

RESEARCH ARTICLE

Open Access



Soybean transcription factor ORFeome associated with drought resistance: a valuable resource to accelerate research on abiotic stress resistance

Chenglin Chai^{1†}, Yongqin Wang^{1†}, Trupti Joshi², Babu Valliyodan¹, Silvas Prince¹, Lydia Michel¹, Dong Xu² and Henry T. Nguyen^{1*}

Abstract

Background: Whole genome sequencing provides the most comprehensive collection of an organism's genetic information. The availability of complete genome sequences is expected to dramatically deliver a high impact on biology. However, to achieve this impact in the area of crop improvement, significant efforts are still required on functional genomics, including the areas of gene annotation, cloning, expression profiling, and functional validation.

Results: Here we report our efforts in generating the first transcription factor (TF) open reading frame (ORF)eome resource associated with drought resistance in soybean (*Glycine max*), a major oil/protein crop grown worldwide. This study provides a highly annotated soybean TF-ORFeome associated with drought resistance. It contains information from experimentally verified protein-coding sequences (CDS), expression profiling under several abiotic stresses (drought, salinity, dehydration and ABA), and computationally predicted protein subcellular localization and *cis*-regulatory elements (CREs) analysis. All the information is available to plant researchers through a freely accessible and user-friendly database, Soybean Knowledge Base (SoyKB).

Conclusions: The soybean TF-ORFeome provides a valuable public resource for functional genomics studies, especially in the area of plant abiotic stresses. It will accelerate findings in the areas of abiotic stresses and lead to the generation of crops with enhanced resistance to multiple stresses.

Keywords: Soybean, Transcription factor, ORFeome, Abiotic stresses, Drought, Salinity, Dehydration, ABA, *cis*-regulatory element

Background

Whole genome sequencing (WGS) provides the most comprehensive collection of an organism's genetic information. Large-scale genome sequencing is expected to change the way in which biology has traditionally been conducted. The ever-decreasing cost of sequencing is moving towards a new era in plant genetic and genomic studies. By taking advantage of large data acquisition platforms, genomes from more than 40 plants of

agronomical importance have been sequenced so far [1]. However, to achieve this promise of WGS in research focused on crop improvement, significant efforts are required in functional genomics that include gene annotation, cloning, expression, and further functional analysis.

Knowledge of gene sequences and of the deduced protein sequences is very important in determining protein functions. In this process, large genomic resources such as expressed sequence tag (EST) databases, full-length complementary DNA (cDNA) libraries, and open reading frame (ORF) collections (ORFeome) have played important roles. Although EST databases and computational

* Correspondence: nguyenhenry@missouri.edu

[†]Equal contributors

¹Division of Plant Sciences and National Center for Soybean Biotechnology, University of Missouri, Columbia, MO 65211, USA

Full list of author information is available at the end of the article

predictions are useful, the EST databases usually provide only partial transcribed sequences that could be misleading, while the automated computational prediction are not fully accurate [2]. Full-length cDNA libraries contain full-ORFs plus 5' and 3' un-translated regions (UTRs), which will allow massive functional screening in various fields of biology. However, the drawback of cDNA libraries has become obvious due to the interference of 5'- and 3'-UTRs, and low coverage of cDNA libraries for total gene transcripts [3]. ORFeome collections not only overcome the problems mentioned above, but also have additional advantages. By using gene-specific primers, genuine full ORFs can be obtained, which assure high coverage and no interference of 5'- or 3'- UTRs. The recombination-based cloning techniques including Gateway cloning [4], have revolutionized the ways of conventional "cut-and-paste" techniques, and greatly expedited high-throughput gene cloning. Furthermore, access to the ORF cDNA clones would facilitate various functional studies of genes and corresponding proteins by transferring ORFs via LR reactions from Entry clones into Gateway-compatible expression vectors [5]. ORFeome resources have been successfully applied in genome annotation, genome-wide protein localization, metabolic structure studies, proteomics, comparative functional genomics, global mapping protein-protein interaction and DNA-protein interactions [6-11]. However, despite all the achievements made so far by plant scientists in building various ORFeomic resources, most existing ORFeomes are too general. This leads to a situation wherein researchers working in a specific area (e.g., drought research) have to spend a significant amount of time finding information in their area of interest.

Soybean (*Glycine max*) is the most important cash crop widely grown for its high protein and oil content, beneficial phytochemicals, and production as biodiesel. However, its growth and grain yield are highly affected by soil water availability. Drought stresses have caused significant yield losses worldwide [12, 13]. Plants respond and adapt to drought stress conditions with an array of molecular, biochemical, and physiological alterations. Despite the fact that the soybean's entire genome was sequenced several years ago [14], the exact transcript structures of the majority of its protein-coding genes remain experimentally unverified. As such, there is an urgent need in the soybean community for ORFeome clones of protein-coding genes. Since TFs are master regulators in controlling many, if not all, of the biological processes such as development, growth, cell division, and responses to environmental stimuli, our efforts in this study are focused on generating the first transcription factor (TF) ORFeome resource associated with drought resistance in soybean. The soybean TF-ORFeome related information has been deposited in the Soybean Knowledge Base (SoyKB) [15-17] and is available to the global research community for comprehensive functional characterization. This will

greatly accelerate findings in the area of drought resistance research.

