Plant Omics Journal POJ

POJ 6(4):278-285 (2013) ISSN:1836-3644

In silico analysis of putative transcription factor binding sites in differentially expressed genes: Study of the turnover of TFBSs under salt stress responsiveness in solanaceae family

Sanchita*, Blessy M. Baby, Ashok Sharma

Biotechnology Division, CSIR-Central Institute of Medicinal and Aromatic Plants, Post Office CIMAP, Lucknow- 226015, India

***Corresponding author: 0804sanchita@gmail.com**

Abstract

The binding of transcription factors in transcription factor binding sites (TFBSs) play a key role in the regulation of different biological processes showing change in the expression of genes in presence of adverse conditions. In this study, an *in silico* analysis of the data of differentially expressed genes of seven plants of family solanaceae under different time periods of salt stress was performed to find out the TFBSs. The data used for the study was retrieved from the public domain. Our analysis revealed up and down expression of genes that might result due to binding of transcription factors in promoter region. The promoter regions of differentially expressed genes were utilized for the prediction of TFBSs. The prediction was done using position weight matrices (PWMs) constructed by taking the data of experimentally validated transcription factors and their respective binding sites. The PWMs were scanned over promoter sequences present in the upstream region of differentially expressed genes. The TFBSs with a threshold of similarity score ≥ 2.97 were selected to get highly up and down expressed genes. These predicted TFBSs would be of help to understand the role of turnover of TFBSs responsible for the change in expression of genes under salt stress.

Keywords: Gene expression, Gene Expression Omnibus, Position Weight Matrix, Salt stress, Solanaceae, Transcription Factor.

Introduction

Microarray technology has become one of the key tools that is being used to monitor expression levels of genes in a given organism. The use of miniaturized microarrays for gene expression profiling was first reported by Schema (Schena et al., 1995). Lashkari et al., 1997 first studied the complete eukaryotic genome of *Saccharomyces cerevisiae* on a microarray. The main objective of a microarray experiment is to identify differentially expressed gene profiles (Yang et al., 2005; Sreekumar et al., 2008; Dudoit et al., 2002). This is being done by comparing the expression of a set of genes in a particular condition with reference genes maintained under normal condition (Babu 2004). The differential expression is the identification of genes, expressed at different levels, at the time of transcription under adverse conditions. The change in expression of genes at which mRNA are synthesised from a DNA template is regulated by different mechanisms. The most widely studied being regulation by transcription factors (TFs) (Romeuf et al., 2010). The activity of TFs is modulated through the signals of external stimuli from signal transduction pathways. Transcription factors (TFs) are DNA binding proteins, which play a central role in gene expression by regulating the process of transcription (Siddharthan, 2010; Won et al., 2010). TFs facilitate or inhibit recruitment of the RNA polymerase by binding to DNA, usually near the gene that they regulate (Siddharthan, 2010). They are the key molecular switches that control or influence several biological processes such as development, growth, cell division and responses to environmental stimuli in a cell or organism. By being capable of activating or repressing the transcription, the stimuli affect the metabolism, physiological balance and progression in cells and the responses of cells to the environment (Mochida et al., 2011). TFs form complex regulatory networks at the transcriptional level through protein-protein interaction. The specific interaction between TFs and a family of cis regulatory elements play a central role in the regulation of expression of genes (Mochida et al., 2011). Protein binding sites in the upstream region represent the regulatory elements to which these TFs bind. These regulatory elements are also known as the transcription factor binding sites (TFBS). Transcription factors have a specific region responsible for binding to the DNA known as DNA binding domain (DBD) (Latchman, 1997). TFBS are usually short (around 5-12 base pair (bp)) and they are frequently found in degenerate sequence motifs. A single TF might have multiple binding sites depending upon the conditions. Although consensus sequences might be frequently used to depict the binding specificities of a TF, a common way of representing the degenerate sequence preferences of a DNA-binding protein is by a position weight matrix (PWM), also known as a position-specific scoring matrix (PSSM). The elements of PWM correspond to scores reflecting the likelihood of observing a nucleotide at a particular position of the known TFBS. In eukaryotic genomes, the regulatory elements are identified in non-coding or upstream regions. These regulatory elements act as binding sites for multiple interacting transcription factors playing a role in the regulation of a single gene. There can be great variability in the binding sites for a single transcription factor (Sinha and Tompa, 2002). The analyses of putative TFBSs are usually based on experimentally verified TFBSs. Numerous studies have been done for transcription factors, that are important in regulating plant responses to stress (Sundar et al., 2008). In the present work, our objective was to find out the TFBSs in the given set of genes, which were showing change in expression under salt stress. Turnover of these binding sites in the responsible genes for a given TF might be the factor

