Original Article

Computational Identification and Comparative Genomic Analysis of Soybean Oxidative Stress-Related Genes

Sharma Arti, Bong-Gyu Mun, Byung-Wook Yun*

School of Applied Biosciences, Kyungpook National University, Daegu 702-701, Republic of Korea

Abstract

Reactive oxygen and nitrogen species (ROS and RNS, respectively) are messengers that carry signals to alter the redox state in order to activate plant responses and other physiological processes, such as differentiation, aging, senescence, and pathogen defense. Quite a large number of genes are involved in this signaling and lead to oxidative stress in plants. Although the role of ROS/RNS during stress conditions is well documented, a comprehensive list of genes and comparative study of these genes has not yet been completed. Accordingly, the *in silico* identification of oxidative stress-related genes was performed for soybeans and Arabidopsis. These genes were also studied in relation to multiple domain prediction. The presence of domains like dehydogenase and ATPase suggests that these genes are involved in various metabolic processes, as well as the transportation of ions under optimal environmental conditions. In addition to a sequence analysis, a phylogenetic analysis was also performed to identify orthologous pairs among the soybean and Arabidopsis oxidative stress-related genes based on neighbor joining. This study was also conducted with the objective of further understanding the complex molecular signaling mechanism in plants under various stress conditions.

Keywords : Arabidopsis, Comparative genomics, Oxidative stress, Phylogenetic analysis, Soybean

Introduction

Plants have sophisticated systems that respond to fluctuations in their environment. Consequences of plant response to adverse environments (biotic and abiotic stresses) are enhancement of ROS/RNS and antioxidant defense, leading to physiological and metabolic changes. An imbalance due to ROS and RNS disturbs the normalredoxstate of cells and modulates the cellular metabolism (Gechev 2006). Traditionally, ROS were considered to be toxic due to the production of peroxides and free radicals, which damage all the cell components, including proteins, lipids, and DNA (Mittler 2002). However, in recent years, it has become evident that increased ROS production is also associated with plant defense responses. ROS act as signaling molecules to mediate cellular processes, such as programmed cell death (Harding et al. 2003), abiotic stress responses (Madhava and Sresty 2000), pathogen defense (Torres et al. 2005), and systemic signaling (Suzuki et al. 2011). Yet, while ROS can be used by plants for cellular homeostasis to monitor their intracellular level of stress, this level has to be kept under tight regulation as over-accumulation of ROS can result in cell death.

Nitric oxide (NO) is already known to serve an important function in various physiological processes, ranging from seed germination (Beligni and Lamattina 2000; Bethke et al. 2007),

regulation of plant maturation and senescence (Guo and Crawford 2005), suppression of floral transition (He et al. 2004), and involvement in light-mediated greening (Zhang et al. 2006), to mediation of stomatal movement as an intermediate downstream of ABA signaling (Bright et al. 2006; García-Mata and Lamattina 2007) and regulation of multiple plant responses toward a variety of abiotic and biotic stresses, such as drought (García-Mata and Lamattina 2001), salt (Zhao et al. 2004), heat (Uchida et al. 2002), and disease infection (Delledonne et al. 2005). Yet, an enhanced NO concentration can cause inhibition of shoot and root development, potential damage to photosynthetic electron transport, DNA damage and cell death (Yun et al. 2011).

One of the mechanisms contributing to oxidative signal induced stress and pathogen tolerance is the activation of a detoxification process and defense gene expression. For example, Arabidopsis plants respond to oxidative stress with an increase in production of antioxidant enzymes, including glutathione- S-transferases (GSTs), peroxidases, superoxide dismutases, and catalases, as well as the activation of protective genes encoding heat shock proteins (HSPs) and pathogenesis-related proteins.

Comparative genomics can be used to gain knowledge of gene organization, and is particularly helpful in examining genome

evolution (Keller and Feuillet 2000; Ellwood et al.2008). Closely related species have extensive regions of gene co-linearity, a phenomenon also known as synteny (Zang et al. 2009), however, as the evolutionary distance between two species increases, the segments of co-linearity become shorter.

The recent availability of complete genomes of several model systems has sparked renewed interest in the study of co-linearity due to the potential of transferring useful information from well-studied small genomes to larger ones (Rubin et al. 2000). Conservation of the structure and function of genetic loci among species has already been documented, even though genome sizes in plant species exhibit large variations, e.g., 125 Mb for *A thaliana* to 125 Gb for *Fritillaria assyriaca* (Bennett and Smith 1976). Extensive gene conservation, both in structure and function, has also been reported in grass genomes (Bennetzen and Ramakrishna 2002), Crucifers (Axelsson 2001), and solanaceous plants (Causse 2007). Plus, comparative genomics is widely used in crops to study various traits (yield, disease resistance etc.). Agalou et al. (2008) predicted 33 drought-responsive genes of the HD-ZIP family in rice by exploiting comparative genomics with Arabidopsis. Therefore, these studies suggest that a comparative analysis of oxidative stress-related genes is a suitable strategy to investigate the molecular basis of plant responses to the disruption of cellular homeostasis in different plant species.

Accordingly, the present study conducted a comprehensive analysis of the oxidative stress-related genes in Arabidopsis and soybeans due to the availability of their genome sequences and several EST and cDNA collections. Moreover, there have already been several reports on stress-related genes in Arabidopsis (Mahantesha et al. 2013 Subhomoi et al. 2013), which can serve as a basis for the manual screening of the Arabidopsis database and also be utilized for soybeans. Soybeans are a major oil crop worldwide and extensively grown around the city of Daegu and in Gyeongbuk Province, South Korea with the land under soybean cultivation being 14,563 ha (Statistics Korea, http://kostat.go.kr/dbro).

