Contribution of Genomics to Gene Discovery in Plant Abiotic Stress Responses

Yun Ma[a], **Feng Qin**[a,2] and **Lam-Son Phan Tran**[b,1]

a Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing, 100039, China
b Signaling Pathway Research Unit, Plant Science Center, RIKEN Yokohama Institute, 1–7–22, Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

Sustaining agricultural production under adverse environmental conditions, such as drought and high salinity, represents a major challenge. The discovery of key genes and signal transduction pathways underlying plant responses to environmental stress will play an important role in developing strategies for the genetic improvement of crops to address this challenge. Crop functional genomics has greatly contributed to the identification of abiotic stress-related genes. Current advances in genomic technologies now provide effective and high-throughput methods for identifying stress-related genes at a genome-wide level, especially with the availability of the complete genomic sequence of several model and crop plant species. The development of genetic database resources has allowed bioinformatic approaches to identify stress-tolerant gene families across species based on homology and synteny. Additionally, genome-wide association studies (GWAS) for complex trait loci in crops have facilitated the discovery of critical stress-related genes and their favorable alleles.

DATABASE RESOURCES

Web-based databases of various plant genomes have become an invaluable resource for plant biologists worldwide. Researchers can readily acquire diverse genetic information on coding and non-coding sequences, regulatory elements, gene families, molecular markers, and genetic variability among different plant species. Regarding the current topic, advanced genome database resources may serve as a substantial resource for the identification of stress-responsive genes. A summary of available database resources and their web addresses are provided in Supplemental Table 1.

NEXT GENERATION SEQUENCING (NGS) TECHNOLOGY

Initial approaches to providing genomic sequences required large investments of human and monetary resources. The development and application of NGS technologies, however, have greatly facilitated the ability to provide genome sequences of more and more plant species. This has created a wide range of opportunities for the discovery of stress-related genes and pathways that can serve as the foundation for crop improvement. For example, NGS technology can be coupled with high-throughput transcriptome profiling to investigate genome-wide changes in transcripts in response to stresses (Molina et al., 2011).

Although NGS technology is still in the early stage of its application, it has proved to be a robust tool for the genome-wide identification of genetic variation. NGS, combined with GWAS, has been used to identify potential molecular markers, such as single nucleotide polymorphisms, insertions and deletions, and copy number variations, which are associated with growth and development and/or stress responses. NGS technology will facilitate the development of comparative genomics and enable explorations of diversity within or across species.

COMPARATIVE GENOMICS

The availability of plant genomes, along with the accumulation of expression data and an increasing number of stress-related cDNA libraries, represent valuable resources for comparative genomics-based discovery of stress-related genes and pathways. Researchers can now transfer gene annotations from model plants to genes of newly sequenced crop species, where functional studies are rudimentary. An important requirement, however, in using comparative genomics is the availability of datasets of orthologs evolved from the same ancestor. These genes frequently execute similar gene functions that are maintained

among species evolved from a common ancestor (Tran and Mochida, 2010a). Regarding stress-related transcription factors (TFs), similarities in sequence and gene expression patterns of orthologs identified in different plant species will enable us to predict orthologous genes with similar functions in newly sequenced crop species. Using this approach, stress-responsive TFs were predicted in soybean, maize, sorghum, barley, and wheat using a comparative analysis of known stress-responsive TFs in *Arabidopsis* and rice (Mochida et al., 2009; Tran and Mochida, 2010b; Mochida et al., 2011). Comparative genomics can also be used to analyze the stress-related expression profiles of less-studied plant species to identify stress-related genes and compare gene expression profiles between the species (Walia et al., 2009). Common features in stress responses among different plant species, identified by comparative analysis of genome databases, will help researchers to predict and understand the function of genes in newly sequenced species. Additionally, differences between and among model and crop species will provide a good opportunity to identify species-specific stress-responsive genes and regulatory mechanisms. Hence, comparative genomics will broaden the ability to transfer information from model plants to other species that are fundamental to food production or as a source of alternative energy.

GENOME-WIDE ASSOCIATION STUDIES

GWAS is a method for high-resolution mapping of complex trait loci based on linkage disequilibrium (Flint-Garcia et al., 2005). GWAS has been used to investigate the association between genetic variation and important agricultural traits, including stress tolerance. An example of GWAS applied to stress tolerance was reported by Famoso et al. (2011), who conducted GWAS among 383 rice accessions in combination with bi-parental QTL mapping to identify genetic variations underlying aluminum (Al) tolerance. Using this approach, 48 genomic regions were identified that were correlated with Al tolerance (Famoso et al., 2011).

Although the application of GWAS to plant species is still in infancy, it has great potential for identifying valuable natural variations in trait-associated loci, as well as allelic variations in candidate genes underlying quantitative and complex traits, including those related to growth, development, stress, and nutritional quality. Candidate genes identified by GWAS can be used to define the priority assigned to genes selected for more intensive biochemical and physiological studies on the mechanisms of stress response. The integration of available database resources, such as geographic information, genetic polymorphism data, and GWAS data, into a common, web-based application will greatly facilitate both forward and reverse genetic approaches to the study of stress tolerance (Huang et al., 2011).

