

# Identification and prediction of abiotic stress responsive transcription factors involved in abiotic stress signaling in soybean

Lam-Son Phan Tran<sup>1,\*</sup> and Keiichi Mochida<sup>2,\*</sup>

<sup>1</sup>RIKEN Plant Science Center; Signaling Pathway Research Unit; Tsurumi, Yokohama Japan; <sup>2</sup>RIKEN Plant Science Center; Gene Discovery Research Group; Tsurumi, Yokohama Japan

**Key words:** soybean, transcription factor, cis-element, abiotic stress, gene ontology

**Abbreviations:** ABA, abscisic acid; TF, transcription factor; DBD, DNA binding domain; TRR, transcriptional regulation region; GO, gene ontology

Submitted: 11/06/09

Accepted: 11/06/09

Previously published online:

[www.landesbioscience.com/journals/psb/article/10550](http://www.landesbioscience.com/journals/psb/article/10550)

\*Correspondence to: Lam-Son Phan Tran/Keiichi Mochida; Email: [tran@psc.riken.jp](mailto:tran@psc.riken.jp)/[mochida@psc.riken.jp](mailto:mochida@psc.riken.jp)

Addendum to: 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; 16:353-69; PMID: 19884168; DOI: 10.1093/dnares/dsp023.

**A**biotic stresses such as extreme temperature, drought, high salinity, cold and waterlogging often result in significant losses to the yields of economically important crops such as soybean (*Glycine max* L.). Transcription factors (TFs) which bind to DNA through specific cis-regulatory sequences either activate or repress gene transcription have been reported to act as control switches in stress signaling. Recent completion of the soybean genomic sequence has open wide opportunities for large-scale identification and annotations of regulatory TFs in soybean for functional studies. Within the soybean genome, we identified 5,035 TF models which grouped into 61 families. Detailed annotations of soybean TF genes can be accessed at SoybeanTFDB ([soybeantfdb.psc.riken.jp](http://soybeantfdb.psc.riken.jp)). Moreover, we have reported a new idea of high throughput prediction and selection of abiotic stress responsive TFs based on the existence of known stress responsive cis-element(s) located in the promoter regions of respective TFs and GO annotations. We, therefore, have provided a basic platform for the genome-wide analysis of regulatory mechanisms underlying abiotic stress responses and a reliable tool for prediction and selection of stress responsive TFs for further functional studies and genetic engineering.

Plants grow in a dynamic and uncontrolled environment that can frequently impose constraints on growth and development, resulting in significant losses to the yields of economically important crops. Among the adverse abiotic factors

commonly encountered by land plants are extreme temperature, water deficit, high salinity and submergence that affect plant growth and productivity. In response to these stresses, plants activate a number of defense mechanisms that function to increase tolerance to the adverse conditions.<sup>1-3</sup>

The genetic basis of stress tolerance is complex and not well understood. Molecular and cellular responses to environmental stimuli have been analyzed extensively at the biochemical, physiological, morphological and metabolic levels. The early events of plant adaptation to environmental stresses are the stress signal perception and subsequent signal transduction through either ABA-dependent or ABA-independent pathway, leading to the activation of various physiological and metabolic responses.<sup>3-5</sup> Furthermore, a large array of genes are activated by stress conditions, meaning several proteins are produced to join the pathways which subsequently lead to synergistic enhancement of stress tolerance.<sup>6-10</sup> These genes are classified into two groups: functional and regulatory genes. The functional group contains genes encoding metabolic components such as sugar, sugar alcohols and amines which play important role in stress tolerance. The regulatory group includes genes encoding various transcription factors (TFs) which can regulate various stress-inducible genes cooperatively or separately, and may constitute gene networks or signaling cascades, in which given TFs may regulate other TFs and/or other regulatory and/or functional genes through their specific binding sequence,

the cis-regulatory sequences located in the promoter regions of target genes.<sup>3,5-11</sup>

Typically, the TFs contain a distinct type of DNA binding domain and transcriptional regulation region (TRR).<sup>3,5,12</sup> Through their DNA binding domains TFs bind to their specific cis-regulatory sequences and activate or repress transcription of multiple target genes depending on the nature of their TRR. Individual members of the same family often respond differently to various stress stimuli; on the other hand, some stress-responsive genes may share the same TF through the conserved cis-acting element in the promoters of respective targets. The TFs and cis-motifs function not only as molecular switches for gene expression, but also as terminal points of signal transduction in the signaling processes.<sup>3,5,11,12</sup>