This study was carried out with following goals: firstly, identification of oxidative stress-related genes in Arabidopsis, and secondly, their orthologs in soybeans for a comparative genomics approach to exploit candidate genetic traits.

Materials and Methods

Retrieval of genes sequences
The genes involved in oxidative stress were retrieved from the Arabidopsis database (TAIR, http://www.arabidopsis.org/) by referring to the comprehensive stress gene catalog provided by Mahanteshaet al.(2013) Subhomoi et al. (2013). The selected genes included all the oxidative stress-related genes available in the TAIR website. The protein sequences of these genes were then used as a query against the soybean database (SoyBase and the Soybean Breeder's Toolbox, http://soybase.org/GlycineBlastPages/). Next, the protein function domains of the identified oxidative stress-related genes were examined based onHidden Markov Model (HMM) searches of the conserved domain databases (CDD) of NCBI (NCBI, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Gene models for the soybean genomes were downloaded from Phytozome (Phytozome, http://www.phytozome.net/). The Arabidopsis gene models were downloaded from TAIR (TAIR, http://www.arabidopsis.org/).

Phylogenetic analysis:
The sequence identities among the soybean and Arabidopsis genes were confirmed using ClustalW in Mega 6. The phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6 (Tamura et al. 2013). The phylogenetic trees were constructed using the neighbor-joining method, including 1000 replications for a bootstrap analysis to generate a dendrogram.

Results
Manual screening of the Arabidopsis genome using previously identified domains was conducted to predict the oxidative stress-related genes. As a result, a total of 252 genes responsible for oxidative stress were identified in the Arabidopsis genome. Plus, all these genes were subjected to a domain analysis in CDD. Therefore, these proteins were used as query sequences to identify the oxidative stress-related genes in the soybean database using a BLAST (Altschul et al. 1990) search. Sequences with a high score, >50% identity and e-value >1^{-30}, were then selected from the soybean database, resulting in 396 gene matches that represented potential oxidative stress-related genes in soybeans.

These genes were further confirmed as oxidative stress-related genes using a conserved domain analysis. Oxidative stress-related genes are highly diverse and mainly include peroxidase, thirodoxin, b-ZIP, Zf-CH, tranferases, kinases, ATPases, Zinc fingers, and LRR domains. A detailed description of the oxidative stress-related genes with the corresponding domain is given on the author's webpage (http://pfg.knu.ac.kr/publish/oxi.pdf). Some of the domains predicted in oxidative stress-related genes indicate their involvement in metabolic processes under favorable and stable environments.

Overexpression of these genes is an indication of oxidative stress. As a result, 108 genes in Arabidopsis and 175 genes in soybeanswere predicted to encode a peroxidase domain, while 40 genes in Arabidopsis and 64 genes in soybeans were predicted to encode a thioredoxin domain (Table 1).

To identify the phylogenetic relationship between the genes, a multiple alignment analysis was performed using the amino acid sequences. The alignment indicated conserved residues for the corresponding domains (Figure 1). Based on these

observations, a phylogenetic alignment was then constructed for Arabidopsis and Soybeans. As shown in Figure 2, the phylogram distinguished 5 groups, namely A to E. The peroxidase domain-harboring oxidative stress-related genes were grouped in clad A. The NJ tree showed that the oxidative stress-related genes in group A could be further divided into 3 subgroups, referred to as 1, 2, and 3. Groups B and C were also divided into two subgroups. The genes in group B corresponded to the thioredoxin domain. Meanwhile, the genes

Table 1. Identified oxidative stress-related genes with their corresponding domains