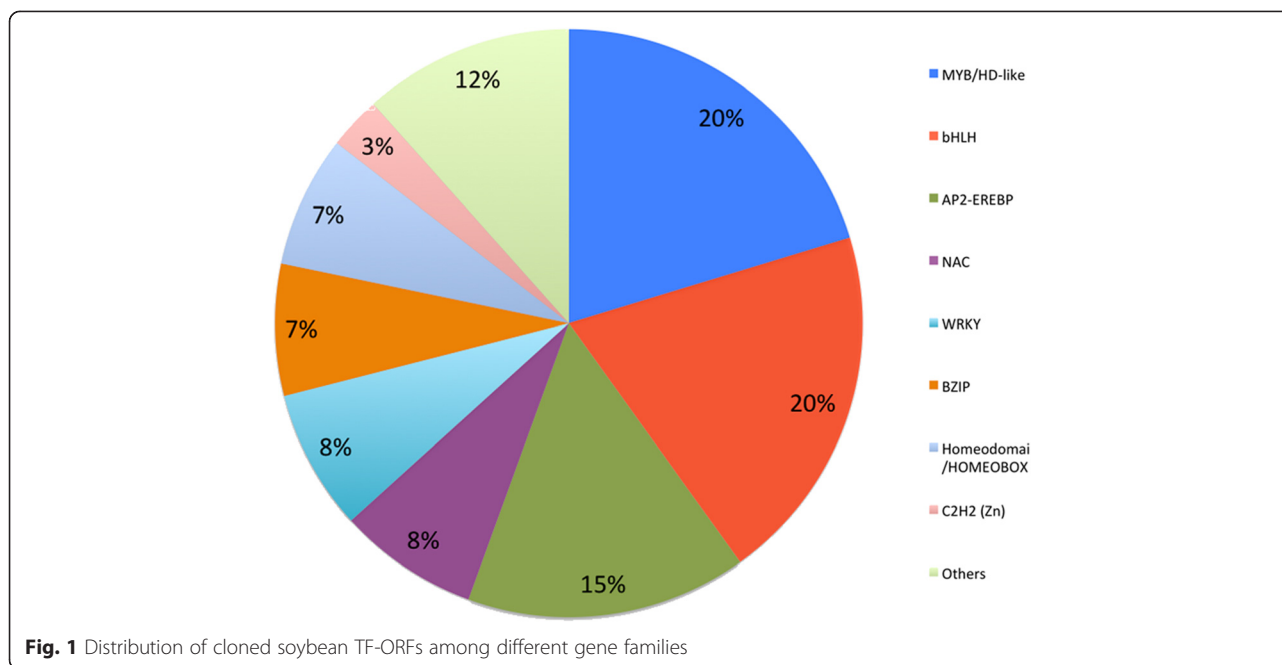
Results and discussion

Soybean TF selection and cloning

Mainly based on microarray data of soybean root and leaf under dehydration and drought conditions generated by our group (Valliyodan et al., unpublished data) and other researchers [18], soybean TFs with a fold change ≥ 1.5 upon treatments and a p -value < 0.05 were selected as candidates for building this soybean TF-ORFeome (Additional file 1). We also included in this TF-ORFeome 19 TFs, which showed a fold change ≥ 1.5 in at least one of the two tissues (shoots and roots) upon mild drought stress in our quantitative reverse transcription-PCR (qRT-PCR) analysis (Additional file 2A) but without support from our microarray data due to lack of probes [18]. A total of 207 soybean full-length TF ORFs were cloned into pENTR™/D TOPO or pDONR™/Zeo vectors, which meets the "gold standard" criteria as previously defined [3]. Detailed information of these clones is provided (Additional file 3), including gene locus number, transcript, GenBank accession number, gene family, gene size, vector, ORF sequence, primer sequences, with/without stop codon, and others. These TFs were not equally distributed among 21 gene families, of which the top seven families are MYB, bHLH, APETALA2 (AP2)-ethylene-responsive element binding protein (EREBP), NAC, WRKY, bZIP and Cys2(C2)His2(H2)-type zinc fingers (ZFs), constituting 88 % of the total clones (Fig. 1). Genes from these TF families were found to play important roles in responding to various abiotic stresses, which was very well summarized in a recent review paper [19]. Several genes from our ORFeome collection were reported as major regulators in the soybean abiotic stress responses, such as *GmbZIP1* (Glyma02g14880) [20], *GmERF3* (Glyma03g42450) [21], *GmDREB2* (Glyma06g04490) [22], and (*GmNAC004* (Glyma12g35000) [23]. However, functions of the vast majority soybean TFs are yet to be explored.

Sequence analysis of cloned TFs

As expected, most (90.3 %) of the soybean TF ORFeome clones matched the gene annotation in the public database Phytozome (v 9.1) [24] based on sequencing results. For clones showing sequence differences, two independent RT-PCRs were performed to make certain that the sequence differences were not caused by errors during RT and/or PCR. At least two clones for each ORFeome were used for sequence verification. However, our sequence analysis revealed differences in 20 clones, 9.7 % of total TF-ORFs cloned in this study (Additional file 3). qRT-PCR analysis of expression changes of these genes upon mild drought stress treatment were conducted



(Additional file 2B), and the results showed that most of them were positively regulated by the stress at the transcriptional level. The sequence differences might be due to alternative splicing, nucleotide replacement, insertion or deletion.

There are several possibilities for the sequence discrepancy in this study. Nearly 75 % of the soybean genes have paralogs, which were probably caused by two whole-genome duplication events that occurred between 59 and 13 million years ago, respectively [14]. Aligning the discrepant sequences back to the soybean genome excluded the possibility that they are one copy of the many duplicated genes, although it is still possible that the duplicated genes are located in the un-sequenced gaps. Another cause of sequence differences of the ORFs might be due to the genomic heterogeneity of Williams 82, which led to the intra-cultivar variations among individuals [25]. However, there is little chance of error from RT-PCR or sequencing due to the stringent conditions set for these processes and the use of multiple clones for sequence verification, as stated above.

Expression profiles of selected TFs from ORFeome collection under drought, dehydration, salt and ABA conditions

Analysis of gene expression in different tissues and under different conditions is a useful way to predict gene functions. By searching the available whole genome profiling data, gene expression profiles of the TFs in 7 soybean tissues/organs (Additional file 4, data are from [26]) and under water deficit conditions (Additional file 1) were collected. Both the tissue expression patterns and the

expression fold changes under water deficit conditions revealed a large amount of variation among different TFs, suggesting their diverse functions during soybean growth, development and adaption to water deficit conditions.