In plants, approximately 7% of the genome encodes for putative TFs.<sup>13</sup> Our basic knowledge of TFs and their role in transcriptional regulation is derived from molecular biological and genetic investigations. Proper characterization of particular TFs often requires a detailed study in the biological context of a whole TF family since functional redundancy is a common occurrence within TF families.<sup>14-18</sup> Furthermore, since TFs control the expression of the genome, it is not possible to completely understand their function without performing detailed functional studies at a genome-wide level.<sup>19</sup> At the present time, the functions for most of stress responsive TF encoding genes are not fully understood. Identification and functional analysis of stress-inducible TFs will provide more information on the complex regulatory gene networks controlling stress responses, and potential tools for genetic engineering of transgenic crop plants with superior yield under stress conditions.<sup>2,3,5</sup>

Soybean (*Glycine max* L.) is a nutritionally important crop which provides an abundant source of oil and protein for worldwide human consumption.<sup>20</sup> Taking advantage of the recent completion of the soybean genomic sequence ([www.phytozome.net/soybean#C](http://www.phytozome.net/soybean#C)), as an initial step to study stress responsive TFs involved in signal transduction networks which control stress response in soybean we carried out a computational analysis of the

soybean genomic sequence to identify the full complement of TF-encoding genes. Through our intensive analysis we identified a total of 5,035 TF models which grouped into 61 families. Relevant annotations of each soybean TF gene can be accessed at SoybeanTFDB ([soybeantfdb.psc.riken.jp](http://soybeantfdb.psc.riken.jp)).<sup>21</sup>

Given the fact that at the present time, the stress-related expression data, which would aid in selection of stress responsive TFs for further functional studies and genetic engineering, are not yet available for most of the soybean TF genes, it was important to find an approach to systematically identify the stress responsive TFs for further characterization and functional analyses in a reliable manner. Recent studies have substantiated that sequence similarity-based clustering of the members of several TF families correlates with their function. For instance, phylogenetic analysis of the AP2\_EREBP and NAC families of soybean and the rice NAC family with orthologs from other plant species whose stress responsive expression pattern and/or function are known, resulted in a nearly perfect match between sequence conservation and function or expression patterns. These similarities clearly demonstrate that this can serve as a reliable approach to rationalize systematic functional predictions of different TF members.<sup>14,17,22</sup> Moreover, increasing evidence indicates that the cis-motifs are highly conserved among orthologous or paralogous genes and coregulated genes, and defined cis-elements can effectively aid in the genome-wide screening of ABA and abiotic stress responsive genes.<sup>23-25</sup> These observations together prompted us to investigate in a comprehensive fashion the relationship between TFs and abiotic stress with the integration of stress responsive cis-element annotation and comparative sequence analysis using stress responsive GO terms which aimed to identifying soybean TFs that may function in abiotic stress response. We, therefore, carried out comparative sequence analysis with stress-responsive Arabidopsis TFs to predict the soybean TFs with stress responsive GO terms. We also searched for the existence of the well-known stress-responsive cis-elements, which were reported previously,<sup>11</sup> in promoter regions of each soybean TF gene.

With the help of our soybean TF database (SoybeanTFDB), at first users can search for TF genes which harbor known stress responsive cis-motif(s) in their promoter regions. Next, users can screen the identified TFs using GO annotation provided for each TF on detailed annotation page. Thus, users will be able to identify the putative stress responsive TFs based on both the existence of major stress responsive cis-motif(s) and the associated stress responsive GO terms.

In conclusion integration of comparative sequence analysis and cis-motif and GO annotations provided through this study could help identify stress responsive TF genes in a systematic manner. Users can easily use the prediction tools provided on SoybeanTFDB ([soybeantfdb.psc.riken.jp](http://soybeantfdb.psc.riken.jp)) to predict stress responsive TFs. The predicted stress responsive function of the identified TFs shall be then confirmed by experiments. SoybeanTFDB has a very user-friendly interface, and aims to meet the broad demands of researchers who strive to perform research with soybean TFs with the goal of gaining greater understanding of their putative roles in different signaling pathways underlying plant development, differentiation and environmental responses. We truly believe that our results will rapidly accelerate the progress in “transcription factoromics” of soybean, comparative genomics of regulatory TF proteins both within legume species and between legumes and other species, as well as facilitate genetic engineering programs to improve the productivity of soybean grown in adverse conditions.<sup>21</sup>

#### Acknowledgements

The soybean sequence data were produced by the US Department of Energy Joint Genome Institute [www.jgi.doe.gov/in](http://www.jgi.doe.gov/in) collaboration with the scientific user community. Funding supports from Grants-in-Aid (Start-up) for Scientific Research, Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 21870046) and from Yokohama Institute Director Discretionary Funds (2009) for TPLS are gratefully appreciated.