leading to differential expression of genes. The identification and analysis of putative TFBSs in differentially expressed genes will further help to understand how an organism perceives and responds to any stress condition. Abiotic stresses are defined as the negative impact of any non-living factors on the living organisms in a specific environment. Abiotic stress, such as drought, high salinity, extreme temperature and flooding is a major cause of crop loss worldwide, reducing average yield for most major crop plants by more than 50% (Ouyang et al., 2007). Plants have the ability to adapt to changes in their environment. The sensing of these changes and the subsequent acclimation to the environment follow a general signal transduction pathway. The signalling pathway is initiated by sensors, which detect the stress and then relay the signal through secondary signalling molecules thereby initiating a phosphorylation cascade and activating transcription factors. Activated transcription factors, in turn, regulate gene expression forming the primary response of the plant that results in the protection and repair of the cell (Rensink et al., 2005). Understanding the mechanisms involved in the response of plants to adverse environmental conditions will help to generate crops with high tolerance to various stresses. The plant family of our interest is solanaceae which is a large family comprising of over 3000 species including many important medicinal plants. It is the third most economically important plant family and ranks first in terms of vegetable crops. In this study, seven genera of this family viz. *Solanum tuberosum*, *S. lycopersicum*, *S. melongena*, *Capsicum annuum*, *Petunia hybrida*, *Nicotiana tabacum*, *N. benthamiana* were undertaken. We utilized the gene expression data already available on these plants in public database. The known and experimentally verified TFBSs were utilized for the prediction of binding sites. The putative TFBSs were searched in the promoter sequences of differentially expressed genes in the above mentioned plants. The turnover of TFBS for corresponding TFs might be the reason for differential expression of genes.

Results

Analysis of differential expression data

A total of 17,453 genes of *Solanum tuberosum* showing differential expression under salt stress condition in seven solanaceous plants were retrieved. We obtained the expression data represented by sample ids provided by the Gene Expression Omnibus (GEO) database corresponding to each plant. Each sample includes the expression values of corresponding genes in successive time periods of salt stress. A sample dataset representing the expression values of 17453 genes at six time periods of salt stress is listed in Table 1. The mean expression values for corresponding genes were analyzed for each plant. Filtering was done to identify highly up and down expression of genes from the retrieved data. A cut off intensity value of 0.322 was used to filter the genes. The genes showing intensity value more than 0.322 and less than -0.322 were taken as highly up and down regulated, respectively in response to salt stress. The number of genes obtained after filtering for each plants has been listed (Table 2). These expressed genes of seven plants were further taken for finding the common genes, showing differential expression in all the seven plants. Forty-three genes were found responsible for change in expression in all seven plants.

Sequence assembly analysis of differentially expressed genes

The microarray data utilized in this study were obtained from GEO database of NCBI. The researchers who submitted this data in GEO database have designed the experiment on ESTs of *Solanum tuberosum* taken as probes for obtaining the signals of the intensity values. The intensity values were of seven solanaceous plants at different time periods of salt stress. The EST sequences of 43 genes of *Solanum tuberosum* were retrieved and assembled into 10 contigs. The contigs were then analyzed for the prediction of TFBSs.

Identification of upstream and promoter sequences in differentially expressed genes

The genome data of an organism could be used for finding the putative TFBSs (Yamamoto et al., 2011). The plants of solanaceae family undertaken in the study have not been fully sequenced yet. Therefore, comparative genomics approach was applied for finding the upstream regions of the expressed genes. The genome of *Solanum tuberosum* is not available until date. *Solanum phureja* is a cultivar of *Solanum tuberosum* with more than 95% similarity (Xu et al., 2011). Therefore, for obtaining upstream regions for our genes, genome of *Solanum phureja*, a cultivar of potato (*Solanum phureja*) was used in the study. *Solanum phureja* is one of the eight cultivar groups of *Solanum tuberosum*, commonly known as the *Solanum tuberosum* L. Phureja group. The contings, representing a gene, were aligned with genome of *Solanum phureja* to find regions, similar to contigs. The length of 5kb upstream of the matching segments in genome was considered as upstream sequences. The promoters are the protein (TF) binding sites present at the upstream region of the gene and considered to be a key player in gene expression regulation (Dieterich et al., 2005). The upstream sequences from ESTs of *Solanum tuberosum* were analyzed for promoter prediction. The promoters having maximum threshold were selected from upstream region for each contig. All the promoters had the threshold score of 0.91. The identified promoter sequences were of 50 bp length. The predicted promoters have a single alphabet in bold, representing the TSS (transcription start site). The promoters were further taken for finding putative transcription factor binding sites for a given set of transcription factors.

Prediction of TFs responsible for the differential expression of genes

TFBSs play an important role in the regulation of genes by TFs for signal transduction in the stress responses. The study of these short DNA sequences is crucial for the understanding of how the gene expression is regulated. There is a scarcity of information on regulatory elements due to the unavailability of genome sequences of considered plants. The experimental validations have proved the involvement of different transcription factors in the responsiveness of genes (Sundar et al., 2008). TFs and their corresponding known TFBSs that were available for seven plants under study were searched from available literatures and databases (Table 3).