In order to provide more experimental support that the cloned TFs are responsive to water stress, expression profiles of 50 randomly selected soybean TFs (generated by the web tool: Research Randomizer [27]) were evaluated using qRT-PCR under conditions of drought, dehydration and salinity (Fig. 5). Upon drought treatments, 98 % of the selected genes were either up- or down-regulated in one or both of the drought conditions (Figs. 2 and 3). The total number of up- and down-regulated genes in roots was much smaller than in shoots under mild drought conditions (62 % vs. 98 %; Fig. 3a, c), while similar numbers of regulated genes were found under moderate stress conditions (72 % in roots vs. 70 % in shoots; Fig. 3b, d). The same TFs showing different expression levels upon drought treatments in different tissues suggested that they might have varying functions in each tissue in response to drought stress. Overall, our qRT-PCR data further confirmed that the TFs in our ORFeome collection were drought responsive.

Compared to drought, dehydration leads to a much lower water potential, and it is also considered as a common stress induced by drought, extreme temperature, or salinity conditions. Under dehydration treatments (Fig. 4), approximately 50 % of the selected genes were regulated the same way as in the drought conditions, while the expression patterns of other genes were quite different, indicating both pathways share some signaling components

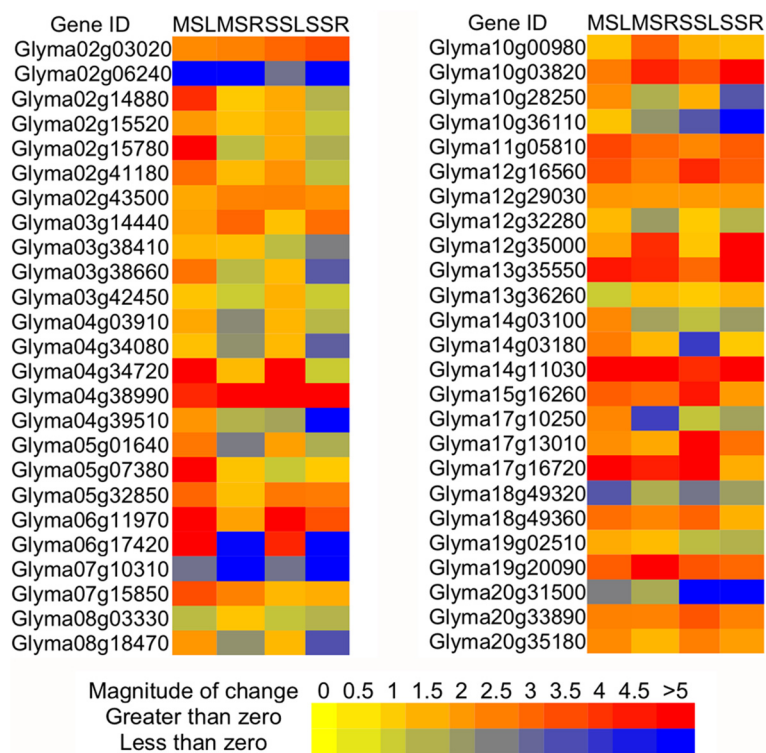


Fig. 2 Expression of randomly selected 50 TFs under mild and moderate drought conditions. MSL, mild drought stress shoots; MSR, mild drought stress roots; SSL, moderate drought stress shoots; SSR, moderate drought stress roots

while remaining relatively independent. Notably, one GmNAC gene, Glyma12g35000.1, showed a dramatic up-regulation upon the dehydration treatment (75 fold change). A very recent study showed that over-expression of Glyma12g35000.1 in *Arabidopsis* enhanced lateral root-

growth under both normal and mild drought stress conditions [23].

Salinity is another abiotic stress that significantly reduces soybean yield, and plant responses to salt stress and drought are very closely related due to their