## References

1. Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K. Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Curr Opin Biotechnol* 2006; 17:113-22.
2. Valliyodan B, Nguyen HT. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr Opin Plant Biol* 2006; 9:189-95.
3. Yamaguchi-Shinozaki K, Shinozaki K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu Rev Plant Biol* 2006; 57:781-803.
4. Tran L-SP, Urao T, Qin F, Maruyama K, Kakimoto T, Shinozaki K, et al. Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought and salt stress in *Arabidopsis*. *Proc Natl Acad Sci USA* 2007; 104:20623-8.
5. Tran L-SP, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K. Plant gene networks in osmotic stress response: from genes to regulatory networks. *Methods in Enzymology* 2007; 428:109-28.
6. Seki M, Kamei A, Yamaguchi-Shinozaki K, Shinozaki K. Molecular responses to drought, salinity and frost: common and different paths for plant protection. *Curr Opin Biotechnol* 2003; 14:194-9.
7. Wang W, Vinocur B, Altman A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 2003; 218:1-14.
8. Bray EA. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. *J Exp Bot* 2004; 55:2331-41.
9. Kacperska A. Sensor types in signal transduction pathways in plant cells responding to abiotic stressors: do they depend on stress intensity? *Physiol Plant* 2004; 122:159-68.
10. Zhang JZ, Creelman RA, Zhu JK. From laboratory to field Using information from *Arabidopsis* to engineer salt, cold and drought tolerance in crops. *Plant Physiol* 2004; 135:615-21.
11. Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci* 2005; 10:88-94.
12. Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. *Plant Physiol* 2009; 149:88-95.
13. Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, et al. Legume transcription factors: global regulators of plant development and response to the environment. *Plant Physiol* 2007; 144:538-49.
14. Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, et al. Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the *early responsive to dehydration stress 1* promoter. *Plant Cell* 2004; 16:2481-98.
15. Fang Y, You J, Xie K, Xie W, Xiong L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Mol Genet Genomics* 2008; 280:535-46.
16. Liao Y, Zou HF, Wei W, Hao YJ, Tian AG, Huang J, et al. Soybean GmbZIP144, GmbZIP162 and GmbZIP178 genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic *Arabidopsis*. *Planta* 2008; 228:225-40.
17. Zhou QY, Tian AG, Zou HF, Xie ZM, Lei G, Huang J, et al. Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21* and *GmWRKY54*, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants. *Plant Biotechnol J* 2008; 6:486-503.
18. Tran L-SP, Quach T, Guttikonda SK, Aldrich DL, Rajesh K, Neelakandan A, et al. Molecular characterization of stress-inducible *GmNAC* genes in soybean. *Mol Gen Genomics* 2009; 281:647-64.
19. 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; 135:773-82.
20. Tran L-SP, Nguyen HT. Future Biotechnology of Legumes. In Emerich WD, Krishnan H, ed. *Nitrogen Fixation in Crop Production*. Madison WI, USA. The American Society of Agronomy, Crop Science Society of America and Soil Science Society of America 2009; 265-308.
21. 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; 16:353-69.
22. Zhang G, Chen M, Chen X, Xu Z, Chen X, Guo J, et al. Phylogeny, gene structures and expression patterns of the ERF gene family in soybean (*Glycine max L.*). *J Exp Bot* 2008; 59:4095-107.
23. Zhang W, Ruan J, Ho TH, You Y, Yu T, Quatrano RS. Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid- and abiotic stress-responsive genes in *Arabidopsis thaliana*. *Bioinformatics* 2005; 21:3074-81.
24. Kim DW, Lee SH, Choi SB, Won SK, Heo YK, Cho M, et al. Functional conservation of a root hair cell-specific cis-element in angiosperms with different root hair distribution patterns. *Plant Cell* 2006; 18:2958-70.
25. Won SK, Lee YJ, Lee HY, Heo YK, Cho M, Cho HT. cis-element- and transcriptome-based screening of root hair-specific genes and their functional characterization in *Arabidopsis*. *Plant Physiol* 2009; 150:1459-73.