In silico prediction of TFBSs in differentially expressed genes

Over several years, scientists have tried to find different methods for the prediction of TFBSs.

Table 1. Gene expression data of *Solanum tuberosum* at different time periods of salt stress.

GENE	GSM201887	GSM201888	GSM201889	GSM201890	GSM201891	GSM201891
BQ118565	0.176	-0.786	-0.8745	-0.8105	1.042	1.042
BQ117934	-0.182	-0.045	0.042	-0.1055	0.015	0.015
BQ117390	0.2075	0.316	-0.2255	0.333	-0.0345	-0.0345
BQ505987	-0.135	-0.3485	0.3275	-0.208	-0.096	-0.096
BQ513942	0.3785	0.254	0.5205	-0.234	0.556	0.556
BQ505287	-0.0145	0.413	-0.0065	0.2985	-0.043	-0.043
BQ504764	-0.0245	0.006	0.1045	0.1695	-0.1775	-0.1775
BQ515458	0.109	0.0825	0.2505	0.152	0.17	0.17
BQ113878	-0.1565	-0.2515	-0.3735	-0.1675	-0.0215	-0.0215
BQ508397	-0.0655	-0.147	0.0625	-0.159	-0.0815	-0.0815
BQ120265	0.029	0.069	0.1165	0.1375	-0.046	-0.046
BQ509973	0.384	0.0285	0.156	0.1215	0.298	0.298
BQ516135	0.138	-0.208	0.103	-0.1195	-0.243	-0.243

The first column shows list of genes of *Solanum tuberosum* representing the GenBank ids of considered genes (ESTs). The column 2-6 represents the differential expression values of genes of the same row at different time period of the salt stress. GSM201887, GSM201888, GSM201889, GSM201890, GSM201891 and GSM201892 are the sample ids provided by the GEO database representing the samples at 0, 2, 12, 24, 48 and 96 hr of time periods. The successive rows of column 2-6 consist of the expression values. The positive and negative values depicted the up and down expression of the gene at particular time period.

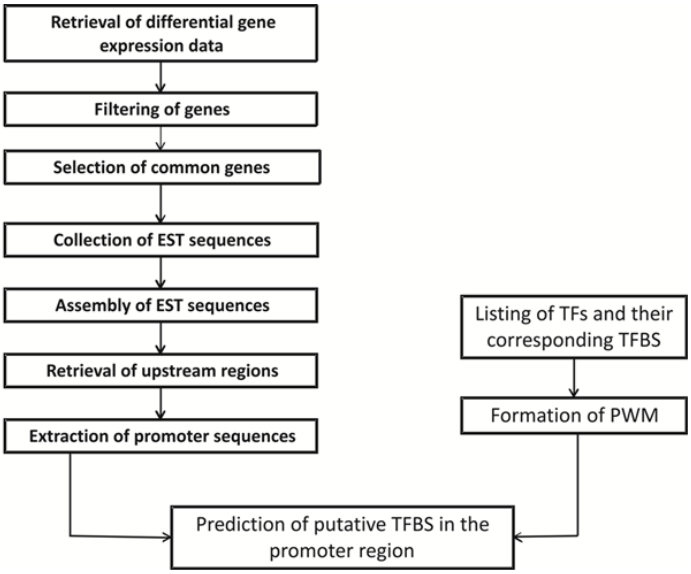

Fig 1. Schematic pipeline of the computational workflow involved in TFBS prediction in differentially expressed genes of *Solanum tuberosum* under different time periods of salt stress.

Earlier experimental methods were used for this purpose but they proved to be quite time-consuming. Fast and efficient computational methods for modelling and identification of DNA regulatory elements have been developed over the past two and a half decades (GuhaThakurta, 2006). The putative TFBSs have been previously identified using different tools and softwares (Arnold et al., 2012; Gonzalez et al., 2012). In our study, TFBSs corresponding to TFs of solanaceous plants were identified, as they may be the key regulators for overall response to the stress treatment. The known TFBSs were used as training dataset for the prediction of new TFBSs in promoters of expressed genEs. Position Weight Matrices (PWMs) were used for this prediction. The RSAT-consensus tool was used to convert the known TFBSs into consensus sequences. The consensus sequences were then taken by RSAT-convert matrix to form PWMs of known TFBSs. The resulted PWMs were validated through D-matrix tool. D-matrix also constructed the PWMs for those TFBSs, that did not contained a specific nucleotide at one or more position like - AAAAAACG/CGTTA and also for those TFBSs that contained an IUPAC code at one or more position of the nucleotide like – GNATATNC (Supplementary Table 1). All the promoter sequences of genes, differentially expressed under salt stress have been scanned by the PWMs. The program RSAT-patser was used to obtain a list of putative TFBSs for corresponding TFs. Score was calculated as per consensus algorithm in RSAT using prior nucleotide frequency as A:T 0.2 and G:C 0.3. Most of the results were obtained above the threshold score value of 2.97 (Supplementary Table 2). Multiple TFBSs have been predicted for each gene corresponding to the TF.