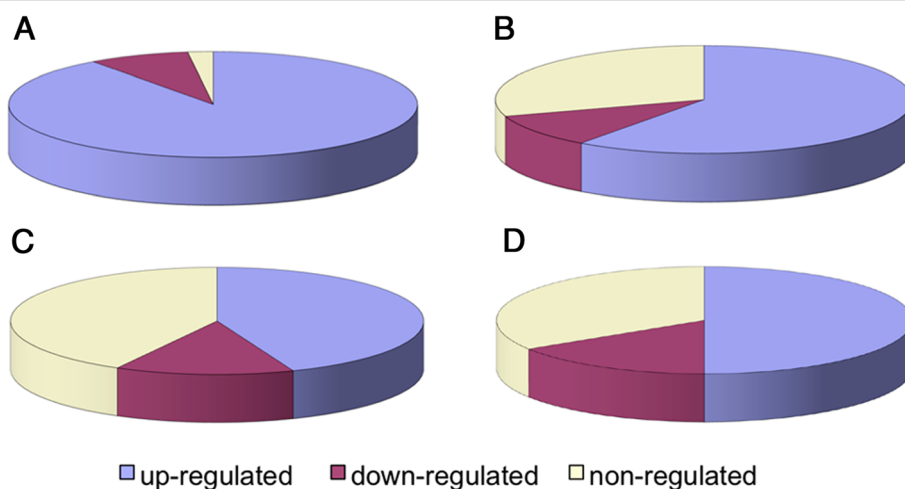
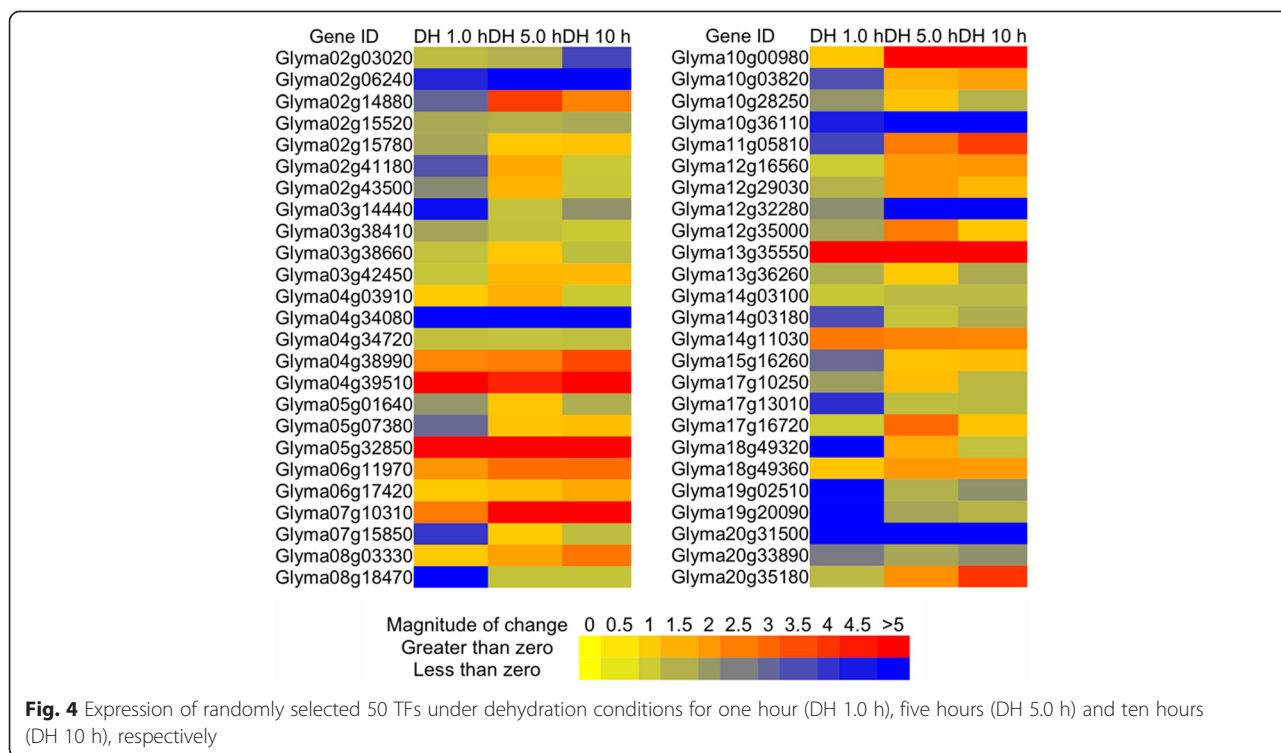
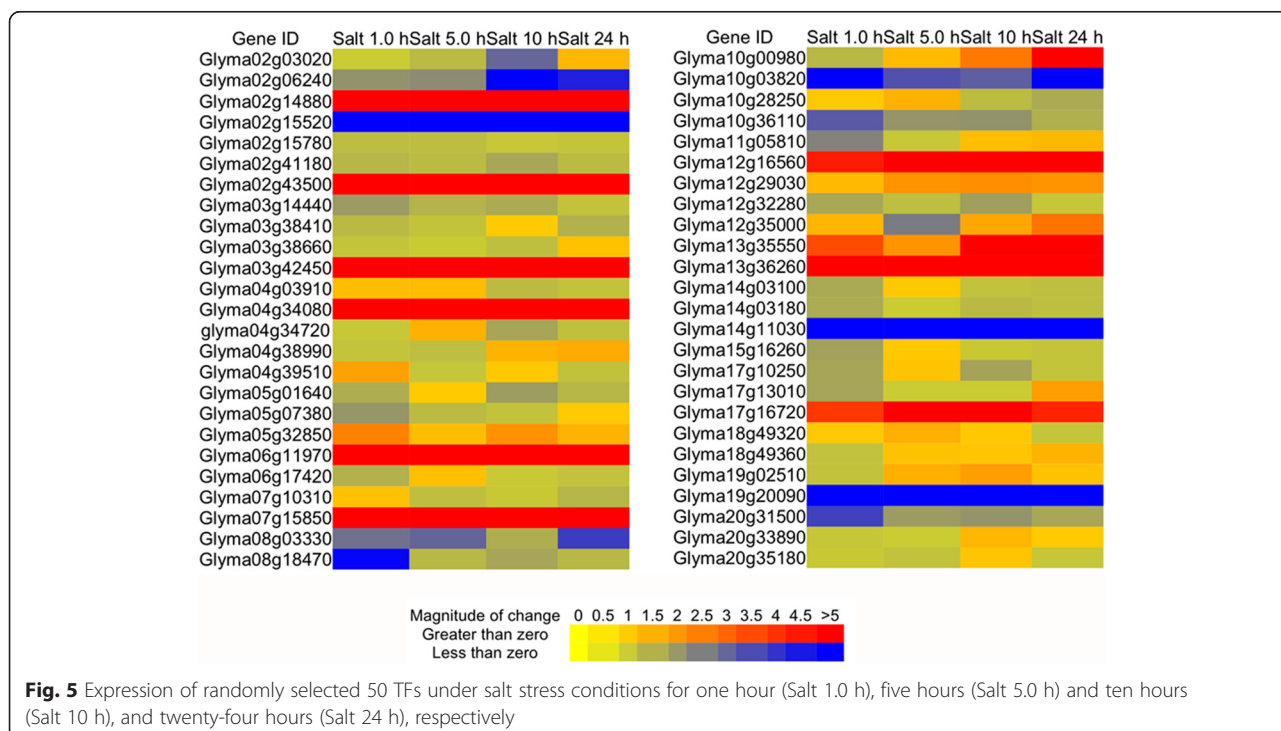


Fig. 3 Soybean TFs up- and down-regulated in mild (a and c) and moderate (b and d) drought stresses. a, mild drought stress shoots; b, moderate drought stress shoots; c, mild drought stress roots; d, moderate drought stress roots



overlapping mechanisms [28, 29]. Surprisingly, about 80 % of the same 50 TFs were differentially expressed upon salt stress (Fig. 5), suggesting their possible role in salt stress adaption and salinity tolerance. While there

was a significant overlap of TFs co-regulated by drought/dehydration/salinity (Figs. 2, 4 and 5), several of them showed opposite expression patterns, such as Glyma19g20090.1, Glyma14g11030.1 and Glyma10g03820.1.



Discrepant expressions under these two conditions suggested their distinct roles in response to different stresses.