PHYLOGENETIC AND EVOUTIONARY ANALYSIS OF GENE FAMILIES

Plant genes, especially stress-related genes, are often clustered in gene families consisting of members with redundant functions. The genomic information available in public databases has made the study of the origin and expansion of gene families possible. Over the course of evolution, plant genomes have undergone large-scale duplication and rearrangement during which some duplicated genes (paralogs) gain new functions (neofunctionalization), some persist as a subset maintaining their ancestral functions (subfunctionalization), while some others completely lose functionality (nonfunctionalization) (Demuth and Hahn, 2009). Advances in genomics and bioinformatics have assisted researchers in elucidating the evolution of plant gene families. This information can contribute to a better understanding of the functional diversity of individual genes belonging to one family.

Comprehensive analysis of gene families, such as TF and two-component system (TCS) families, within different genotypes of a species has been carried out in several plant species to identify stress-related genes. For instance, phylogenetic analysis of the soybean *GmNAC* TF family predicted 58 dehydration-responsive genes (Le et al., 2011b). Using RT–qPCR, the expression of 31 *GmNAC* genes out of 38 genes examined was further confirmed to be altered to varying degrees in shoots or roots in response to dehydration. This study demonstrated a highly accurate rate of phylogenetic analysis-based functional prediction and identified dehydration-related candidate genes for further functional studies aimed at the improvement of drought tolerance of soybean cultivars (Le et al., 2011b). A combination of comparative and phylogenetic analyses of TCS elements in soybean, *Arabidopsis*, and rice was used to identify and predict the putative functions of TCS members, including putative stress-related functions, in soybean (Mochida et al., 2010). Following this study, expression profiles of all soybean TCS encoding genes were also characterized in shoots and roots of soybean seedlings under both normal and dehydrated conditions in order to identify potential dehydration-responsive TCS genes for further *in planta* functional studies (Le et al., 2011a; Ha et al., 2012).

CONCLUDING REMARKS

Unprecedented growth in human population, accompanied by climate change and a decrease in arable land, will increase the demand for crops that can sustain high yields under adverse conditions. Therefore, understanding mechanisms underlying stress tolerance in plants is not only an academic interest, but also has great socioeconomic significance. Identifying and mining genes involved in stress response represents a key step to unraveling and manipulating stress

tolerance in plants. Recent advances in genomics make identifying stress response genes at a whole-genome level possible in both model plants that have been studied extensively, as well as both native and crop species that have been recently sequenced or sequenced in the future. Uncultivated, native plants that are endemic to harsh environments may represent a rich source of novel, stress-tolerant genes. NGS technology has brought about a revolution in gene sequencing, accelerating the development of genomic information that can be used to identify stress-related genes in an ever-widening array of species. Sequence data can be combined with diverse bioinformatic approaches to accelerate stress tolerance research.

Comparative analysis among genotypes within the same species and between species will enable us to identify species-specific genes underlying stress responses. GWAS, which has been recently applied to plant research, has greatly identified loci and allelic variation associated with complex, quantitative traits, including stress tolerance in a more rapid and cost-efficient manner. Various '-omic' technologies are being used in a systems biology approach to better understand plant processes to obtain a comprehensive knowledge of how plants respond and adapt to environmental stress. The integration of diverse data sets and emerging technologies represents an effective and promising approach to obtaining the critical knowledge necessary to sustainably increase food production under adverse conditions through the genetic improvement of crops.

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online.*

REFERENCES

Demuth, J.P., and Hahn, M.W. (2009). The life and death of gene families. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology. **31,** 29–39.

Famoso, **A.N**., et al. (2011). Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. PLoS Genet. **7,** e1002221.

Flint–Garcia, **S.A**., et al. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. Cell. Mol. Biol. **44,** 1054–1064.

Ha, S., Vankova, R., Yamaguchi–Shinozaki, K., Shinozaki, K., and Tran, L.S. (2012). Cytokinins: metabolism and function in plant adaptation to environmental stresses. Trends Plant Sci. **17,** 172–179.

Huang, Y.S., et al. (2011). Analysis and Visualization of *Arabidopsis thaliana* GWAS Using Web 2.0 Technologies (Oxford: Database).

Le, **D.T**., et al. (2011a). Genome-wide expression profiling of soybean two-component system genes in soybean root and shoot tissues under dehydration stress. DNA Res. **18,** 17–29.

Le, **D.T**., et al. (2011b). Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. DNA Res. **18,** 263–276.

Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi–Shinozaki, K., Shinozaki, K., and Tran, L.S. (2009). *In silico* analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. DNA Res. **16,** 353–369.

Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi–Shinozaki, K., Shinozaki, K., and Tran, L.S. (2010). Genome-wide analysis of two-component systems and prediction of stress-responsive two-component system members in soybean. DNA Res. **17,** 303–324.

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

Molina, **C**., et al. (2011). The salt-responsive transcriptome of chickpea roots and nodules via deepSuperSAGE. BMC Plant Biol. **11,** 31.

Tran, L.S., and Mochida, K. (2010a). A platform for functional prediction and comparative analyses of transcription factors of legumes and beyond. Plant Signal Behav. **5,** 550–552.

Tran, L.S., and Mochida, K. (2010b). Identification and prediction of abiotic stress responsive transcription factors involved in abiotic stress signaling in soybean. Plant Signal Behav. **5,** 255–257.

Walia, H., Wilson, C., Ismail, A.M., Close, T.J., and Cui, X. (2009). Comparing genomic expression patterns across plant species reveals highly diverged transcriptional dynamics in response to salt stress. BMC Genomics. **10,** 398.