Discussion

In this study, we performed an *in silico* analysis of differential expression data of seven solanaceous plants under salt stress at different time periods and one salt concentration. The data taken for the present *in silico* analysis was retrieved from GEO database of NCBI. Similar reports are already available for potato (Rensink et al., 2005) and barley (Walia et al., 2006). The mean of expression values from all time periods was analyzed representing the average expression of

Table 2. Plant wise number of genes left after filtering.

Sr. no.	Plants	No. of genes
1.	*Solanum tuberosum*	1015
2.	*Solanum lycopersicum*	879
3.	*Solanum melongena*	1464
4.	*Capsicum annuum*	1024
5.	*Petunia hybrida*	477
6.	*Nicotiana tabacum*	1339
7.	*Nicotiana benthamiana*	1460

Table 3. TFs and their experimentally validated TFBSs plants of solanaceae family.

Sn.	TFs	TFBSs	References
Capsicum annuum			
1.	CaWRKYb	T/TGAC/C	Lim et al., 2011
2.	CaPF1	CCGAC	Yi et al., 2004
3.	CaERFLP1	TAAGAGCCGCC	Jae-Hoon Lee et al., 2004
4.	Ca-DREBLP1	TACCGACAT	Hong and Kim, 2005
Nicotiana benthamiana			
1.	bHLH	GCACGTTG	Todd et al., 2010
2.	WRKY8	TTGACC/T	Ishihama et al., 2011
3.	AP2/ERF	GCAGGCC	Andriankaja et al., 2007
Nicotiana tabacum			
1.	bZIP	TGACGTCA	Schiermeyer et al., 2003
2.	WRKY	TTGAC	Yoda et al., 2002
3.	JERF	AGCCGCC	Zhang et al., 2004
4.	NtWRKY12	TTTTCCAC	Marcel et al., 2011
5.	EREBP/AP2	TAAGAGCCGCC	Park et al., 2001
6.	MYC2	GCACGTTG	Zhang et al., 2012
Petunia hybrida			
1.	MYB.Ph3	AAAAAACG	Solano et al., 1995
Solanum lycopersicum			
1.	SlRAV2	CAACA	Li et al., 2011
2.	VSF-1	GCTCCGT	Ringli and Keller, 1998
3.	LpWRKY1	TTTGACT/C	Hofmann et al., 2008
Solanum melongena			
1.	SmCP	CACGTG	Xu et al., 2003
2.	MYCS	TTTCTTGTTT	Chen et al., 2011
3.	P1BS	GNATATNC	Chen et al., 2011
Solanum tuberosum			
1.	*St-WRKY1*	TGAC	Dellagi et al., 2000
2.	StEREBP1	AGCCGCC CCGAC	Lee et al., 2007
3.	StWhy1	GTCAAAAA/T	Desveaux et al., 2004
4.	*StMYB1R-1*	G/AGATAA	Shin et al., 2011

all the samples at 0, 6, 12, 24, 48 and 96 hours of the salt stress for each plant species separately. A cut off $\log_2$ value of expression ratio, 0.322 was selected to filter out the significant genes as at this cut off value, we obtained maximum number of significant genes. Walia et al., 2006 has earlier used 0.585 as cut off $\log_2$ value of expression ratio to get the significant genes. Our analysis revealed that *Solanum melongena* showed maximum number of significant genes, i.e., 1464 followed by *Nicotiana tabacum*. This indicated that *Solanum melongena* showed higher response towards salt stress in comparison to other six plants under study. Our analysis further identified 43 common genes which were found in all the seven plants simultaneously. These 43 common genes might be responsible for differential expression in all the seven plants and can be used for development of salt tolerant plant varieties. We used the genome data of *Solanum phureja* to identify the promoter sequence as the genome of *Solanum tuberosum* is currently not available. Genome of *Solanum phureja* and *Solanum tuberosum* has 95% similarity (Xu et al., 2011). Previous workers have also used the genome data of the wild grass *Brachypodium distachyon* for the analysis of transcription factors of economically important pooideae grasses, including wheat and barley (Mochida et al., 2011) as the genomes of those plants were not available at that time.