The plant hormone abscisic acid (ABA) plays a pivotal role in plant responses to biotic and abiotic stresses [30-32]. When plants are exposed to abiotic stress such as drought and salinity, ABA regulates stomata aperture to limit water loss through transpiration [33]; on the other hand, the localized ABA signaling, by working together with other phytohormones, regulates root growth, especially lateral root growth plasticity [34]. qRT-PCR analysis was performed in order to investigate whether the same set of selected genes are involved in the ABA-dependent signaling pathway (Fig. 6). Expression of the selected genes showed dramatic changes in both roots and shoots under ABA treatments, and 98 % of them showed ≥ 1.5 fold change at one or more of the time points. This result indicated that most of the TFs might function dependently on the ABA signal transduction pathway. More than half of the genes showed a similar expression pattern in roots and shoots, while some other TFs exhibited an opposite pattern in different tissues (such as Glyma05g32850.1 and Glyma20g31500.1), suggesting different roles in shoots and roots (Fig. 6).

Discovery of cis-regulatory elements (CREs) in soybean TF promoters

Although other alternative mechanisms of gene expression regulation exist, the control of gene transcription via

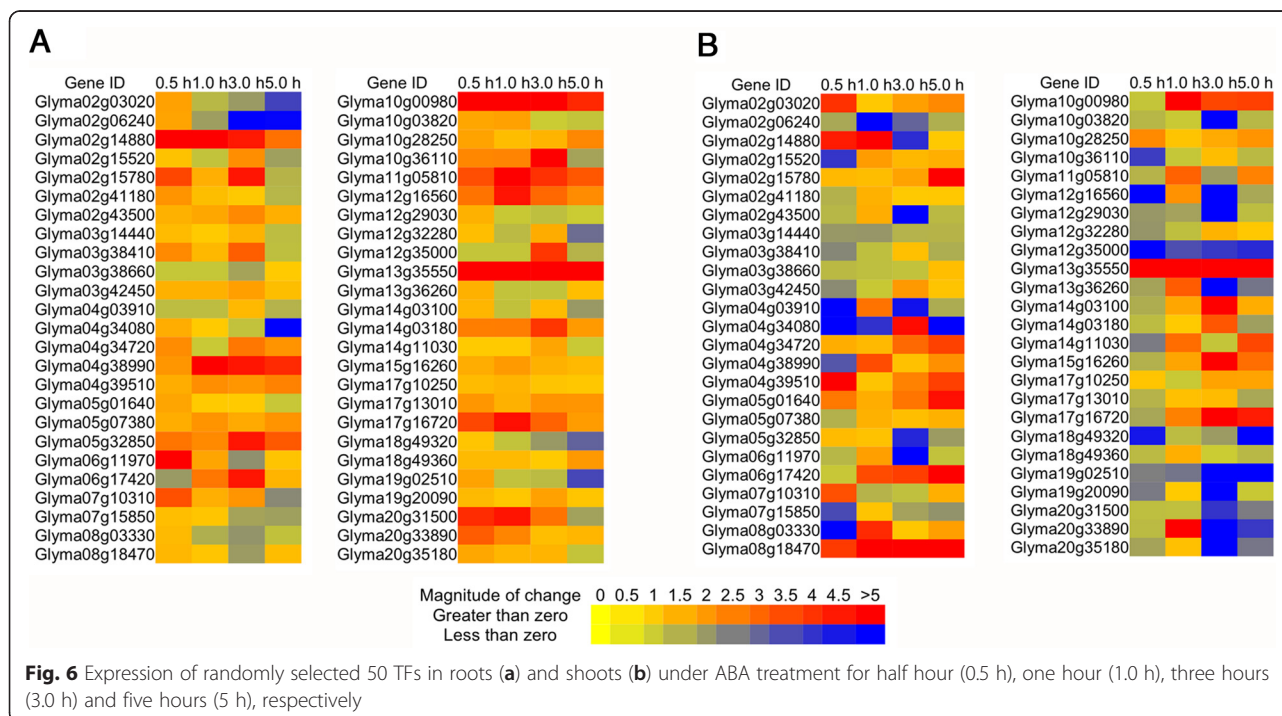
CREs in promoters is still a primary mode of gene expression regulation. Our interest in abiotic stress prompted us to investigate abiotic stress responsive CREs, which may be bound and regulated by other TFs, in the genetic upstream regions in our soybean TF-ORFeome collection. A total of 21 CREs responsive to abiotic stresses were identified among 200 TF promoters (Additional file 3). However, over-representation analysis did not show any of these CREs significantly enriched in the 1 kb promoters of the 200 TFs.

Integration of TF-ORFeome resource into SoyKB website

The TF-ORFeome data has been incorporated into SoyKB [15-17]. The data can be directly accessed via the URL [35] after registration. The genes have been linked to the gene card pages (Additional file 5A, B), where users can access other relevant genomic information (Additional file 5C), and multi-omics expression datasets (Additional file 5D, E) available in SoyKB. The motif locations can also be browsed in tabular format or using the graphical visualization Motif Viewer tool. All the results can be downloaded as a CSV file.

Subcellular localization prediction of cloned TF-ORFs

Protein subcellular localizations are closely linked to their biological functions, and precisely predicting protein subcellular localizations is important for gene function prediction and genome annotation. To maximize the prediction accuracy, results were derived from adopting several publicly available tools [36-39] and carefully



analyzed, compared, and combined. Consistent with the putative function of cloned genes in this study as TFs, most of them were predicted to reside in the nucleus (Additional file 3). However, Glyma13g27280.1 was predicted to be localized in the nucleus or chloroplast. Multiple subcellular localizations or altered subcellular localization of proteins are believed to be associated with multiple or altered functions, which have been observed in both mammals and plants [40-44]. Several lines of evidence also showed that nucleus encoded TFs might regulate gene expression, directly or indirectly, in other organelles such as mitochondria and chloroplasts [40, 44]. Furthermore, with the aid of another protein, a TF is able to shuttle dynamically between the nucleus and cytoplasm [45]. It is, therefore, possible that Glyma13g27280.1 functions in both of the organelles. However, experimental investigation is needed for validation of such an assumption.