| Protein motif in oxidative stress-related genes	Number of genes harboring corresponding motifs	
	Arabidopsis	Soybean
Peroxidase	108	175
Thioredoxin	44	64
Kinases	4	18
Glutathione S-transferase family	9	3
Leucine Rich Repeats	3	1
Hydrolase	4	6
Reductase	12	6
Aldehyde dehydrogenase	4	2
The Basic Leucine Zipper	3	8
Zinc Finger	13	7
ATPases Associated with diverse cellular Activities	9	13
Glutamine Synthetase	2	23
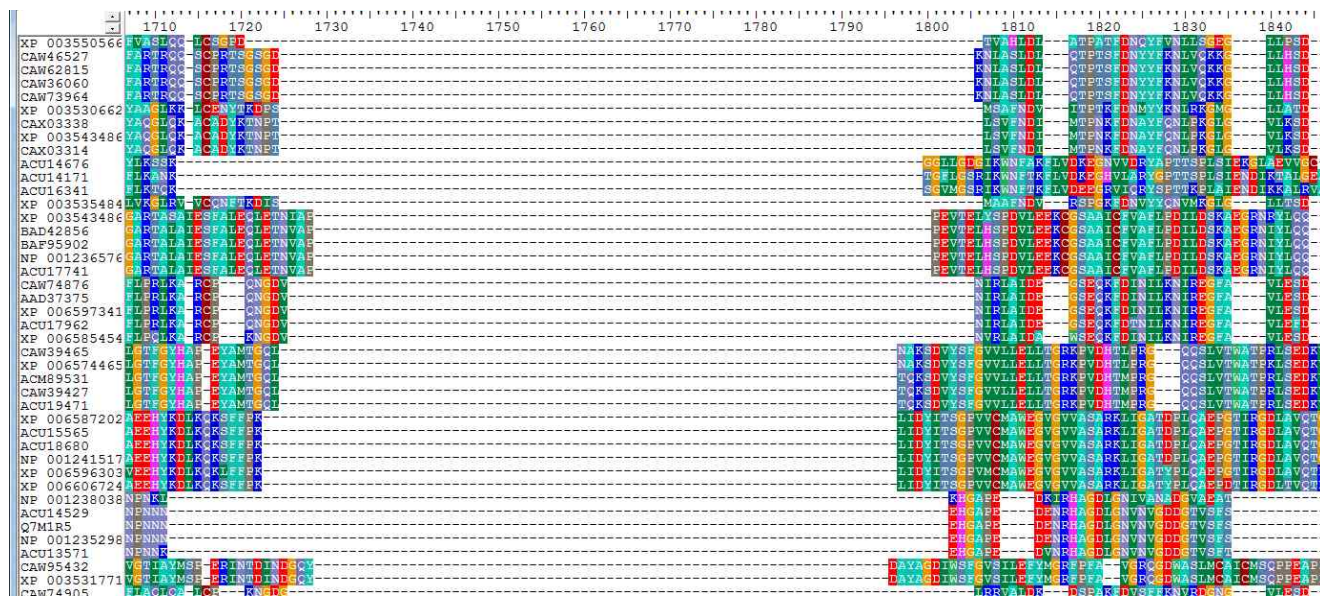
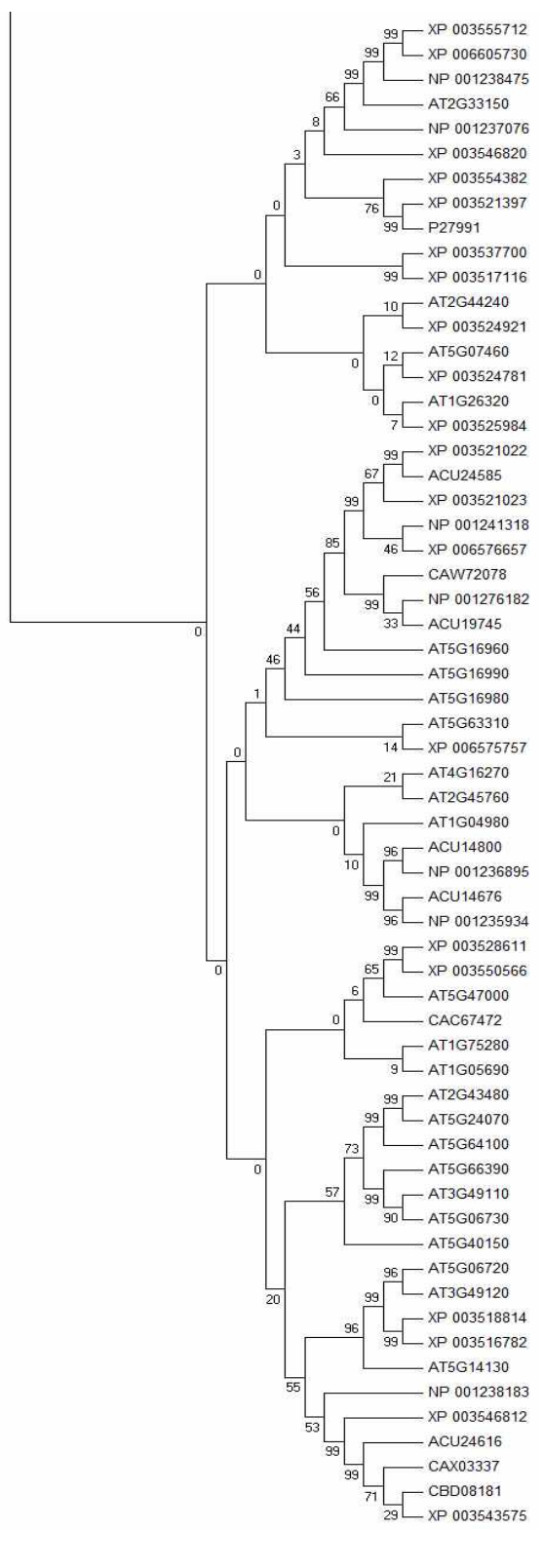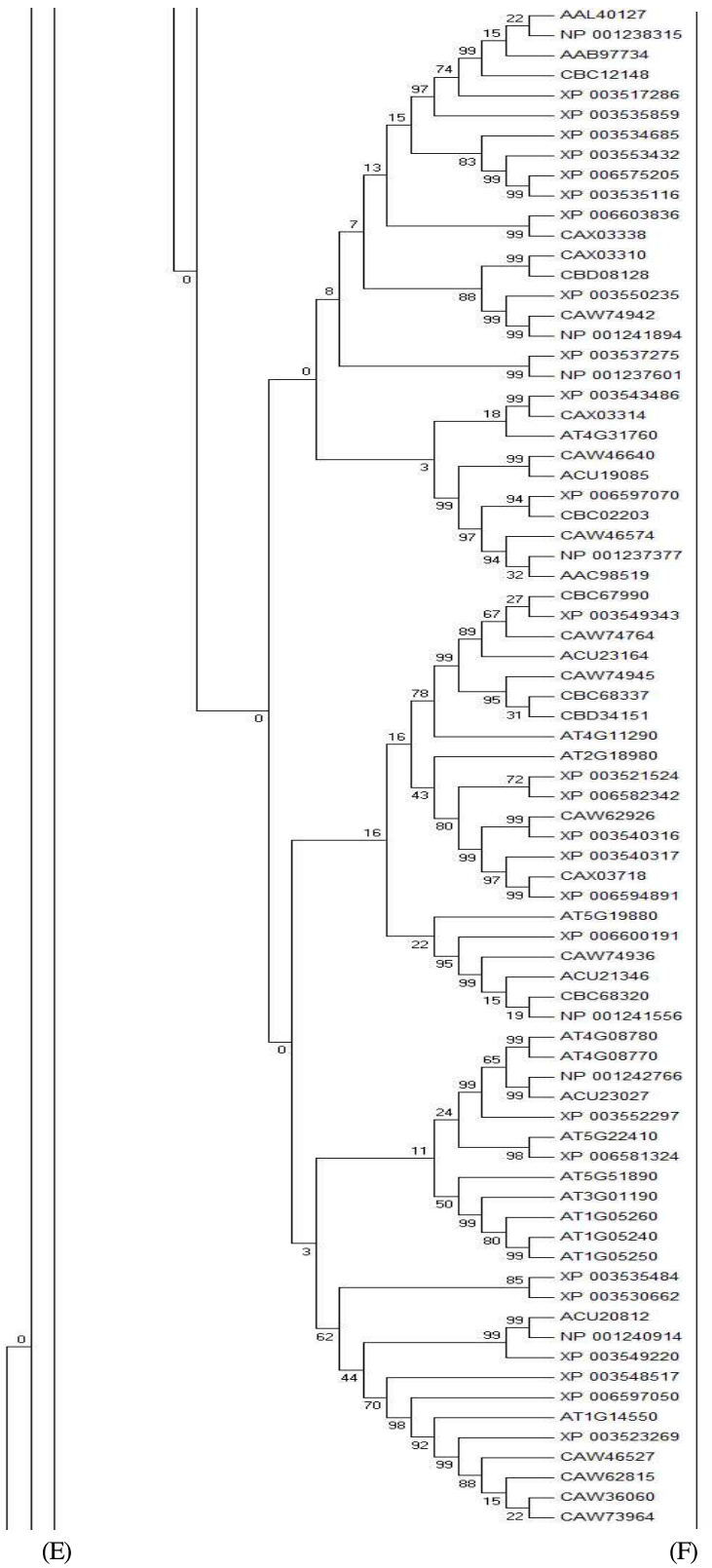**Figure 1. Amino acid conservation within signature sequence of peroxidase family in soybean proteins**

(A)   (B)

(C)

(D)

(E)   (F)
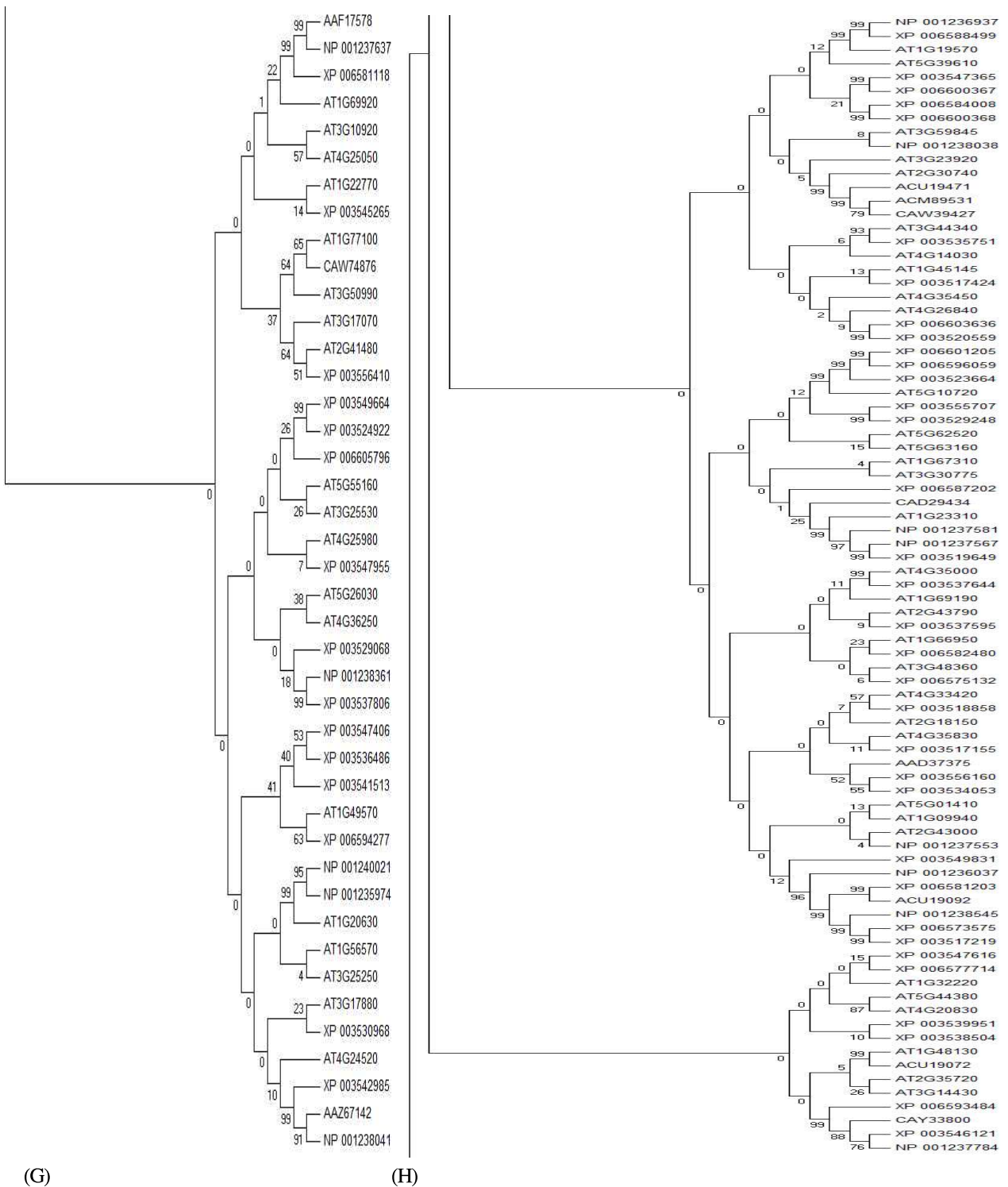
(G)   (H)
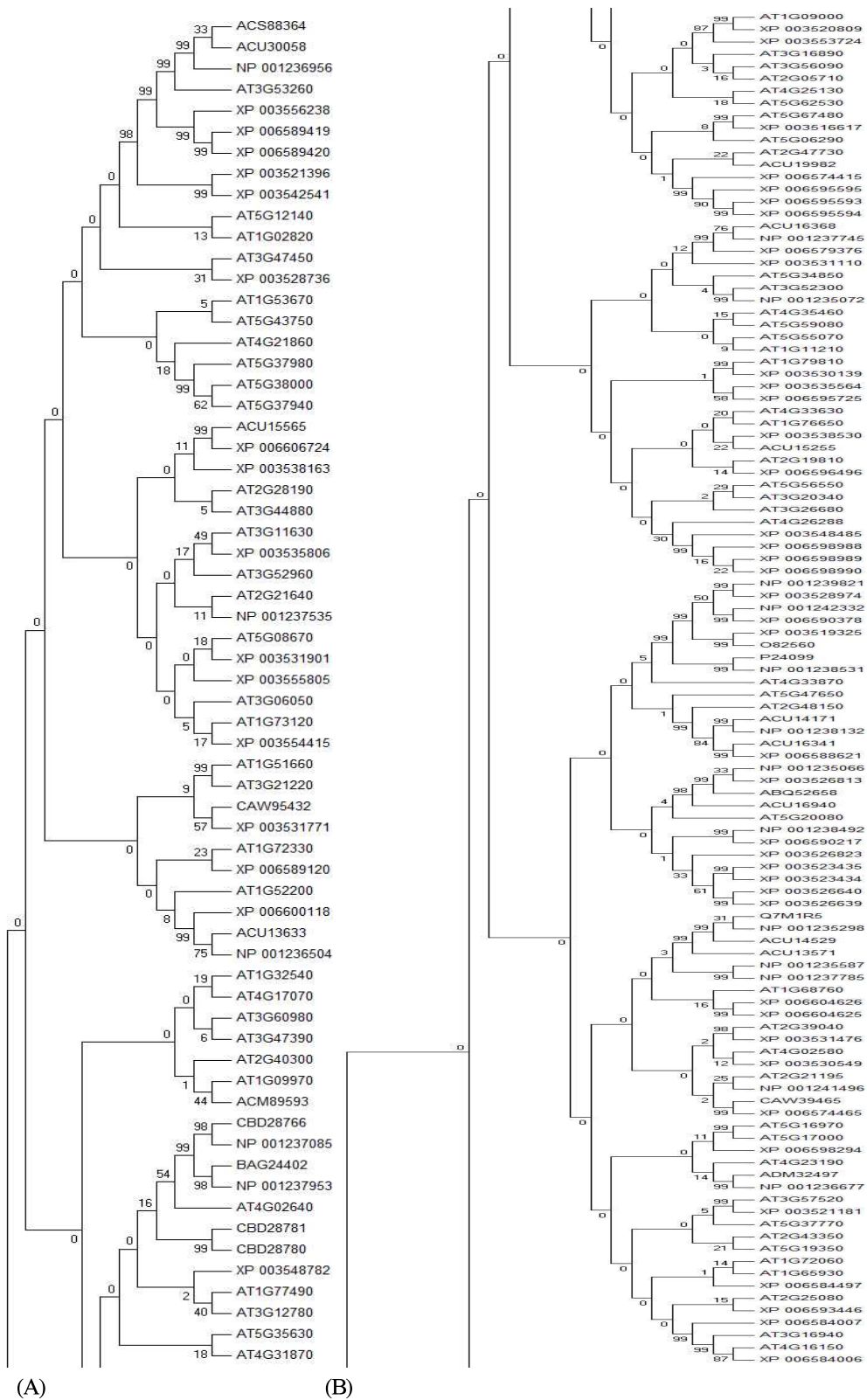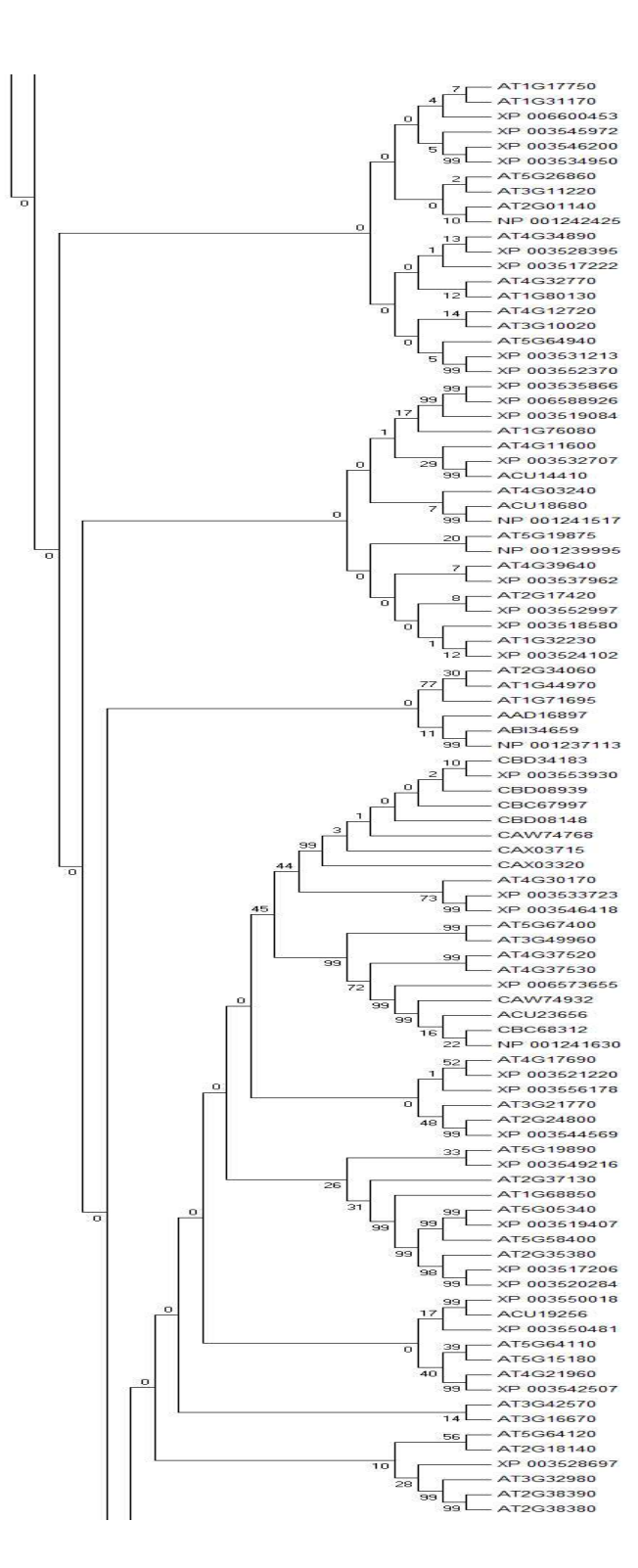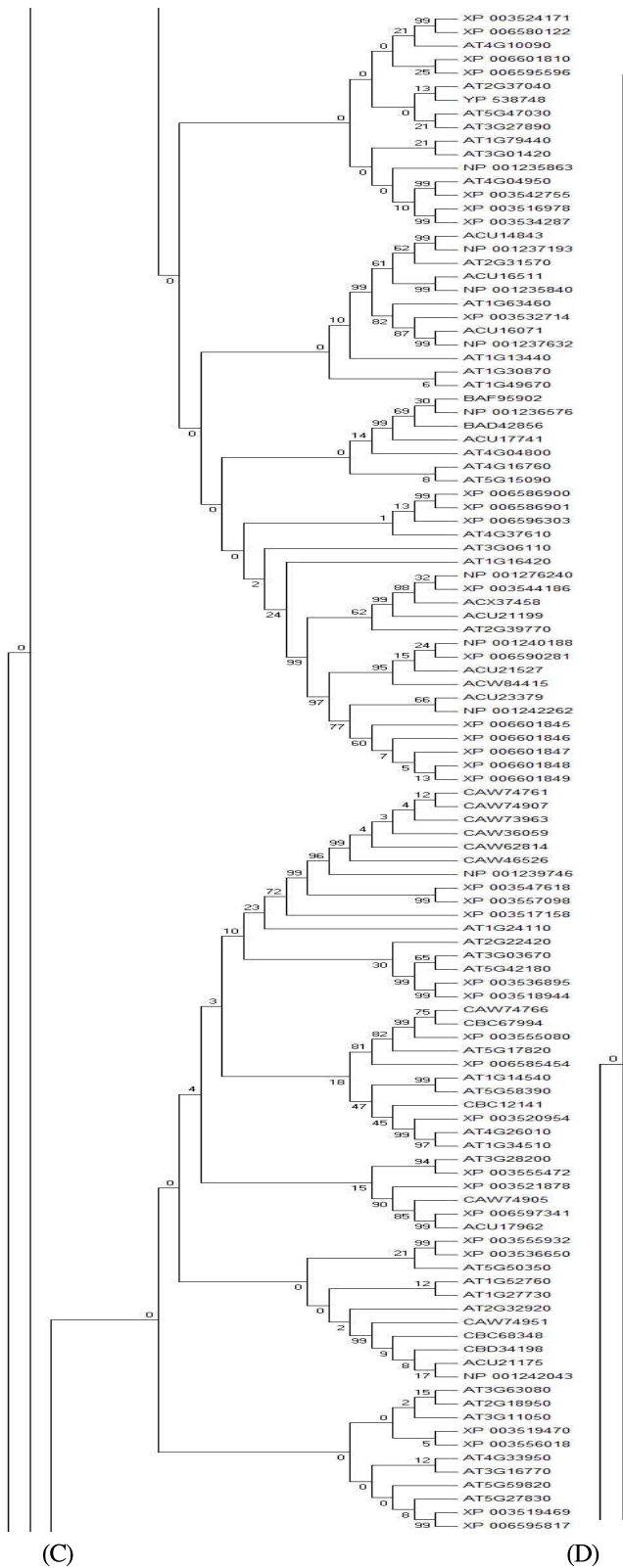
Figure 2. Phylogenetic analysis (A-H). The sequences determined were aligned using MEGA program version 6.0 with grouping based on NJ method using p-distance matrix for nucleotides, with pairwise gap deletion option adopted with 1000 bootstrap repetitions.

in clads C and D corresponded to multiple domains and were further divided into 3 subgroups. Clad E also harbored a kinase, as well as other ion-specific domains with a further 2 subgroups. All clads have included both soybean and Arabidopsis genes. The higher number of groups and multiple subgroups suggested a higher level of diversity among the oxidative stress-related genes.

Discussion

The adaptation mechanisms of plants to adverse and extreme environment fluctuation are of major interest to plant biologists as regards growth and yield. Uncommon stresses such as drought, salt, cold, pathogen etc. have a common feature in plants: the generation of oxidative stress and a change in the cellular redox potential referred to as an oxidative burst. However, the molecular basis for an oxidative burst in plants is still poorly understood. The availability of full-length genome databases has now facilitated the utilization of comparative genomics in plants, and since the Arabidopsis oxidative stress responsive dataset includes 252 genes, this provides a basis for comparative genomics in plant stress physiology. These genes were exploited for identification of oxidative stress genes in soybean. The response of different plants to similar stress conditions includes similar proteins. Agalou et al. (2008) previously identified homeodomain leucine zipper (HD-Zip) genes in rice using Arabidopsis HD-Zip gene information. Thus, the HD-Zip family plays a common role in both plants as a drought responsive family.

As a first line of plant defense, the induction of a common set of stress proteins is the molecular basis for both tolerance and stress hardening. After the initial stress, these proteins remain active/elevated for a time period that varies depending on the species, cell-type, history of prior stress exposure, gene-environment interactions during development, and stress severity(Kultz 2005). Virtually every gene response to stress is also affected by modulation of the redox state or the free radical levels (Asada 2006). Thus, a number of different domain carrying genes are involved in oxidative stress, unlike other diseases (e.g. NBS domain) and particular stress responsive genes (Miller et al. 2010). The current phylogenetic analysis of the predicted dataset for Arabidopsis and soybeans showed quite diverse oxidative stress-related genes with 5 groups that were further sub-grouped depending on various motifs and other conserved sequences.

Alteration of the cellular redox potential is a major trigger for the stress response in plants. During different types of stress, cellular oxidases, such as the plasma membrane NADPH oxidase, are rapidly activated, which may explain the increased ROS levels (Miller et al. 2009). All cells have various antioxidant enzymes to minimize and repair oxidative damage. Many oxidoreductases present in the minimal stress proteome are dehydrogenases, some of which are elements of basic metabolic pathways; including glycolysis, the pentose phosphate cycle, and Krebs cycle, making them essential even in the absence of stress. However, these dehydrogenases also influence the cellular redox potential and oxidative damage repair by generating reducing equivalents for antioxidant enzymes that depend on NADPH as a cofactor, including thioredoxin reductase, glutathione reductase, and aldehyde reductase. Aldehyde dehydrogenase and aldehyde reductase are important for the detoxification of aldehydes, which are common toxic intermediary metabolites during oxidative stress. The current study identified thirodoxin, several peroxidases, oxido-reductases, dehydrogenases, and transferases, and a few transcription factors with a DNA binding domain.

Oxidative stress is the execution of a complex signaling mechanism that forces the destruction of the affected part of the plant to protect the survival of the healthy part. The oxidative gene dataset identified in the present study will help in understanding and manipulating the response of plants to biotic and abiotic stresses. Moreover, this strategy will provide an edge for inferring the molecular basis of stress specificity in different plants over traditional costly and time-consuming techniques.

Acknowledgements

References

Agalou A, Sigit P, Elin O, Henrik J, Xiaoyi Z, Amy E, Rolf J de K, Peter E, Inez H SL, Zhen Z, Mei W, Lizhong X, Annemarie HM, Pieter BFO (2008) A genome wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. *Plant Mol Biol* 66: 87-103.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.

Asada K (2006) Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiol* 141: 391-396.

Axelsson T, Shavorskaya O,Lagercrantz U (2001) Multiple

flowering time QTLs within several Brassica species could be the result of duplicated copies of one ancestral gene. *Genome* 44: 856-864.

Beligni MV, Lamattina L (2000) Nitric oxide stimulates seed germination and de-etiolation, and inhibits hypocotyl elongation, three light-inducible responses in plants. *Planta* 210: 215-221.

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. *Proc R Soc B* 274: 227-274.

Bennetzen JL,Ramakrishna W (2002) Numerous small rearrangements of gene content, order, and orientation differentiate grass genomes. *Plant Mol Biol* 48: 821-827.

Bethke PC, Libourel IGL, Aoyama N, Chung Y, Still DW, Jones RL (2007) The Arabidopsis thaliana aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. *Plant Physiol* 143: 1173-1188.

Bright J, Desikan R,Hancock JT, Weir IS, Neill SJ (2006) ABA induced NO generation and stomatal closure in Arabidopsis are dependent on H2O2 synthesis. *Plant J* 45: 113‑122.

Causse M, Damidaux R,Rousselle P (2007) Traditional and enhanced breeding for quality traits in tomato. In: Genetic improvement of Solanaceous crops; tomato Enfield, NH: Science Publishers 2: 153-192.

del-Río LA, Sandalio LM, Corpas FJ, Palma JM, Barroso JB (2006) Reactive oxygen species and reactive nitrogen species in peroxisomes. Production, scavenging, and role in cell signaling. *Plant Physiol* 141: 330‑335.

Delledonne M (2005) NO news is good news for plants. *Curr Opin Plant Biol* 8: 390-396.

Ellwood SR, Phan HTT, Jordan M, Hane J, Torres AM, Avila CM, Serafín Cruz-Izquierdo S, Oliver RP (2008) Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. *BMC Genomics* 9: 380-391

García-Mata C, Lamattina L (2001) Nitric oxide induces stomatal closure and enhances the adaptive plant responses against drought stress. *Plant Physiol* 126: 1196‑1204.

García-Mata C, Lamattina L (2007) Abscisic acid (ABA) inhibits light-induced stomatal opening through calcium- and nitric oxide-mediated signaling pathways. *Nitric Oxide* 17: 143-151.

Gechev TS, Van Breusegem F, Stone JM, Denev I, Laloi C (2006) Reactive oxygen species as signals that modulate plant stress responses and programmed cell death. *Bioessays* 28: 1091-1101.

Guo FQ, Crawford NM (2005) Arabidopsis nitric oxide

synthase1 is targeted to mitochondriaand protects against oxidative damage and dark-induced senescence. *Plant Cell* 17: 3436-3450.

Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calfon M, Ron D (2003) An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell* 11: 619-633.

He YK, Tang RH, Yi H, Stevens RD, Cook CW, Ahn SM, Jing L, Yang Z, Chen L, Guo F, Fiorani F, Jackson RB, Crawford NM, Pei ZM (2004) Nitric oxide represses the Arabidopsis floral transition. *Science* 305: 1968-1971.

Keller B, Feuillet C (2000) Colinearity and gene density in grass genomes. *Trends Plant Sci* 5: 246-251.

Kultz D (2005) Molecular and evolutionary basis of the cellular stress response. *Ann Rev Physiol* 67: 225-257.

Madhava Rao KV, Sresty TVS (2000) Antioxidative parameters in the seedlings of pigeonpea (*Cajanus cajan* (L.) Millspaugh) in response to Zn and Ni stresses. *Plant Sci* 157: 113-128.

Mahantesha N, Khader S, Oommen K. M, Ramanjini G, Ramanathan S (2013) STIFDB2: An Updated Version of Plant Stress-Responsive Transcription Factor Database with Additional Stress Signals, Stress-Responsive Transcription Factor Binding Sites and Stress-Responsive Genes in Arabidopsis and Rice. *Plant Cell Physiol* 54: 1-15.

Miller G, Schlauch K, Tam R, Cortes D, Torres MA, Shulaev V, Dangl JL, Mittler R. (2009) The plant NADPH oxidase RBOHD mediates rapid systemic signaling in response to diverse stimuli. *Science Signaling* 2: 45.

Miller GAD, Suzuki N, Ciftci-Yilmaz S, Mittler R. (2010)Reactive oxygen species homeostasis and signaling during drought and salinity stresses. *Plant Cell Environ* 33: 453-467.

Mittler R (2002) Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci* 7: 405-410.

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GR, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, *et al.* (2000) Comparative genomics of the eukaryotes. Science 287: 2204-2215.

Subhomoi B, Vijayakumar S, Amit J, *et al.* (2013) The Arabidopsis Stress Responsive Gene Database, *Int J Plant Genomics*, vol. 2013, Article ID 949564, 3 pages, 2013. doi:10.1155/2013/949564.

Suzuki N, Miller G, Morales J, Shulaev V, Torres, MA, Mittler R (2011) Respiratory burst oxidases: the engines of ROS signaling. *Curr Opin Plant Biol* 14: 691-699.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013)

MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.

Torres MA, Jones JD, Dangl JL (2005) Pathogen-induced, NADPH oxidase-derived reactive oxygen intermediates suppress spread of cell death in *Arabidopsis thaliana. Nature Genet* 37: 1130-1134.

Uchida A, Jagendorf AT, Hibino T, Takabe T, Takabe T (2002) Effects of hydrogen peroxide and nitric oxide on both salt and heat stress tolerance in rice. *Plant Sci* 163: 515-523.

Yun BW, Feechan A, Yin M, Saidi NB, Le Bihan T, Yu M et al. (2011) S-nitrosylation of NADPH oxidase regulates cell death in plant immunity. *Nature* 478:264-268.

Zang YX, Kim HU, KimJA, Lim MH, Jin M, Lee SC, Kwon SJ, Lee SI, Hong JK, Park TH, Mun JH, Seol YJ, Hong SB, Park BS (2009) Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa. FEBS J* 276: 3559-3574.

Zhang Y, Wang L, Liu Y, Zhang Q, Wei Q, Zhang W (2006) Nitric oxide enhances salt tolerance in maize seedlings through increasing activities of proton-pump and Na+/H+ antiport in the tonoplast. *Planta* 224: 545-555.

Zhao L, Zhang F, Guo J, Yang Y, Li B, Zhang L (2004) Nitric oxide functions as a signal in salt resistance in the calluses from two ecotypes of reed. *Plant Physiol* 134: 849-857.