We further predicted TFBSs through *in silico* method. The promoter sequences are present in upstream region of any gene. These sequences are the locations where a transcription factor binds to regulate the expression of genes. Position Weight Matrices (PWMs) were used to predict the TFBSs. The use of PWMs for finding the putative TFBSs in the promotes have earlier been reported (Grau et al., 2006). The predicted TFBSs were further analyzed for corresponding transcription factors, which might be responsible for change in the expression of genes in stress condition. Our results revealed 3-5 TFBSs for each transcription factor. Retrieved

references on available TFs and TFBSs are listed for all seven plants separately (Table 3). WRKYb, TF of *Capsicum annuum* has been reported to recognize Wbox element, T/TGAC/C present in pathogenesis related genes (Lim et al., 2011). WRKY proteins also regulate the expression of biotic (Chujoet al., 2009) and abiotic (Pandey and Somssich, 2009) stresses of genes in other plants. The CRT/DRE or GCC box has been found to be responsible for binding of ERF/AP2-type transcription factor (CaPF1) (Yi et al., 2004). The ethylene-responsive factor like protein 1 (CaERFLP1) (Jae-Hoon et al., 2004) was found to bind in GCC box where as dehydration-responsive element binding-factor-like protein 1(Ca-DREBLP1) (Hong and Kim, 2005) binds at CRT/DRE box in *Capsicum annuum*. The bHLH play a role in regulation of nicotine biosynthesis by binding to G-box of promoter of putrescine N-methyltransferase gene (Todd et al., 2010). WRKY8 TF binds on Wbox response to expression of defense related genes in *Nicotiana benthamiana*. AP2/ERF binds at NF box. These known TFBSs were used as training dataset for the prediction of new TFBS in the genes expressed under salt stress. The bZIP TF of *Nicotiana tabacum* regulates the signal transduction pathways. WRKY of *Nicotiana tabacum* regulates the transcriptional activation of defense-related genes (Yoda et al., 2002). NtWRKY12 is a variant of WRKY protein and EREBP/AP2 TF that activate pathogenesis related genes by binding to WK box (Marcel et al., 2011) and GCC box (Park et al., 2001) respectively. MYC2 is responsible for the regulation of jasmonic acid pathway by binding to G-box (Zhang et al., 2012). VSF-1 TF is a vascular specific protein, a type of bZIP TF that interacts with vs-1 element controls the vascular gene expression (Ringli and Keller, 1998). LpWRKY1 binds at W-box of genes involve in defense response (Hofmann et al., 2008). SmCP is cysteine protenase involve in regulation of expression of genes involved in developmental events in *Solanum melongena*. StMYB1R-1 is a MYB like TF regulate the tolerance of drought stress in *Solanum tuberosum* (Shin et al., 2011). The predicted TFBSs differed from experimentally validated TFBSs for corresponding TFs indicating that the turnover of binding site may be responsible for differential expression of genes. Thus, the binding of transcription factor to different TFBSs at different time periods of salt stress mey be the reason for the differential expression of genes. Further studies on these predicted TFBSs will provide a better insight into the role of turnover of TFBSs responsible for the change in expression of genes.

Materials and Methods

Retrieval of gene expression data

The gene expression data showing differential expression under different time periods of salt stress was retrieved from GEO (http://www.ncbi.nlm.nih.gov/geo/) database of NCBI (Barrett et al., 2005). The expression data of our interest were obtained from the series id GSE8158 in the database. As per experimental details available in database, ESTs of Solanum tuberosum has been used as spotted PCR amplified cDNA array on glass. The researchers have subjected seven different Solanaceae species, Potato (*Solanum tubersosum*), Tomato (*Solanum lycopersicum*), Eggplant (*Solanum melongena*), Pepper (*Capsicum annuum*), Tobacco (*Nicotiana tabaccum*), *Petunia hybrida* and *Nicotiana benthamiana* to salt stress. The experimental conditions included the application of one salt concentration of 150mM NaCl and the control plants were watered without the additional salt. Samples were collected at 0, 6, 12, 24, 48 and 96 hours after the first application of the salt. The RNA of theses experimental plants were isolated and microarray hybridization was performed. The expression profiling was performed by microarray technique. We retrieved expression values for seven solanaceous species viz. potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*), eggplant *(Solanum melangena*), pepper (*Capsicum annuum*), tobacco (*Nicotiana tabaccum*), petunia (*Petunia hybrida*) and *Nicotiana benthiamana.* The expression values of 17,453 genes showing differential expression in all the seven plants were obtained.

Sequence assembly

The genes were filtered with cutoff of mean intensity values 0.322 in each plant. The cutoff of mean intensity value of 0.322 was selected because at this value maximum number of highly up and down expression of genes in all the seven plants was obtained. The common genes were selected from all the seven plants (Table 2). These common genes were characterized for their response to salinity stress in all plants. The expressed sequence tags (ESTs) corresponding to the common genes were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) database at NCBI. The EST sequences were assembled to form the contigs using CAP3 sequence assembly (http://pbil.univ-lyon1.fr/cap3.php) program (Huang and Madan, 1999).

Upstream sequence mining

After sequence assembly, the contigs were used to find the upstream regions. For finding the upstream regions, the genome of *Solanum phureja*, a cultivar of *Solanum tuberosum* was used. 5kb upstream of the contig sequence, showing an exact alignment was taken as the upstream regions.

Prediction of promoters

The Neural Network Promoter Prediction (NNPP) (http://www.fruitfly.org/seq_tools/promoter.html) tool was used for searching promoter regions in the upstream sequences (Reese, 2001). A threshold value of 0.9 was used for promoter finding which lies between 0.1 to 1. The transcription start sites are also analyzed in the given upstream sequence.

Literature search for known TFs & TFBSs

As the analysis of putative TFBSs was based on experimentally verified TFBSs, the collection and listing of the experimentally validated TFs and their corresponding TFBSs that are present either in a database or in literature was done. The TFs of seven species of solanaceae family were searched in literature as well as in databases like TRANSFAC (http://www.gene-regulation.com/pub/databases.html) (Wingender et al., 1996), JASPER (http://jaspar.genereg.net/) (Sandelin et al., 2004) and PLANT TFDB (http://planttfdb.cbi.edu.cn/) (Zhang et al., 2010) etc.

Formation of PWMs

The TFBSs retrieved for seven species of solanaceae family were used to form PWMs. The TFBSs were first converted into consensus sequence using the Regulatory Sequence Analysis Tool (RSAT-consensus) (http://rsat.ulb.ac.be/

consensus_form.cgi) (Thomas-Chollier et al., 2011; van Helden et al., 1998). The resulting consensus sequences were used as input to the RSAT-convert matrix (http://rsat.ulb.ac.be/convert-matrix_form.cgi) (Thomas-Chollier et al., 2011; van Helden et al., 1998). RSAT-convert matrix converts the consensus sequences into position weight matrices (PWMs). The resulted PWMs were cross validated by the program D-matrix (http://203.190.147.116/dmatrix/home.aspx) (Sen et al., 2009).

Prediction of putative TFBSs

The analyzed promoter sequences from differentially expressed genes were used for prediction of TFBSs. RSAT-patser tool (http://rsat.ulb.ac.be/patser_form.cgi) was used to scan these promoter sequences by PWMs generated from known TFBSs (Thomas-Chollier et al., 2011; van Helden et al., 1998). Multiple putative TFBSs for corresponding TFs were obtained. The workflow of all steps involved in the study has been shown in Fig1.

Acknowledgements

We would like to thanks to C Robin Buell who have submitted her experiments in GEO database of NCBI and make them freely available to the scientific community. One of the authors, (Sanchita) is thankful to CSIR, New Delhi, India for CSIR-SRF fellowship.

Supplementary data online only

Supplementary Table 1. PWMs for the known TFs and TFBSs.
Supplementary Table 2. Details of the predicted TFBSs in the promoter regions based on the PWMs having their position and scores.

References

Andriankaja A, Boisson-Dernier A, Frances L, Sauviac L, Jauneau A, Barker DG, de Carvalho-Niebel F (2007) AP2-ERF transcription factors mediate Nod factor dependent Mt ENOD11 activation in root hairs via a novel cis-regulatory motif. Plant Cell 19 (9):2866-2885

Arnold P, Erb I, Pachkov M, Molina N, van Nimwegen E (2012) MotEvo: integrated Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences. Bioinformatics 28 (4):487-494

Babu M (2004) An introduction to microarray data analysis. Computational Genomics, Horizon Press, Richard P. Grant Laboratory of Molecular Biology, Cambridge, UK

Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R (2005) NCBI GEO: mining millions of expression profiles—database and tools. Nucleic Acids Res. 33 (1): D562-D566

Chen AQ, Gu MA, Sun SB, Zhu LL, Hong SA, Xu GH (2011) Identification of two conserved cis-acting elements, MYCS and P1BS, involved in the regulation of mycorrhiza-activated phosphate transporters in eudicot species. New Phytol 189 (4):1157-1169

Chujo T, Sugioka N, Masuda Y, Shibuya N, Takemura T, Okada K, Nojiri H, Yamane H (2009) Promoter analysis of the elicitor-induced WRKY gene OsWRKY53, which is involved in defense responses in rice. Biosci Biotech Bioch 73 (8):1901-1904

Dellagi A, Heilbronn J, Avrova AO, Montesano M, Palva ET, Stewart HE, Toth IK, Cooke DEL, Lyon GD, Birch PRJ (2000) A potato gene encoding a WRKY-like transcription factor is induced in interactions with *Erwinia carotovora* subsp atroseptica and *Phytophthora infestans* and is coregulated with class I endochitinase expression. Mol Plant Microbe In 13 (10):1092-1101

Desveaux D, Subramaniam R, Despres C, Mess JN, Levesque C, Fobert PR, Dangl JL, Brisson N (2004) A "whirly" transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*. Dev Cell 6 (2):229-240

Dieterich C, Grossmann S, Tanzer A, Ropcke S, Arndt PF, Stadler PF, Vingron M (2005) Comparative promoter region analysis powered by CORG. BMC Genomics 6:24

Dudoit S, Yang YH, Callow MJ, Speed TP (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Stat Sinica 12 (1):111-13

Gonzalez S, Montserrat-Sentis B, Sanchez F, Puiggros M, Blanco E, Ramirez A, Torrents D (2012) ReLA, a local alignment search tool for the identification of distal and proximal gene regulatory regions and their conserved transcription factor binding sites. Bioinformatics 28 (6):763-770

Grau J, Ben-Gal I, Posch S, Grosse I (2006) VOMBAT: prediction of transcription factor binding sites using variable order Bayesian trees. Nucleic Acids Res 1:34 (Web Server issue):W529-33.

GuhaThakurta D (2006) Computational identification of transcriptional regulatory elements in DNA sequence. Nucleic Acids Res 34 (12):3585-3598

Hofmann MG, Sinha AK, Proels RK, Roitsch T (2008) Cloning and characterization of a novel LpWRKY1 transcription factor in tomato. Plant Physiol Biochem 46 (5-6):533-540

Hong JP, Kim WT (2005) Isolation and functional characterization of the Ca-DREBLP1 gene encoding a dehydration-responsive element binding-factor-like protein 1 in hot pepper (*Capsicum annuum* L. cv. Pukang). Planta 220 (6):875-888

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9(9):868-77

Ishihama N, Yamada R, Yoshioka M, Katou S, Yoshioka H (2011) Phosphorylation of the *Nicotiana benthamiana* WRKY8 Transcription Factor by MAPK Functions in the Defense Response. Plant Cell 23 (3):1153-1170

Jae-Hoon Lee J-PH, Sang-Keun Oh, Sanghyeob Lee, Doil Choi and Woo, Kim T (2004) The ethylene-responsive factor like protein 1 (CaERFLP1) of hot pepper (*Capsicum annuum* L.) interacts *in vitro* with both GCC and DRE/CRT sequences with different binding affinities: Possible biological roles of CaERFLP1 in response to pathogen infection and high salinity conditions in transgenic tobacco plants. Plant Mol Biol 55:61-81

Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci USA 94 (24):13057-13062

Latchman DS (1997) Transcription factors: an overview. Int J Biochem Cell Biol 29 (12):1305-1312

Lee HE, Shin D, Park SR, Han SE, Jeong MJ, Kwon TR, Lee SK, Park SC, Yi BY, Kwon HB, Byun MO (2007) Ethylene responsive element binding protein 1 (StEREBP1) from *Solanum tuberosum* increases tolerance to abiotic

stress in transgenic potato plants. Biochem Biophys Res Commun 353 (4):863-868

Li CW, Su RC, Cheng CP, Sanjaya, You SJ, Hsieh TH, Chao TC, Chan MT (2011) Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. Plant Physiol 156 (1):213-227

Lim JH, Park CJ, Huh SU, Choi LM, Lee GJ, Kim YJ, Paek KH (2011) *Capsicum annuum* WRKYb transcription factor that binds to the CaPR-10 promoter functions as a positive regulator in innate immunity upon TMV infection. Biochem Bioph Res Co 411 (3):613-619

van Verk MC, Neeleman L, Bol JF, and Linthorst HJM (2011) Tobacco transcription factors NtWRKY12 and TGA2.2 interact *in vitro* and *in vivo*. Front Plant Sci 2 (32):1-10

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LSP (2011) *In Silico* analysis of transcription factor repertoires and prediction of stress-responsive transcription factors from six major gramineae plants. DNA Res 18 (5):321-332

Ouyang B, Yang T, Li H, Zhang L, Zhang Y, Zhang J, Fei Z, Ye Z (2007) Identification of early salt stress response genes in tomato root by suppression subtractive hybridization and microarray analysis. J Exp Bot 58 (3):507-520

Pandey SP, Somssich IE (2009) The Role of WRKY Transcription Factors in Plant Immunity. Plant Physiol 150 (4):1648-1655

Park JM, Park CJ, Lee SB, Ham BK, Shin R, Paek KH (2001) Overexpression of the tobacco Tsi1 gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. Plant Cell 13 (5):1035-1046

Rensink WA, Iobst S, Hart A, Stegalkina S, Liu J, Buell CR (2005) Gene expression profiling of potato responses to cold, heat, and salt stress. Funct Integr Genomics 5 (4):201-207

Reese MG (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput Chem 26(1):51-56.

Ringli C, Keller B (1998) Specific interaction of the tomato bZIP transcription factor VSF-1 with a non-palindromic DNA sequence that controls vascular gene expression. Plant Mol Biol 37 (6):977-988

Romeuf I, Tessier D, Dardevet M, Branlard G, Charmet G, Ravel C (2010) wDBTF: An integrated database resource for studying wheat transcription factor families. BMC Genomics 11:185

Sandelin A, Alkema W, Engström P, Wasserman1 WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32 (1): D91-D94

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270 (5235):467-470

Schiermeyer A, Thurow C, Gatz C (2003) Tobacco bZIP factor TGA10 is a novel member of the TGA family of transcription factors. Plant Mol Biol 51 (6):817-829

Sen N, Mishra M, Khan F, Meena A, Sharma A (2009) D-MATRIX: a web tool for constructing weight matrix of conserved DNA motifs. Bioinformation 3 (10):415-418

Shin D, Moon SJ, Han S, Kim BG, Park SR, Lee SK, Yoon HJ, Lee HE, Kwon HB, Baek D, Yi BY, Byun MO (2011) Expression of StMYB1R-1, a novel potato single MYB-like domain transcription factor, increases drought tolerance. Plant Physiol 155 (1):421-432

Siddharthan R (2010) Dinucleotide weight matrices for predicting transcription factor binding sites: generalizing the position weight matrix. PLoS One 5 (3):e9722

Sinha S, Tompa M (2002) Discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res 30 (24):5549-5560

Solano R, Nieto C, Paz-Ares J (1995) MYB.Ph3 transcription factor from *Petunia hybrida* induces similar DNA-bending/distortions on its two types of binding site. Plant J 8 (5):673-682

Sreekumar J, Jose KK (2008) Statistical tests for identification of differentially expressed genes in cDNA microarray experiments. Indian J Biotechnol 7 (4):423-436

Sundar AS, Varghese SM, Shameer K, Karaba N, Udayakumar M, Sowdhamini R (2008) STIF: Identification of stress-upregulated transcription factor binding sites in *Arabidopsis thaliana*. Bioinformation 2 (10):431-437

Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J (2011) RSAT 2011: regulatory sequence analysis tools. Nucleic Acids Res 39:W86-W91

Todd AT, Liu EW, Polvi SL, Pammett RT, Page JE (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. Plant J 62 (4):589-600

van Helden J, Andre B, Collado-Vides J (1998) Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. J Mol Biol 281 (5):827-842

Walia H, Wilson C, Wahid A, Condamine P, Cui X, Close TJ (2006) Expression analysis of barley (*Hordeum vulgare* L.) during salinity stress. Funct Integr Genomics. 6 (2):143-56

Wingender E, Dietze P, Karas H, Knüppel R (1996) TRANSFAC: A Database on Transcription Factors and Their DNA Binding Sites. Nucleic Acids Res 24 (1):238-241.

Won KJ, Ren B, Wang W (2010) Genome-wide prediction of transcription factor binding sites using an integrated model. Genome Biol 11:R7

Xu X, Pan SK, Cheng SF, Zhang B, Mu DS, Ni PX et al. (2011) Genome sequence and analysis of the tuber crop potato. Nature 475 (7355):189-U194

Xu ZF, Chye ML, Li HY, Xu FX, Yao KM (2003) G-box binding coincides with increased *Solanum melongena* cysteine proteinase expression in senescent fruits and circadian-regulated leaves. Plant Mol Biol 51 (1):9-19

Yamamoto YY, Yoshioka Y, Hyakumachi M, Maruyama K, Yamaguchi-Shinozaki K, Tokizawa M, Koyama H (2011) Prediction of transcriptional regulatory elements for plant hormone responses based on microarray data. BMC Plant Biol 11:39

Yang YH, Xiao YY, Segal MR (2005) Identifying differentially expressed genes from microarray experiments via statistic synthesis. Bioinformatics 21 (7):1084-1093

Yi SY, Kim JH, Joung YH, Lee S, Kim WT, Yu SH, Choi D (2004) The pepper transcription factor CaPF1 confers pathogen and freezing tolerance in *Arabidopsis*. Plant Physiol 136 (1):2862-2874

Yoda H, Ogawa M, Yamaguchi Y, Koizumi N, Kusano T, Sano H (2002) Identification of early-responsive genes associated with the hypersensitive response to tobacco mosaic virus and characterization of a WRKY-type transcription factor in tobacco plants. Mol Genet Genomics 267 (2):154-161

Zhang H, Huang Z, Xie B, Chen Q, Tian X, Zhang X, Lu X, Huang D, Huang R (2004) The ethylene-, jasmonate-, abscisic acid- and NaCl-responsive tomato transcription factor JERF1 modulates expression of GCC box-containing genes and salt tolerance in tobacco. Planta 220 (2):262-270

Zhang HB, Bokowiec MT, Rushton PJ, Han SC, Timko MP (2012) Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. Mol Plant 5 (1):73-84

Zhang H, Jin J, Tang L, Zhao YI, Gu X, Gao G, Luo J (2010) PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res 2010, 1–4