Application of soybean TF-ORFeome resources to stresses studies

Since the results presented here are from various comprehensive analyses, plant biologists, especially researchers in the field of abiotic stresses, may find our genomic resources very informative in their search for candidate genes as a starting point. Two examples are given below to demonstrate what function a certain soybean TF may have by putting all data together. Glyma06g17420, one TF from our ORFeome collection, is annotated as a member of the bHLH superfamily, of which 393 members have been *in-silico* characterized in the soybean genome but until now, none have been functionally characterized in terms of drought resistance [46]. Its subcellular localization in the nucleus suggested it might function as a TF (Additional file 3). Its expression was highly up-regulated in shoots upon drought and ABA treatments (Figs. 2 and 6), indicating a role in responding to drought and probably through an ABA dependent pathway. Since it has little similarity with well characterized MYC2 or ICE1, which are positive regulators of drought tolerance [47, 48], exploring the possible novel function of Glyma06g17420 might be interesting.

NAC is one of the largest plant-specific gene families with 152 genes in soybean, and 58 of them are putative stress-responsive genes [49]. Ectopic expression of several of these stress-responsive genes in *Arabidopsis* enhanced resistance to salinity and freezing [50]. According to our qRT-PCR analysis, Glyma13g35550 (GmNAC101) was highly up-regulated by drought, dehydration, salt and ABA. Recent studies reported that higher expressions of this gene were detected in both shoots and roots of the drought-tolerant cultivar DT51 in comparison with the drought-sensitive cultivar MTD720 under drought conditions [51, 52]. More

interestingly, a total of 10 CRE motifs were identified within its 1 kb promoter sequence, indicating that this gene is under complex regulations. All of this evidence suggests that Glyma13g35550 is a potential candidate for in-depth investigation.

Conclusions

The soybean TF-ORFeome provides a valuable public resource for functional genomics studies, especially in the area of plant abiotic stresses, and will facilitate accelerating the findings in the area of abiotic stresses and in generating crops with enhanced resistance to multiple stresses.

Methods

Plant growth, treatments, and tissue collections

Soybean (cv. Williams 82) seedlings were grown in 4-gallon pots containing a mixture of turf and sand (3:1) under the same growth chamber conditions [53]. Drought treatments were initiated by withholding water at the VC stage (stage that cotyledons and unifoliates are fully expanded), while water was provided daily to the well-watered control seedlings. The water potentials for mild and moderate drought were -7 bar and -13 bar, respectively. Dehydration and salt treatments were conducted as previously described [53]. For ABA treatments, two-week-old seedlings were irrigated and sprayed with $200 \mu\text{M}$ ABA (or a mock solution without ABA as control) and incubated for certain period of times (0.5, 1, 3, and 5 h). After treatment, tissues were harvested and frozen immediately in liquid nitrogen and stored at -80°C . All samples were collected in biological triplicates.

RNA isolation and qRT-PCR

Total RNA isolation and qRT-PCR were carried out as described previously [53]. Three biological and two technical replications were conducted in all the qPCR experiments. Gene-specific primers (Additional file 6) for qRT-PCR were designed using Primer3 (version 0.4.0) [54]. The efficacy of primers for qRT-PCR was tested and desirable results were obtained. Soybean Ubiquitin3 gene (Glyma20g27950.1) was used as an internal control for all qRT-PCR analysis.

Soybean TF-ORF gene cloning

PCR was performed using Phusion high-fidelity DNA polymerase (Thermo Scientific, Pittsburgh PA, USA). PCR products were purified with a gel extraction kit (Epoch Life Sciences, Sugar Land, TX, USA), cloned into pENTR[™]/D-TOPO[®] vector or pDONR[™]/Zeo vector (Invitrogen, Carlsbad, CA, USA), and verified by sequencing using M13 forward and reverse primers, and additional gene specific primers if necessary. Primers were designed based on sequence information obtained from the Phytozome (v. 9.1) [24].

TF promoter putative CRE analysis

One thousand base pairs (bps) of the TF promoter sequences retrieved from Phytozome (version 9.1) were subjected to CRE analysis through DNA Pattern Search [55] by referring to the literature [56, 57] and the Stress Responsive Transcription Factor Database (STIFDB) [58].

TF subcellular localization prediction

Deduced TF protein sequences from experimentally verified ORF sequences were used for predicting TF proteins' subcellular localization by adopting on-line tools, including WoLF PSORT [36], PlantLoc [37], Cell-PLoc [38], and Euk-mPLoc2.0 [39].

Availability of supporting data

The data sets supporting the results of this article are included within the article and its additional files.

Additional files

Additional file 1: Fold change of expression of TF genes under water stress.

The expressions (shown as fold change) of soybean TF-ORFeome genes upon drought stress were based on publicly available data [18] and unpublished data (Valliyodan et al.). 5hR1, 5 h of dehydration stress in primary root region 1; 5hR2, 5 h of dehydration stress in primary root region 2; 48hR1, 48 h of dehydration stress in primary root region 1; 48hR2, 48 h of dehydration stress in primary root region 2; SSR, drought stressed roots; SSL, drought stressed leaves; V6L, drought stressed leaves at V6 stage; R2L, drought stressed leaves at R2 stage. The dehydration treatments and soybean primary root region 1 (apical 4 mm) and root region 2 (apical 4–8 mm) were referred to from previous definitions [59]. All of the heat maps in this article were generated using BAR HeatMapper Plus Tool [60].

Additional file 2: Validation of expressions of selected soybean TFs in shoots and roots upon drought treatment.

A, genes do not show differential expression upon drought from literature; B, genes show sequence discrepancies compared with genome annotation (Phytozome v9.1). MSL, mild drought stressed shoots; MSR, mild drought stressed roots. qRT-PCR analysis (shown as fold change) of selected soybean TFs for soybean TF-ORFeome construction under mild drought stress (see Methods).

Additional file 3: Detailed information of soybean TF-ORFs cloned in this study.

Additional file 4: Tissue/organ expression pattern of TF genes. The expression of soybean TF-ORFeome candidates in seven soybean organs including root, root tip, leaf, shoot apical meristem (SAM), nodule, flower and green pod were based on published RNA-Seq data [26]. The color scale indicates the degree of gene expression levels (yellow, low expression level; red, high expression level).

