR5021</td>
</tr>
<tr>
<td>AT1G06150</td>
<td>bHLH</td>
<td>ath-miR172a,b,c,d</td>
</tr>
<tr>
<td>AT1G10120</td>
<td>bHLH</td>
<td>ath-miR396a</td>
</tr>
<tr>
<td>AT1G14200</td>
<td>Zinc finger</td>
<td>ath-miR5021</td>
</tr>
<tr>
<td>AT1G14920</td>
<td>GA1</td>
<td>ath-miR5658</td>
</tr>
<tr>
<td>AT1G16060</td>
<td>AP2</td>
<td>ath-miR5021</td>
</tr>
<tr>
<td>AT1G17590</td>
<td>NF-YA8</td>
<td>Ath-miR169a,b,c,h,i,j,k,l,m,n</td>
</tr>
<tr>
<td>AT1G18750</td>
<td>AGL65</td>
<td>ath-miR5658</td>
</tr>
<tr>
<td>AT1G20696</td>
<td>HMGB3</td>
<td>ath-miR5021</td>
</tr>
<tr>
<td>AT1G27360</td>
<td>SPL11</td>
<td>ath-miR156a,b,c,d,h,i,j,ath-</td>
</tr>
<tr>
<td>AT1G27370</td>
<td>SPL10</td>
<td>miR157a,b,c,d</td>
</tr>
<tr>
<td>AT1G28040</td>
<td>Zinc finger</td>
<td>ath-miR854a,b,c,d,e</td>
</tr>
<tr>
<td>AT1G28420</td>
<td>RLTI</td>
<td>ath-miR5021</td>
</tr>
<tr>
<td>AT1G30210</td>
<td>ATTCP24</td>
<td>ath-miR319c</td>
</tr>
<tr>
<td>AT1G30330</td>
<td>ARF6</td>
<td>ath-miR167c</td>
</tr>
<tr>
<td>AT1G30490</td>
<td>ATHB9</td>
<td>ath-miR165a,b, ath-miR166a,b,c,d,e,f,g</td>
</tr>
<tr>
<td>AT1G30650</td>
<td>WRKY14</td>
<td>ath-miR5021</td>
</tr>
<tr>
<td>AT1G32360</td>
<td>Zinc finger</td>
<td>ath-miR5658</td>
</tr>
<tr>
<td>AT1G33240</td>
<td>ATGTL1</td>
<td>ath-miR854a,b,c,d,e</td>
</tr>
</tbody>
</table>

Fig 1. Regulatory network including miRNA-TF's targets as hub genes under cold, salt, high light intensity, heat and Flagellin treatments inferred by ARACNE. Seven commonly expressed miRNA regulated TFs were placed in the centers of seven non-overlap modules targeted a total number of 1697 genes predicted by ARACNE. TFs and target genes were shown in red and green respectively.

Fig 2. ROC curves of GRN obtained by ARACNE algorithm versus corresponding network derived by CLR. Horizontal and vertical axis represent false positive and true positive rates respectively. GRN derived by ARACNE contained more true positives compared to network by CLR and therefore ARACNE more fits to be used as GRN inference method in this study.
involved in transcriptome changes in response to single and combined stresses (Wang et al., 2013; Hummel et al., 2010; You et al., 2015; Ben Rejeb et al., 2014; Del Carmen Martínez-Ballesta et al., 2013) (Supplementary Table 4).

**Analysis of cis-regulatory elements in putative target genes**

TFs regulate gene expression in different tissues and at various developmental stages in plants and as important regulatory components of the genome are the main targets for engineering stress tolerance. Usually, TFs are comprised of a DNA-binding domain that interacts with the cis-regulatory elements of its target genes (Katagiri and Chua, 1997) and a protein-protein interaction domain that facilitates oligomerization between TFs and other regulators. A number of previous studies have demonstrated regulatory networks of TF-miRNA leading meaningful transcriptional changes during stress-response processes (Bandyopadhyay et al., 2009; Gennarino et al., 2012; Liu et al., 2010; Zhang et al., 2015). As a validation step of target genes, motifs within the promoter of these genes was enriched using PlantRegMap database. P-value and q-value in promoter analysis show the probability of TF occurrence at binding sites (Supplementary Table 5). A total of 409 binding sites were identified in the upstream of miRNA regulated TF's target genes corresponding to 152 TFs (Supplementary Table 5) from AP2, C2H2, bHLH, B3, ERF, Dof, h, ZIPBBR-BPC and E2F/DP families. Among these families bHLH and bZIP TFs, were previously reported to control a wide range of stress signaling events in response to multiple biotic and abiotic stresses in plants (Balderas-Hernández et al., 2013, Singh et al., 2002). Next several of TFs belong to zinc finger family were identified that link to basal defense against pathogen attacks (Bianchi et al., 2015). Dof TFs have been functionally characterized in Arabidopsis and a plenty number of studies have shown that Dof TFs are involved in plant growth and development, seed germination, photosynthesis, and biotic/abiotic stress responses in many species (Ma et al., 2015). Moreover, Dof TFs can be used as biomarkers in response to heat and salt stresses (Wu et al., 2016). Furthermore, AT1G06160, AT2G44840 and AT5G43410 (Fig 3) were observed as the most significance TFs that bind to the cis-regulatory elements in the promoter of potential target genes. These TFs encode members of ERFs subfamily B-3 of ERF/AP2 family containing one AP2 domain. Interestingly, AT5G43410 encodes members which bind to GCC elements at the promoter of their target genes and are induced by pathogens. The AP2/ERF family is one of the largest of TFs family that are involved in plant development and response to abiotic stresses as well as pathogen attack via ethylene signaling pathway (Wu et al., 2015; Guo and Ecker, 2002). They act as a trans-acting factor.

![AT1G06160](image1)

![AT2G44840](image2)

![AT5G43410](image3)

Fig 3. Three motifs were found as the most probable binding sites in upstream of miRNA-TFs target genes by using the PlantRegMap server matched to AT1G06160, AT2G44840 and AT5G43410 TFs belong to AP2 family. These motifs were evidently found to be correlated with not only different abiotic stresses but also with response to pathogen attack.
interact with drought-responsive elements triggering downstream regulation events to improve stress tolerance in plants. Reportedly, TFs contained B3 DNA binding domains have the ability to modulate the transcription levels of a subset of other ERF TFs (Liu et al. 2013) that may biologically be relevant with complex regulatory circuits of ERF/AP2 TFs with another TFs in occupying binding sites in response to multiple stresses.

Materials and Methods

Data collected and pre-processing

We first downloaded XYS files for GSE41935 series from NCBI (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. This data consists of a total of 207 samples which was based on the platform of GPL16226 correspond to cold, heat, high-light intensity, salt, and Flagellin treatments with matched controls. Next, raw files were preprocessed with quantile method. After quantile normalization of the raw data, limma R package was applied to extract edges between control genotype and treatment effect (p-value < 0.01, fold change ≥ 2.0). Giving a total of 207 arrays for the analysis, the normalized log₂ expression values of 30380 transcripts filtered for the analysis.

The identification of TFs that are putatively targeted by miRNAs

We extracted 94 stress-assocciate miRNAs belong to 23 different families in Arabidopsis from PMTED database (http://pmted.agrinome.org/) and study by Xie et al. (2015). Subsequently, their sequences were retrieved from mirTarbase datasets (http://mirTarbase.nbc.ncu.edu.tw/). We used psRNATarget, a small plant RNA target prediction server (http://plantgrn.noble.org/psRNATarget/) for predicting the putative target genes of mentioned stress-related miRNAs. For the identification of transcription factors among the predicted targets, we used a list of 2576 Arabidopsis transcription factor AGI IDs compiled from AGRIS (Yilmaz et al., 2011), DAFT (Guo et al., 2005), PlantTFDB (Jin et al., 2014) and RARFT (Iida et al., 2005) (Supplementary Table 6). Notably we only focused on the TFs which were expressed above the defined threshold (p-value ≤ 0.01). We have used STIIFDB2 server (http://caps.ncbi.res.in/stifdb2/) for the identification of stress responsive TFs among the extracted TFs.

Transcriptional network inferred by ARACNE

Using ARACNE (http://wiki.c2b2.columbia.edu/workbench/index.php/ARACNe) we generated a GRN by setting p-value to 0.05, Adaptive Partitioning and DPI tolerance to 0 (Fig 1). ARACNE uses an information-theoretic approach to eliminate indirect links. In this approach first a pair-wise MI matrix is been calculated between all possible pairs of genes afterward the resulted matrix manipulated for identifying regulatory interactions between nodes (Margolin et al., 2006).

Statistical evaluation of generated GRNs

To assess the accuracy and validation of constructed networks we used CLR, another MI-based algorithm embedded in minet R package (https://www.bioconductor.org/). The performance of the generated networks was compared on the basis of 3521 edges of reference network derived from gold standard list obtained by AGRIS (http://arabidopsis.med.ohio-state.edu/). The GRN inferred by ARACNE algorithm is already sparse we then used whole of the interactions against reference network when about the CLR network we tested the first 1000 highly ranked edges. The perfore of ARACNE and CLR derived networks was illustrated as ROC curves by R package minet. To further quantifying the accuracy of algorithms we computed AUPR and AUROC values by auc.pr(table) and auc.roc(table) functions existing in minet R package when higher values shows higher true positive rates.

GO annotation analysis

Here to understand the functions of the potential target genes and classify them into three main categories namely biological process, molecular function and cellular localization, GO analysis was performed by using the PlantRegMap database (http://plantRegMap.cbi.pku.edu.cn/prediction.php) at p-values ≤ 0.01.

Promoter sequence analysis for finding out the cis-regulatory elements

We extracted one kb upstream region of common target genes of TFs from TAIR database and analyzed them for finding the commonly elicited cis-regulatory elements conserved among promoter sequences of the selected genes by utilizing the PlantRegMap database.

Conclusion

To sum TFs and miRNAs can be considered as two classes of regulators that their regulatory loops may play important roles in governing hostile conditions by regulating the driver genes in multiple biological processes. In the frame of accurate approaches of data mining we identified a small number of regulatory motifs, miRNAs and transcription factors whose interplays can lead to broad spectrum resistance to biotic and abiotic stresses. We noted that in agreement with previous experimental confirmations, how bioinformatics is robust to choose of marker genes and depict the structural and functional features of biological networks and compared to the methods merely based on single genes derived by differential expression analysis, regulatory networks may represent more credible information. Of note, network analysis at transcriptome level could be more intensified through merging studies with protein networks to draw more precise conclusions regarding predicted master regulators. Finally, we inferred an undirected network while connectivity between nodes does not mean the causal relationships. An expected next step of such a work would possibly be exploring of interactions among the regulators in vivo that shed light the bridges between biological regulators and response to single and combined stresses.

Competing interests

The authors declared that they have no competing interests.

Acknowledgements

We thank Dr Nooshin Omranian, scientific staff in Systems Biology and Mathematical Modelling Group, Max Planck Institute for Molecular Plant Physiology, Potsdam, Germany for her precious assistances.
References


189