Additional file 5: Integration of TF ORFeome information into Soykb.

On the TF-ORFeome data page in SoyKB (A), clicking a gene of interest will lead to its gene card page (B), where the relevant genomic information (C) and multi-omics expression datasets (D, E) can be browsed.

Additional file 6: Primers used for qRT-PCR in this study.

Abbreviations

ABA: Abscisic acid; AP2: APETALA2; bHLH: Basic helix-loop-helix; BR: Brassinosteroids; bZIP: Basic zipper; CDS: Coding sequence; CRE: cis-regulatory element; DRE: Dehydration-responsive element; DREB: Binding to DRE; EREBP: Ethylene-responsive element binding protein; EST: Expressed sequence tag; FPKM: Fragments per kilobase of exon per million fragments mapped; GA: Gibberellin; ICE1: Inducer of CBF expression; JA: Jasmonates;

MYB: Myeloblastosis; MYC2: Mylocytomatosis oncogene homolog 2; NAC: NAM/ATAF/CUC; ORF: Open reading frame; PCR: Polymerase chain reaction; RT: Reverse transcription; SA: Salicylic acid; SoyKB: Soybean Knowledge Base; TF: Transcription factor; UTR: Un-translated regions; WGS: Whole genome sequencing.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HTN and BV provided the original concept of the study and supervised the study; CC, YW, and HTN designed the research; CC, YW, SP, and LM performed research; CC, YW, and TJ analyzed data; and CC, YW, TJ, and HTN wrote the paper. DX and HTN provided comments as well as scientific support and important revisions to the manuscript. All authors read and approved the manuscript.

Acknowledgements

We would like to thank Dr. Scott Jackson (University of Georgia, USA) for helpful discussion on the source of soybean ORF sequence discrepancy. We thank Theresa Musket for carefully editing this manuscript. We thank Jiaojiao Wang for adding the data to SoyKB. We thank the two anonymous reviewers for their constructive comments, which helped us to improve the manuscript. Funding support from United Soybean Board Grant number: 1204 (High-Impact Research for Soybean Improvement Using Genetics and Genomics) is gratefully appreciated.

Author details

¹Division of Plant Sciences and National Center for Soybean Biotechnology, University of Missouri, Columbia, MO 65211, USA. ²Department of Computer Science, Informatics Institute, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA.

Received: 23 September 2014 Accepted: 30 June 2015

Published online: 13 August 2015

References

- Li MW, Qi X, Ni M, Lam HM. Silicon era of carbon-based life: application of genomics and bioinformatics in crop stress research. *Int J Mol Sci*. 2013;14(6):11444–83.
- Yamamoto K, Sasaki T. Large-scale EST sequencing in rice. *Plant Mol Biol*. 1997;35(1–2):135–44.
- Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, et al. Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science*. 2003;302(5646):842–6.
- Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, et al. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol*. 2000;328:575–92.
- Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, et al. Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant Physiol*. 2004;135(2):773–82.
- Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, et al. C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat Genet*. 2003;34(1):35–41.
- Matsuyama A, Arai R, Yashiroda Y, Shirai A, Kamata A, Sekido S, et al. ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol*. 2006;24(7):841–7.
- Ghamsari L, Balaji S, Shen Y, Yang X, Balcha D, Fan C, et al. Genome-wide functional annotation and structural verification of metabolic ORFeome of *Chlamydomonas reinhardtii*. *BMC Genomics*. 2011;12 Suppl 1:S4.
- Pellet J, Tafforeau L, Lucas-Hourani M, Navratil V, Meynier L, Achaz G, et al. ViralORFeome: an integrated database to generate a versatile collection of viral ORFs. *Nucleic Acids Res*. 2010;38(Database issue):D371–8.
- Rajagopala SV, Yamamoto N, Zweifel AE, Nakamichi T, Huang HK, Mendez-Rios JD, et al. The *Escherichia coli* K-12 ORFeome: a resource for comparative molecular microbiology. *BMC Genomics*. 2010;11:470.
- Walhout AJ. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res*. 2006;16(12):1445–54.

12. Liu F, Andersen MN, Jensen CR. Loss of pod set caused by drought stress is associated with water status and ABA content of reproductive structures in soybean. *Funct Plant Biol.* 2003;30(3):271–80.
13. Brevedan RE, Egli DB. Short periods of water stress during seed filling, leaf senescence and yield of soybean. *Crop Sci.* 2003;43:2083–8.
14. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature.* 2010;463(7278):178–83.
15. Joshi T, Patil K, Fitzpatrick MR, Franklin LD, Yao Q, Cook JR, et al. Soybean Knowledge Base (SoyKB): a web resource for soybean translational genomics. *BMC Genomics.* 2012;13 Suppl 1:S15.
16. Joshi T, Fitzpatrick MR, Chen S, Liu Y, Zhang H, Endacott RZ, et al. Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. *Nucleic Acids Res.* 2014;42(Database issue):D1245–52.
17. Soybean Knowledge Base (SoyKB). [<http://soykb.org>].
18. Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M, le Ham H, et al. Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. *PLoS One.* 2012;7(11), e49522.
19. Lindemose S, O'Shea C, Jensen MK, Skriver K. Structure, function and networks of transcription factors involved in abiotic stress responses. *Int J Mol Sci.* 2013;14(3):5842–78.
20. Gao SQ, Chen M, Xu ZS, Zhao CP, Li L, Xu HJ, et al. The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants. *Plant Mol Biol.* 2011;75(6):537–53.
21. Zhang G, Chen M, Li L, Xu Z, Chen X, Guo J, et al. Overexpression of the soybean GmERF3 gene, an AP2/ERF type transcription factor for increased tolerances to salt, drought, and diseases in transgenic tobacco. *J Exp Bot.* 2009;60(13):3781–96.
22. Chen M, Wang QY, Cheng XG, Xu ZS, Li LC, Ye XG, et al. GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. *Biochem Biophys Res Commun.* 2007;353(2):299–305.
23. Quach TN, Tran LS, Valliyodan B, Nguyen HT, Kumar R, Neelakandan AK, et al. Functional analysis of water stress-responsive soybean gmnac003 and gmnac004 transcription factors in lateral root development in arabidopsis. *PLoS One.* 2014;9(1), e84886.
24. Phytozome (v 9.1). [<http://www.phytozome.net/>].
25. Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, et al. The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiol.* 2011;155(2):645–55.
26. Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, et al. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* 2010;63(1):86–99.
27. Urbaniak GC, Plous S. Research Randomizer (Version 4.0). <http://www.randomizer.org/>. Retrieved on 22 June 2013.
28. Valliyodan B, Nguyen HT. Genomics of Abiotic Stress in Soybean. In: *Genetics and Genomics of Soybean*. Edited by Stacey G, vol. 2: Springer; 2008. p. 343–72.
29. Zhu JK. Salt and drought stress signal transduction in plants. *Annu Rev Plant Biol.* 2002;53:247–73.
30. Nakashima K, Yamaguchi-Shinozaki K. ABA signaling in stress-response and seed development. *Plant Cell Rep.* 2013;32(7):959–70.
31. Sreenivasulu N, Harshavardhan VT, Govind G, Seiler C, Kohli A. Contrapuntal role of ABA: does it mediate stress tolerance or plant growth retardation under long-term drought stress? *Gene.* 2012;506(2):265–73.
32. Lee SC, Luan S. ABA signal transduction at the crossroad of biotic and abiotic stress responses. *Plant Cell Environ.* 2012;35(1):53–60.
33. Acharya BR, Assmann SM. Hormone interactions in stomatal function. *Plant Mol Biol.* 2009;69(4):451–62.
34. Duan L, Dietrich D, Ng CH, Chan PM, Bhalerao R, Bennett MJ, et al. Endodermal ABA signaling promotes lateral root quiescence during salt stress in *Arabidopsis* seedlings. *Plant Cell.* 2013;25(1):324–41.
35. Chai C, Wang Y, Joshi T, Valliyodan B, Prince S, Michel L, et al. Soybean transcription factor ORFeome associated with drought. [http://soykb.org/TF_ORFeome/].
36. Horton P, Park K, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007;35:W585–7.
37. Tang S, Li T, Cong P, Xiong W, Wang Z, Sun J. PlantLoc: an accurate web server for predicting plant protein subcellular localization by substantiality motif. *Nucleic Acids Res.* 2013;41(Web Server issue):W441–7.
38. Chou KC, Shen HB. Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat Protoc.* 2008;3(2):153–62.
39. Chou KC, Shen HB. A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS One.* 2010;5(4), e9931.
40. Leigh-Brown S, Enriquez JA, Odom DT. Nuclear transcription factors in mammalian mitochondria. *Genome Biol.* 2010;11(7):215.
41. Lucas CH, Calvez M, Babu R, Brown A. Altered subcellular localization of the NeuN/Rbfox3 RNA splicing factor in HIV-associated neurocognitive disorders (HAND). *Neurosci Lett.* 2014;558:97–102.
42. Patel VP, Defranco DB, Chu CT. Altered transcription factor trafficking in oxidatively-stressed neuronal cells. *Biochim Biophys Acta.* 2012;1822(11):1773–82.
43. Home P, Saha B, Ray S, Dutta D, Gunewardena S, Yoo B, et al. Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. *Proc Natl Acad Sci U S A.* 2012;109(19):7362–7.
44. Liere K, Weihe A, Borner T. The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J Plant Physiol.* 2011;168(12):1345–60.
45. Camarata T, Bimber B, Kulisz A, Chew TL, Yeung J, Simon HG. LMP4 regulates Tbx5 protein subcellular localization and activity. *J Cell Biol.* 2006;174(3):339–48.
46. Wang Z, Libault M, Joshi T, Valliyodan B, Nguyen HT, Xu D, et al. SoyDB: a knowledge database of soybean transcription factors. *BMC Plant Biol.* 2010;10:14.
47. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell.* 2003;15(1):63–78.
48. Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, et al. ICE1: A regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes Dev.* 2003;17(8):1043–54.
49. Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K, et al. Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Res.* 2011;18(4):263–76.
50. Hao YJ, Wei W, Song QX, Chen HW, Zhang YQ, Wang F, et al. Soybean NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants. *Plant J.* 2011;68(2):302–13.
51. Thu NB, Hoang XL, Doan H, Nguyen TH, Bui D, Thao NP, et al. Differential expression analysis of a subset of GmNAC genes in shoots of two contrasting drought-responsive soybean cultivars DT51 and MTD720 under normal and drought conditions. *Mol Biol Rep.* 2014;41(9):5563–9.
52. Thao NP, Thu NB, Hoang XL, Van Ha C, Tran LS. Differential expression analysis of a subset of drought-responsive GmNAC genes in two soybean cultivars differing in drought tolerance. *Int J Mol Sci.* 2013;14(12):23828–41.
53. Tran LS, Quach TN, Guttikonda SK, Aldrich DL, Kumar R, Neelakandan A, et al. Molecular characterization of stress-inducible GmNAC genes in soybean. *Mol Genet Genomics.* 2009;281(6):647–64.
54. Primer3 (version 0.4.0). [<http://frodo.wi.mit.edu/>].
55. DNA Pattern Search. [http://www.geneinfinity.org/sms/sms_DNApatterns.html].
56. Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS. In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. *DNA Res.* 2009;16(6):353–69.
57. Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci.* 2005;10(2):88–94.
58. Stress Responsive Transcription Factor Database (STIFDB, V2.0). [<http://caps.ncbs.res.in/stifdb>].
59. Yamaguchi M, Valliyodan B, Zhang J, Lenoble ME, Yu O, Rogers EE, et al. Regulation of growth response to water stress in the soybean primary root. I. Proteomic analysis reveals region-specific regulation of phenylpropanoid metabolism and control of free iron in the elongation zone. *Plant Cell Environ.* 2010;33(2):223–43.
60. BAR HeatMapper Plus Tool. [http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi].