

## ***In silico* comparative analysis and expression profile of antioxidant proteins in plants**

S. Sheoran<sup>1</sup>, B. Pandey<sup>1</sup>, P. Sharma<sup>1</sup>, S. Narwal<sup>2</sup>, R. Singh<sup>1</sup>, I. Sharma<sup>4</sup> and R. Chatrath<sup>3</sup>

<sup>1</sup>Plant Biotechnology, Directorate of Wheat Research, Karnal, India

<sup>2</sup>Quality and Basic Sciences, Directorate of Wheat Research, Karnal, India

<sup>3</sup>Crop Improvement, Directorate of Wheat Research, Karnal, India

<sup>4</sup>Project Director, Directorate of Wheat Research, Karnal, India

Corresponding author: S. Sheoran

E-mail: sheoran\_sonia@yahoo.co.in

Genet. Mol. Res. 12 (1): 537-551 (2013)

Received March 5, 2012

Accepted October 7, 2012

Published February 27, 2013

DOI <http://dx.doi.org/10.4238/2013.February.27.3>

**ABSTRACT.** The antioxidant system in plants is a very important defensive mechanism to overcome stress conditions. We examined the expression profile of antioxidant enzymes superoxide dismutase (SOD), catalase (CAT) and ascorbate peroxidase (APX) using a bioinformatics approach. We explored secondary structure prediction and made detailed studies of signature pattern of antioxidant proteins in four plant species (*Triticum aestivum*, *Arabidopsis thaliana*, *Oryza sativa*, and *Brassica juncea*). Fingerprinting analysis was done with ScanProsite, which includes a large collection of biologically meaningful signatures. Multiple sequence alignment of antioxidant proteins of the different plant species revealed a conserved secondary structure region, indicating homology at the sequence and structural levels. The secondary structure prediction showed that these proteins have maximum tendency for  $\alpha$  helical structure. The sequence level similarities were also analyzed with a phylogenetic tree using neighbor-joining method. In the antioxidant enzymes SOD, CAT and APX, three major families of signature were predominant and common; these were PKC\_PHOSPHO\_SITE, CK2\_PHOSPHO\_SITE and N-myristoylation

site, which are functionally related to various plant signaling pathways. This study provides new strategies for screening of biomodulators involved in plant stress metabolism that will be useful for designing degenerate primers or probes specific for antioxidant. These enzymes could be the first line of defence in the cellular antioxidant defence pathway, activated due to exposure to abiotic stresses.

**Key words:** Antioxidant proteins; Comparative analysis; Signature pattern; Evolutionary trend

## INTRODUCTION

In plants, several enzymes act jointly to maintain redox homeostasis. In addition to scavenging reactive oxygen species (ROS) produced during regular metabolic processes, antioxidant metabolism also has a major role in plant protection against stressful environmental conditions that induce ROS production and accumulation (Dat et al., 2000). Plants possess very efficient enzymatic antioxidant defense systems, which work in concert to control the cascades of uncontrolled oxidation (superoxide dismutase, SOD; catalase, CAT; ascorbate peroxidase, APX; glutathione reductase; monodehydroascorbate reductase; dehydroascorbate reductase; glutathione peroxidase; guaiacol peroxidase, and glutathione-S-transferase). SODs (EC 1.15.1.1) are a family of metalloenzymes that constitute the frontline defense by catalyzing the dismutation reaction of toxic superoxide radicals to molecular oxygen and hydrogen peroxide, thus forming a crucial part of the cellular antioxidant defense mechanism by preventing the oxidation of various biological molecules, either by the radicals themselves or by their derivatives (Bowler et al., 1992, 1994). In *Arabidopsis*, 7 SOD genes, including 3 CuZnSODs, 1 MnSOD, and 3 FeSODs, have been isolated through analysis of large numbers of cDNA and genomic DNA sequences (Delseny et al., 1997; Kliebenstein et al., 1998). Five wheat SOD genes including 2 CuZnSODs (SOD1.1 and SOD1.2) (Wu et al., 1996) and 3 MnSODs [SOD3.1 and SOD3.2 (Wu et al., 1997) and another with GenBank accession No. AF092524] have been isolated and characterized. Based on the metal cofactor, SOD enzymes are classified into FeSOD in chloroplasts, CuZnSOD in chloroplasts and cytosol, and MnSOD in the mitochondria and peroxisomes (Kanematsu and Asada, 1990). Catalase is essential for the removal of H<sub>2</sub>O<sub>2</sub> produced by photorespiration in the peroxisomes (Noctor et al., 2000). Plant catalases are encoded by a small gene family of 3 isozymes, which exhibit fairly complex spatial and temporal patterns of expression throughout the plant life cycle (Scandalios et al., 1997; Willekens et al., 1997). Ascorbate peroxidase plays a major role in scavenging H<sub>2</sub>O<sub>2</sub> in chloroplast cytosol and glyoxysomes (Asada, 2006). The chloroplastic APXs, including thylakoid-bound and stromal isoforms, scavenge the H<sub>2</sub>O<sub>2</sub> produced during photosynthesis. In comparison to the cytosolic and peroxisomal isoforms, chloroplastic enzymes are more labile and are rapidly inactivated in the absence of ascorbate (Miyake et al., 1993; Ishikawa et al., 1998; Shigeoka et al., 2002).

Highly conserved sequences in protein families are generally important for protein function and/or for maintenance of its 3-dimensional structure. These motifs, typically around 10 to 20 amino acids in length, arise because specific residues and regions thought or proven to be important to the biological function of a group of proteins are conserved in both structure and sequence during evolution. As the sequence of biologically meaningful motifs is evolution-

arily conserved, a multiple alignment can be reduced to a consensus expression called a regular expression or pattern (Pei et al., 2008). Protein signatures are efficient mining tool capable of identifying related protein sequences having the same functional residues, belonging to the same class of proteins from the abundant sequences in the non-redundant databases. Over the last decade, the sensitivity of sequence searching techniques has been improved by profile-based analysis, which uses information derived from multiple sequence alignments to construct and search for sequence patterns (Pei et al., 2008). Plants with the ability to scavenge and/or control cellular ROS may be useful in the future for their ability to withstand harsh environmental conditions. Computer-assisted characterization of the features of the proteins found or predicted in complete sequence proteomes is an important task in the search for knowledge of protein function. This study aimed to identify conserved motifs that are expressed in different plant species. Further *in silico* structural and functional analysis of sequences can provide insight into the correlation between the antioxidant gene family and plant oxidative stress with respect to certain specific conserved motifs, and may facilitate screening of biomodulators of plant stress metabolism.

## MATERIAL AND METHODS

The plants selected for this study were *Triticum aestivum*, *Arabidopsis thaliana*, *Oryza sativa*, and *Brassica juncea*. The sequences of the selected antioxidant proteins were retrieved from the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>). Only full-length protein sequences were considered for *in silico* analysis. The retrieved data were submitted to the GOR IV secondary structure prediction online software (<http://abs.cit.nih.gov/gor/>). GOR is based on information theory and has a mean accuracy of 64.4% for a 3-state prediction. The data obtained were submitted for further analysis and the curated data in its FASTA format were submitted to ScanProsite (<http://www.expasy.org/tools/scanprosite/>) to search for fingerprints. From the ScanProsite results, conserved motifs with a minimum length of 8 amino acids were considered for further analysis. The BLAST program was used to search orthologous sequences of selected antioxidant proteins in other plant species for evolutionary analysis. In order to infer the phylogenetic relationship between members of the antioxidant gene family, the protein multiple sequence alignment was performed using MUSCLE (Edgar, 2004) and T-Coffee (Notredame et al., 2000). The final alignment was completed through manual editing. The FASTA multiple sequence alignment was used to infer the neighbor-joining (NJ) phylogenetic tree (Saitou and Nei, 1987) using the MEGA 4.0 software (Tamura et al., 2007). The bootstrap interior branch test, superior to bootstrap analysis in testing confidence in a given tree node (Zvelebil and Baum, 2008), was applied to test the statistical significance of the antioxidant gene family NJ phylogenetic tree using 1000 replicates. Evolutionary distances were calculated following the Poisson correction evolutionary model, and excessive gaps/missing data were handled using the pairwise deletion option. The method considers interior branch lengths and gives the confidence probability levels of each node in the given tree topology.

## RESULTS AND DISCUSSION

### Secondary structure analysis

Antioxidant enzymes comprise an interesting family of proteins, members of which

play critical roles in diverse biological processes in plants. A total of 29 SOD, 11 CAT, and 16 APX protein sequences from 4 different plant species were considered for *in silico* analysis. The percentage of occurrence of secondary structures in all the antioxidant proteins is shown in Table 1. The results revealed that in CuZnSOD, the random coil percentage dominated with 62% in *A. thaliana*, followed by  $\beta$ -sheet and  $\alpha$ -helix in all selected plants. For MnSOD,  $\alpha$ -helix predominated with 46.76% in *O. sativa* and 41.85% in *A. thaliana*, followed by random coil and  $\beta$ -sheet. In FeSOD, the  $\alpha$ -helix predominated (52.48%) in *A. thaliana*, while in *O. sativa* the random coil percentage (51.20%) was higher. This finding was consistent with Bowler et al. (1992) who found that MnSOD and FeSOD are structurally similar while CuZnSOD is structurally unrelated. In the catalase enzyme, the random coil percentage was higher in all selected plant species, ranging from 61.48% (*B. juncea*) to 58.29% (*O. sativa*). A similar pattern was apparent in ascorbate peroxidase, in which the random coil percent was highest (49.0 to 51.73%), while the occurrence of  $\beta$ -sheet was lower (11.45% in *A. thaliana* to 6.53% in *T. aestivum*). The results showed the tendency of these proteins to have  $\alpha$ ,  $\beta$  and random coil structures, but the frequency of these secondary structures varies between antioxidants in the selected plant species. Modern protein secondary structure prediction methods exploit the evolutionary information contained in multiple sequence alignments. Therefore, identifying secondary structural elements can help guide sequence alignment of distantly related sequences (Chothia and Lesk, 1986).

**Table 1.** Comparative analysis of secondary structures of antioxidant proteins.

Class	Organism	GOR IV (%)		
		$\alpha$ -helix (%)	$\beta$ -sheet (%)	Random coils (%)
CuZnSOD	<i>Triticum aestivum</i>	15.50	29.22	55.28
	<i>Arabidopsis thaliana</i>	13.15	24.4	62.45
	<i>Oryza sativa</i>	12.52	30.12	57.36
	<i>Brassica juncea</i>	2.63	41.45	55.92
MnSOD	<i>T. aestivum</i>	42.86	14.72	42.42
	<i>A. thaliana</i>	41.85	18.88	39.27
	<i>O. sativa</i>	46.76	14.43	38.81
FeSOD	<i>A. thaliana</i>	52.48	11.11	36.40
	<i>O. sativa</i>	34.93	13.87	51.20
CAT	<i>T. aestivum</i>	17.14	22.61	60.25
	<i>A. thaliana</i>	22.16	18.40	59.44
	<i>O. sativa</i>	19.32	22.39	58.29
	<i>B. juncea</i>	21.24	17.28	61.48
APX	<i>T. aestivum</i>	44.06	6.53	49.41
	<i>A. thaliana</i>	39.55	11.45	49.00
	<i>O. sativa</i>	41.51	8.19	50.30
	<i>B. juncea</i>	40.40	7.87	51.73

## Fingerprint analysis

ScanProsite was used for fingerprinting analysis; it is a large collection of biologically meaningful signatures that are described as patterns (regular expressions), used for short motif detection, or generalized profiles (weight matrices) for sensitive detection of larger domains. Each signature is linked to detailed annotation that provides useful biological information on the protein family, domain, or functional site identified by the signature. After detailed analysis of different classes of antioxidant proteins we found the sequences belonging to the following families: PKC\_PHOPHO\_SITE (protein kinase C (PKC) phosphorylation site),

CK2\_PHOSPHO\_SITE (casein kinase II (CKII) phosphorylation site), and N-myristoylation (N-MYR) site. These sites were present in all 3 antioxidant enzymes of the selected plant species (see [Table S1](#)). These families have a more or less significant relationship with the antioxidant roles in the cell signaling, regulation, and metabolism. In CuZnSOD, the families PKC\_PHOPHO\_SITE, N-myristoylation site, and CK2\_PHOSPHO\_SITE were of 3, 6 and 4 amino acids in length and present in all 4 plant species (Table 2). The PKC\_PHOPHO\_SITE sequence was variable at the 2nd position and more or less conserved at the 1st and 3rd positions in *T. aestivum*, *A. thaliana*, and *O. sativa*, while this signature pattern was absent in *B. juncea*. Plant PKC is involved in many aspects of cellular regulation and metabolism (Stone and Walker, 1995). The CK2\_PHOSPHO\_SITE sequence of 4 amino acids was found conserved in *T. aestivum* and *O. sativa*, while its signature pattern was variable at the 2nd and 3rd positions in the other plant species. In plants, both monomeric and oligomeric forms of CKII have been purified. *Arabidopsis* subunits have demonstrated formation of a tetrameric  $\alpha 2\beta 2$  complex and the regulatory  $\beta$  subunit stimulates catalytic activity (Riera et al., 2001). CKII appears to be involved in circadian clock regulation in *A. thaliana* (Takahashi et al., 2001) and photoperiod sensitivity in rice (Resh, 1999). The signature pattern of N-myristoylation site was variable in all selected plant species. N-myristoylation is a form of lipid modification that targets a wide variety of eukaryotic proteins and plays important roles in cell physiology. In many instances, N-myristoylation alters the lipophilicity of the target protein and facilitates its interaction with membranes, thereby affecting its subcellular localization (Andriotis and Rathjen, 2006; Batistic et al., 2008; Benetka et al., 2008). Within *Arabidopsis*, N-myristoylation of several proteins mediates plant rescue metabolism to survive damage produced by external unstable conditions. N-MYR in plant cells have a critical role in controlling membrane signaling pathways that lead to specific plant immunity (Hanks, 2003). In MnSOD, the PKC\_PHOPHO\_SITE signature was 3 amino acids long and conserved only in *A. thaliana* and *O. sativa*. Phosphorylation by protein kinases is a major mechanism by which virtually every eukaryotic cellular activity is regulated, including proliferation, gene expression, metabolism, motility, membrane transport, and apoptosis (Katyshev et al., 2006). The CK2\_PHOSPHO\_SITE signature was conserved in *T. aestivum* and *A. thaliana*. Variables were found at the 3rd, 4th, 5th, and 6th positions of the N-myristoylation site in *T. aestivum*, *A. thaliana*, and *O. sativa* and the signature was not present in *B. juncea*. Ishitani et al. (2000) demonstrated the essential role of protein N-myristoylation in plant salt tolerance. The AMP\_PHOSPHO\_SITE was more or less conserved in *T. aestivum*, *A. thaliana*, and *O. sativa*. In FeSOD, the PKC, CK2\_PHOSPHO\_SITE, N-myristoylation site, and AMP\_PHOSPHO\_SITE signatures were absent in *T. aestivum* and *B. juncea*, but were more or less conserved in *O. sativa* and *A. thaliana*. The most intriguing subfamily of plant SODs are FeSODs. To date, no convincing direct experimental evidence has been provided for the existence of FeSOD genes in extensively studied plant species such as *T. aestivum* and *Zea mays* (del Río et al., 1985). In higher plants, CuZnSODs and MnSODs are uniformly distributed throughout the plant, but FeSODs were thought to exist only in prokaryotic organisms and tree species from the Ginkgoaceae, Nymphaeaceae, and Cruciferae (Murao et al., 2004). However, more recently, genes belonging to FeSOD have been isolated from angiosperm species that lack close phylogenetic relationships to these families such as tobacco, *A. thaliana*, rice, and soybean (Hidalgo et al., 2001). In catalase, PKC\_PHOPHO\_SITE, CK2\_PHOSPHO\_SITE, N-myristoylation, and TYR\_PHOSPHO\_SITE were present with conserved sequences of 3, 4, 7, and 6 amino acids in length in all selected plant species.

**Table 2.** Fingerprint analysis for different antioxidant enzymes.

Signature	Class	Plant species	Sequence
PKC_PHOSPHO_SITE	CuZnSOD	<i>Triticum aestivum</i> <i>Arabidopsis thaliana</i> <i>Oryza sativa</i>	[S/T][A/V/K/F][R/K] T[L/G][K/R] [T/S][G/A/I/K][K/R]
CK2_PHOSPHO_SITE		<i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i>	[S/T][Q/A/G/K/S/Y][V/E/S/Y][E/D] [S/T][D/Q/I/F/S/V][V/D/G/A/T][E/D] [T/S][A/G/Q/I][G/K/V/D][E/D]
N-myristoylation site		<i>Brassica juncea</i> <i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i> <i>A. thaliana</i> <i>O. sativa</i>	T[V/I][G/T]D G[G/A/L/N/V/S/M][A/T/S/T/R/Q/D/L/F][A/P/T/H/V/E/G/L/R/S][A/G/A/T/G][A/N/L/R/T/F/Q/S/I] G[V/N/G/L/T/L/M][S/A/R/L/T/A/V/K][N/G/E/G/L/R/P/V/S/T][N/G/A/T/S][N/T/R/V/F/L] G[A/L/N/V/G/M/S][A/T/S/R/I/K/V][A/P/T/V/E/G/L/S][A/G/T][R/L/N/T/I] G[V/N/T/L/M/G][K/G/V/S/A/R][G/T/S/P/V/X][T/G/A][T/L/R]
PKC_PHOSPHO_SITE	MnSOD	<i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i> <i>A. thaliana</i> <i>O. sativa</i>	S[R/S][K/R] SRK [T/S][I/K][P/G]D [T/S][I/A][P]D
CK2_PHOSPHO_SITE		<i>T. aestivum</i> <i>A. thaliana</i> <i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i>	G[L/G][A/H/L][L/R/V/Q][G/N][G/V/H/S] G[L/G][K/H/L][E/V/G][T/N/S/G][S/H/A/L] G[G/S][H/F][V/E][N/A][H/L]
N-myristoylation site		<i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i> <i>A. thaliana</i> <i>O. sativa</i>	K[K/R][I/M]S KKMS KK[M]JS S[R/S][K/R]
AMP_PHOSPHO_SITE		<i>A. thaliana</i> <i>O. sativa</i> <i>A. thaliana</i> <i>O. sativa</i>	S[T][P/L/R/M/A][R/K] SR
PKC_PHOSPHO_SITE	FeSOD	<i>A. thaliana</i> <i>O. sativa</i> <i>A. thaliana</i> <i>O. sativa</i>	[T/S][Y/D/L/F/R/E/M/S][V/L/E/M/T/D/M/Y][E/D] TIEE
CK2_PHOSPHO_SITE		<i>A. thaliana</i> <i>O. sativa</i> <i>A. thaliana</i> <i>O. sativa</i>	G[T/V/G/S/A][K/I/G/F][V/T/P/W][A/T/N/S][V/W/F/G] G[G/S][S/F][P/T][G/N][R/F]
N-myristoylation site		<i>A. thaliana</i> <i>O. sativa</i> <i>A. thaliana</i> <i>O. sativa</i>	RRS RRYS
AMP_PHOSPHO_SITE		<i>A. thaliana</i> <i>O. sativa</i> <i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i>	[S/T][A/V/S/Q][K/R] [S/T][A/I/Q/D/G][K/R] [S/T][A/L/P/Q][I][K/R] [S/T][A/L/G][K/R]
PKC_PHOSPHO_SITE	CAT	<i>B. juncea</i> <i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i>	[T/S][C/L/W/Y/R/Q/K][A/R/P/V/C/S/L/N][D/E] [S/T][S/A/L/D/Y/Q/W][H/A/R/E/L/P][D/E] [T/S][V/C/L/W/Y/R/Q/S][G/A/R/P/F/C][D/E] [T/S][C/L/W/Y/R/Q][A/R/L/P/Y][D/E]
CK2_PHOSPHO_SITE		<i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i>	G[A/S/T/Q/V/M][N/P/C/H/K/G/N/Y/T][S/E/H/Y/L/V/Y/C][A/T/N/S][Y/L/H/D/S/T/I] G[S/A/Q/T][P/G/N/K/H][E/V/H/E/V/Y][T/N/S/A][L/T/H/D/S]
N-myristoylation site		<i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i>	G[S/T/Q/V][N/P/C/H/W/K/S/Y][S/E/L/H/Y/I][T/N/S/A][F/L/Y/H/D/S/N] G[S/T/Q][P/G/N/H/K][E/V/H/Y/L][T/N/S/A][L/T/H/D/S]

Continued on next page

Table 2. Continued.

Signature	Class	Plant species	Sequence
TYR_PHOSPHO_SITE		<i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i>	Rde.Evn.Y Rde.Evn.Y R[D/N][E,Y]PE[V,W][N,K][L/Y]Y
PKC_PHOSPHO_SITE	APX	<i>B. juncea</i> <i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i>	[S/T][T/R/A/L/F/Y/M][R/K] [S/T][D/H/P/T/K/M/L/F/G/R/A][R/K] [S/T][S/M/D][K/R] [S/T][A/Y/C/L/R][R/A/P/D][E/D]
CK2_PHOSPHO_SITE		<i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i>	[S/T][Y/D/G/K/F][A/K/F/S/E][D/E] [T/S][P/Y/D/G][A/Q/E][E/D] [S/T][V/F/D/G][V/A/Q][E/D]
N-myristoylation site		<i>T. aestivum</i> <i>A. thaliana</i> <i>O. sativa</i> <i>B. juncea</i>	G[G/A/L/Q][A/G/D/N/T/S/V/S][G/A/D/N/Y/W][A/G/S/T/N][A/G/S/L/D/K/E] G[G/A/S/L/V][A/N/K/F/H/W/D/V/S][N/G/A/V/F/D/I/Y][G/S/A/C/S][S/L/I/A/Q/T/N/V] G[G/L/A/C/NV][P/D/Q/W/T/I/A/R/F/N][F/I/M/T/P/S/V/E/G/L][G/A/S][T/V/L/N/R/C] G[G/L/A/C/NV][P/D/H/W/I/A/R/F/N][F/I/D/T/P/S/V/E/G/L][C/G/A/S][T/L/N/R] [R/K][R][K][D]S
CAMP_PHOSPHO_SITE		<i>O. sativa</i> <i>B. juncea</i> <i>A. thaliana</i>	

SOD = superoxide dismutase; CAT = catalase; APX = ascorbate peroxidase.

The PKC phosphorylation signature was variable at the 2nd position and present in all selected plants. PKC enzymes play important roles in several signal transduction cascades. In plants, protein phosphorylation has been implicated in responses to many signals, including light, pathogen invasion, hormones, temperature stress, and nutrient deprivation. The activities of several plant metabolic and regulatory enzymes are controlled by reversible phosphorylation of PKCs (Stone and Walker, 1995). The 4-amino acid sequence of the CKII protein is known to be activated by salicylic acid in tobacco (La Du and Zannoni, 1955) and this sequence was variable at the 2nd and 3rd positions in the selected plant species. Another family found in antioxidant enzyme catalase was tyrosine kinase, which was 7 amino acids in length and highly conserved in *T. aestivum*, *A. thaliana*, and *B. juncea*. Moreover, in catalase, the signature pattern of the TYR\_PHOSPHO\_SITE was found in all selected plant species. This finding is consistent with previous reports in which catalase is required in one of the steps of tyrosine metabolism, namely, the conversion of p-hydroxy-phenylpyruvic acid to homogentisic acid (Moorhead et al., 2009). *A. thaliana* has a few protein Tyr-specific phosphatases and 22 dual-specificity phosphatases (Spiteri et al., 1989; Rayapureddi et al., 2005). The role of tyrosine phosphatases in stomatal regulation provides critical evidence that tyrosine phosphatases not only exist but also play an important role in higher plants. In ascorbate peroxidase, the PKC signature was well conserved in *O. sativa* and *B. juncea* and variable at the 2nd position in *T. aestivum* and *A. thaliana*. The conserved sequences of CKII were variable at the 2nd and 3rd positions. The cAMP- and cGMP-dependent protein kinase phosphorylation signature of 4 amino acids was observed only in *A. thaliana*. Reports of cAMP in plants have been disputed (Zamocky et al., 2008) although cGMP may function in phytochrome responses (Van Camp et al., 1997). This study provides clear evidence of the structural and functional correlation of different plant antioxidant protein classes in relation to plant metabolism.

### Sequence alignment and phylogenetic analysis

Sequence alignment of CuZnSOD (Figure 1A) indicated consensus secondary structure elements with 4  $\alpha$ -helices and 4  $\beta$ -sheets, and 2 conserved histidine residues that bind a copper atom. The C-terminal domain (residues 1 to 80) consists of a  $\beta\beta\alpha\alpha$  structural motif, which contributes to the framework stability and specificity of the protein fold and dimer assembly. In the Mn/FeSOD structures, the polypeptide chain is divided into N-terminal helices and a C-terminal  $\alpha/\beta$  domain. The active site of Mn/FeSOD is located between the N- and C-terminal domains, and differs from that of CuZnSOD by containing a single metal ion. In MnSOD, D (asparagine) and H (histidine) manganese ligands were conserved in the C-terminus in all selected plant species (Figure 1B) while in FeSOD, the iron ligands were conserved in *Arabidopsis* and *O. sativa* (Figure 1C). Nine  $\alpha$ -helix and 7  $\beta$ -sheets were well conserved in the CAT sequence and alignment revealed a conserved C-terminal tyrosine, which serves as the heme proximal side ligand. The N-terminal domain of catalase is well conserved and contains the conserved histidine that is crucial for catalytic activation (Gratao et al., 2006) (Figure 2A). In APX, a more variable C-terminal domain (domain II, approximately 2/3 of the protein) that is entirely helical suggests that differences in the C-terminal domain may be responsible for differences in substrate specificity between antioxidant classes. Ten  $\alpha$ - and only one  $\beta$ -sheet had a conserved histidine residue, which serves as an acid-base catalyst in the reaction with hydrogen peroxide (Figure 2B). Protein secondary structure prediction remains an important step in full-tertiary structure prediction in computational biology.





**Figure 1.** Amino acid sequence alignment of antioxidant proteins. **A.** CuZnSOD, **B.** MnSOD and **C.** FeSOD of selected plants. Alpha helices and beta strands are represented as rods and arrows. Conserved metal binding residues in plants are shown by shaded region.

A



B

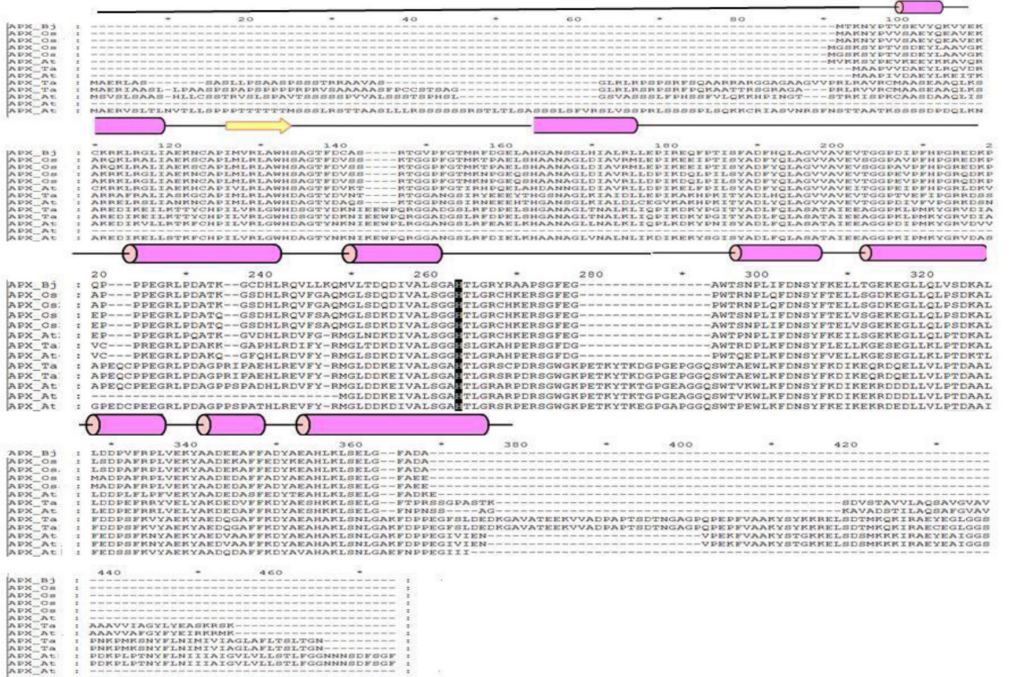
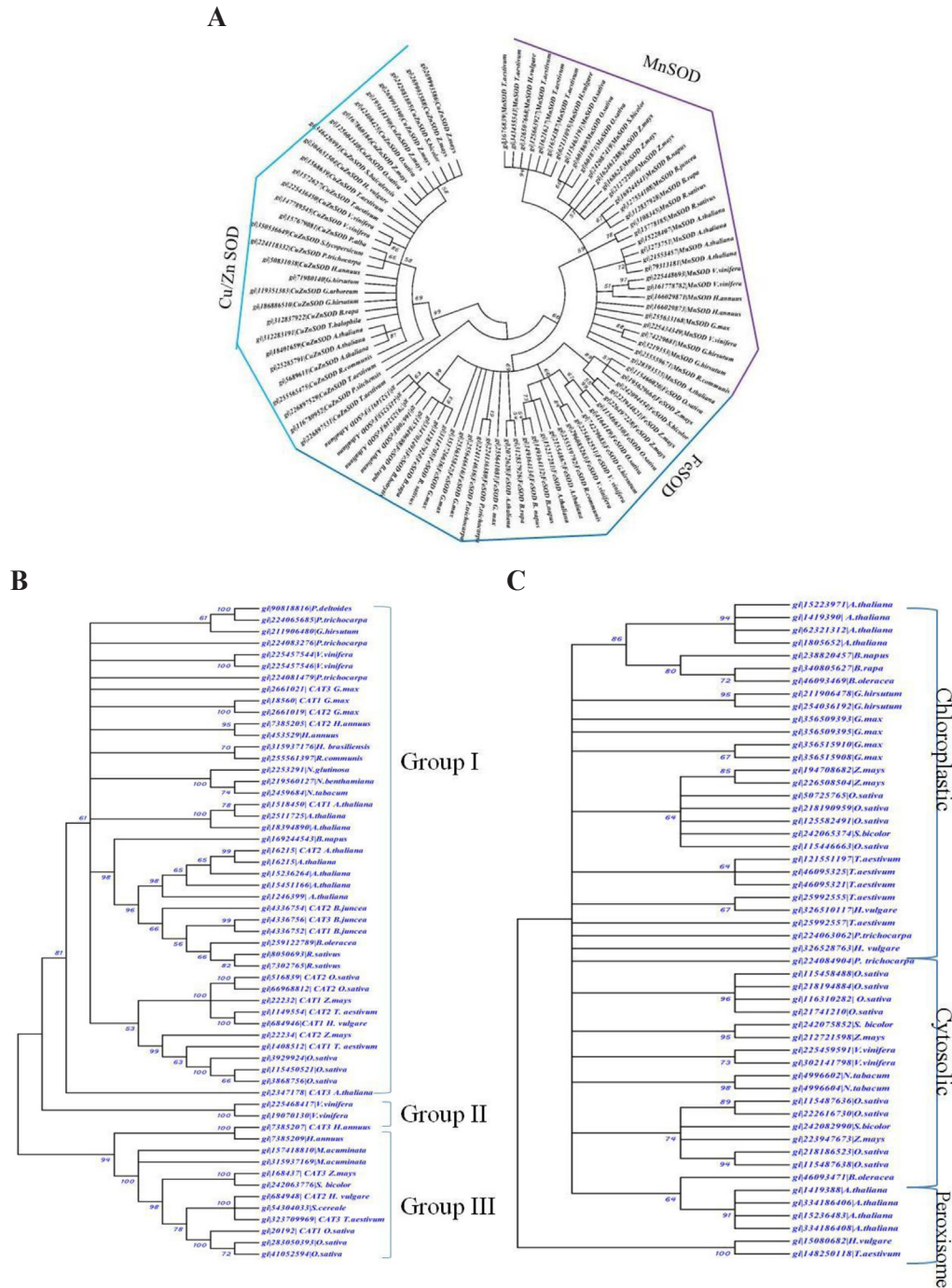


Figure 2. Amino acid sequence alignment of antioxidant proteins. A. Catalase (CAT) and B. ascorbate peroxidase (APX) of selected plants. Alpha helices and beta strands are represented as rods and arrows. Conserved metal binding residues in all plants are shown by shaded region.

Phylogenetic analysis was used to infer evolutionary relatedness among the members of the antioxidant family. Seven SOD-encoding clusters were identified. The phylogenetic analyses revealed that 4 clusters correspond to mitochondrial MnSOD, 1 encoding FeSOD isoforms and 3 clusters encoding CuZnSOD. SOD showed a distinct metal-binding specific cluster of sequences, indicating clear metal-specific sequence differences. It has been suggested that the most ancient FeSOD enzymes were replaced by CuZnSODs in some plant species (Hidalgo et al., 2001). Comparison of deduced amino acid sequences from these different types of SODs suggested that CuZnSODs have no detectable sequence similarity to Mn and FeSODs. Thus, FeSOD and MnSOD may have evolved from a common ancestor while CuZnSODs evolved recently and independently (Hidalgo et al., 2001). We found the CuZnSOD was highly conserved, and that the evolution of this type of SOD in plant species was largely divergent (Figure 3A). Multiple structure alignment and phylogenetic analysis of antioxidants in 4 plant species indicated that SOD sequences possess high sequence and structural similarity irrespective of their dicot or monocot origin, suggesting they shared a common evolutionary relationship. The phylogenetic analysis of catalase revealed 3 major groups. The first group showed a clear demarcation of catalase into 3 clades. Clade I comprised dicot sequences (*Glycine max*, *A. thaliana*, *Helianthus annuus*, *Nicotiana tabacum*, *Populus deltoides*, *N. glutinosa*, and *Vitis vinifera*) and clade II included monocots (*Hordeum vulgare*, *O. sativa*, *Z. mays* and *T. aestivum*), whereas cluster III included sequences from *Raphanus sativus*, *Brassica napus*, *B. juncea*, and *B. oleracea*). The second group included *V. vinifera*, supported by 100% bootstrap values. The monocot-specific catalases (group III) form a single group including *O. sativa* CAT1, *T. aestivum* CAT3, *Z. mays* CAT3, *H. vulgare* CAT2, and *Sorghum bicolor* CAT3 supported by bootstrap values of 100%. These results clearly indicated that most of the dicot catalases and some monocot catalases are very closely related (group I). Monocot-specific catalases (group III) are closely related, but have low sequence identities with other plant catalases (Figure 3B). The wide diversity of plant CAT sequences has led to some discrepancies in the isozyme correlation in phylogenetic trees. The phylogenetic analysis of APX showed 3 clades based on subcellular localization, i.e., cytosolic, peroxisomal, and chloroplast as shown in Figure 3C. The chloroplast branch has apparently diverged only recently into stromal and thylakoid-bound types in *A. thaliana* and *T. aestivum*. The cytosolic and peroxisome were more closely related than the cytoplasmic sequences. The different APX isozymes are very similar, sharing 50 to 70% sequence homology, so a common evolutionary route for different APX genes has been proposed. Furthermore, the phylogenetic and structural organization of APX genes indicated that different APX isoforms arose by a complex evolutionary process involving several gene duplications, suggesting a close relationship among proteins located in the same subcellular compartment (Teixeira et al., 2004).

## CONCLUSION

In addition to some common signatures related to a particular family, some unique families were present in particular antioxidants and isozymes. For instance, AMP\_PHOSPHO\_SITE was present in mitochondrial MnSOD and FeSOD, TYR\_PHOSPHO\_SITE in catalase, and CAMP\_PHOSPHO\_SITE in ascorbate peroxidase, which have significant roles in oxidative stress in the selected plant species. With the help of secondary



**Figure 3.** Phylogenetic tree of **A.** superoxide dismutase (SOD), **B.** catalase (CAT) and **C.** ascorbate peroxidase (APX) was constructed using the neighbor-joining method.

structure prediction and phylogenetic analysis, we found that CuZnSODs share greater sequence and structural similarity than MnSOD and FeSOD. Our results showed the sequence, structural, and functional relationship of antioxidant sequences with biologically significant signature patterns from different species, which further reveals how these genes might have evolved. The phylogenetic tree outlined the development of antioxidants in wheat, *Arabidopsis*, rice, *Brassica*, and other plant species; many exhibited orthologous and paralogous relationships with each other. The comparative analysis showed footprints for a complex evolutionary scenario through species-specific expansion and contraction of gene families. At the molecular level, conserved signature patterns shared by these protein sequences from diverse plant species suggested specific functions in the plant. Sequences that share conserved patterns are crucial to protein structure and function and could provide a signature for the family. These signatures can be very useful for the annotation of uncharacterized proteins, specific retrieval of protein sequences of interest from non-redundant databases, and screening of biomodulators involved in plant stress metabolism. Lastly, the information could be used to study the comparative molecular evolution of the one of the oldest supergene families across diverse lineages.

## ACKNOWLEDGMENTS

Research supported by the Agri Bioinformatics Promotion Program by the Bioinformatics Initiative Division, Department of Information Technology, Ministry of Communications & Information Technology, Government of India, New Delhi. The authors are grateful to the Project Director, DWR, Karnal, for providing the necessary facilities, constant support, and encouragement throughout the study.

## [Supplementary material](#)

## REFERENCES

- Andriotis VM and Rathjen JP (2006). The Pto kinase of tomato, which regulates plant immunity, is repressed by its myristoylated N terminus. *J. Biol. Chem.* 281: 26578-26586.
- Asada K (2006). Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiol.* 141: 391-396.
- Batistic O, Sorek N, Schultke S, Yalovsky S, et al. (2008). Dual fatty acyl modification determines the localization and plasma membrane targeting of CBL/CIPK Ca<sup>2+</sup> signaling complexes in *Arabidopsis*. *Plant Cell* 20: 1346-1362.
- Benetka W, Mehlmer N, Maurer-Stroh S, Sammer M, et al. (2008). Experimental testing of predicted myristoylation targets involved in asymmetric cell division and calcium-dependent signalling. *Cell Cycle* 7: 3709-3719.
- Bowler C, Van Montagu M and Inzé D (1992). Superoxide dismutase and stress tolerance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 43: 83-116.
- Bowler C, Neuhaus G, Yamagata H and Chua NH (1994). Cyclic GMP and calcium mediate phytochrome phototransduction. *Cell* 77: 73-81.
- Chothia C and Lesk AM (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5: 823-826.
- Dat J, Vandenabeele S, Vranová E, Van Montagu M, et al. (2000). Dual action of the active oxygen species during plant stress responses. *Cell Mol. Life Sci.* 57: 779-795.
- del Río LA, Sandalio LM, Yanez J and Gomez M (1985). Induction of a manganese-containing superoxide dismutase in leaves of *Pisum sativum* L. by high nutrient levels of zinc and manganese. *J. Inorg. Biochem.* 24: 25-34.
- Delseny M, Cooke R, Raynal M and Grellet F (1997). The *Arabidopsis thaliana* cDNA sequencing projects. *FEBS Lett.* 403: 221-224.

- Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792-1797.
- Gratao PL, Gomes Junior RA, Delite FS, Lea PJ, et al. (2006). Cadmium Toxicity and Tolerance in Plants. Alpha Sciences International Ltd., Oxford.
- Hanks SK (2003). Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol.* 4: 111.
- Hidalgo P, Garretton V, Berrios CG, Ojeda H, et al. (2001). A nuclear casein kinase 2 activity is involved in early events of transcriptional activation induced by salicylic acid in tobacco. *Plant Physiol.* 125: 396-405.
- Ishikawa T, Yoshimura K, Sakai K, Tamoi M, et al. (1998). Molecular characterization and physiological role of a glyoxysome-bound ascorbate peroxidase from spinach. *Plant Cell Physiol.* 39: 23-34.
- Ishitani M, Liu J, Halfter U, Kim CS, et al. (2000). SOS3 function in plant salt tolerance requires N-myristoylation and calcium binding. *Plant Cell* 12: 1667-1678.
- Kanematsu S and Asada K (1990). Characteristic amino acid sequences of chloroplast and cytosol isozymes of CuZn-Superoxide dismutase in spinach, rice and horsetail. *Plant Cell Physiol.* 31: 99-112.
- Katyshev AI, Konstantinov I and Kobzev VF (2006). Characterization of Mn- and Cu/Zn-containing superoxide dismutase gene transcripts in *Larix gmelinii*. *Mol. Biol.* 40: 372-374.
- Kliebenstein DJ, Monde RA and Last RL (1998). Superoxide dismutase in *Arabidopsis*: an eclectic enzyme family with disparate regulation and protein localization. *Plant Physiol.* 118: 637-650.
- La Du BN and Zannoni VG (1955). The tyrosine oxidation system of liver. II. Oxidation of p-hydroxyphenylpyruvic acid to homogentisic acid. *J. Biol. Chem.* 217: 777-787.
- Miyake C, Cao WH and Asada K (1993). Purification and molecular properties of the thylakoid-bound ascorbate peroxidase in spinach chloroplast. *Plant Cell Physiol.* 34: 881-889.
- Moorhead GB, De Wever V, Templeton G and Kerk D (2009). Evolution of protein phosphatases in plants and animals. *Biochem J.* 417: 401-409.
- Murao K, Takamiya M, Ono K, Takano H, et al. (2004). Copper deficiency induced expression of Fe-superoxide dismutase gene in *Matteuccia struthiopteris*. *Plant Physiol. Biochem.* 42: 143-148.
- Noctor G, Veljovic-Jovanovic S and Foyer CH (2000). Peroxide processing in photosynthesis: antioxidant coupling and redox signalling. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355: 1465-1475.
- Notredame C, Higgins DG and Heringa J (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302: 205-217.
- Pei J, Kim BH and Grishin NV (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36: 2295-2300.
- Rayapureddi JP, Kattamuri C, Chan FH and Hegde RS (2005). Characterization of a plant, tyrosine-specific phosphatase of the aspartyl class. *Biochemistry* 44: 751-758.
- Resh MD (1999). Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochim. Biophys. Acta* 1451: 1-16.
- Riera M, Peracchia G and Pages M (2001). Distinctive features of plant protein kinase CK2. *Mol. Cell Biochem.* 227: 119-127.
- Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- Scandalios JG, Guan L and Polidoros AN (1997). Catalases in Plants: Gene Structure, Properties, Regulation, and Expression. In: *Oxidative Stress and the Molecular Biology of Antioxidant Defenses* (Scandalios JG, ed.). Cold Spring Harbor Laboratory, New York, 343-406.
- Shigeoka S, Ishikawa T, Tamoi M, Miyagawa Y, et al. (2002). Regulation and function of ascorbate peroxidase isozymes. *J. Exp. Bot.* 53: 1305-1319.
- Spiteri A, Viratelle OM, Raymond P, Rancillac M, et al. (1989). Artfactual origins of cyclic AMP in higher plant tissues. *Plant Physiol.* 91: 624-628.
- Stone JM and Walker JC (1995). Plant protein kinase families and signal transduction. *Plant Physiol.* 108: 451-457.
- Takahashi Y, Shomura A, Sasaki T and Yano M (2001). Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proc. Natl. Acad. Sci. U. S. A.* 98: 7922-7927.
- Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
- Teixeira FK, Menezes-Benavente L, Margis R and Margis-Pinheiro M (2004). Analysis of the molecular evolutionary history of the ascorbate peroxidase gene family: inferences from the rice genome. *J. Mol. Evol.* 59: 761-770.
- Van Camp W, Inzé D and Van Montagu M (1997). The regulation and function of tobacco superoxide dismutases. *Free Radic. Biol. Med.* 23: 515-520.
- Willekens H, Chamnongpol S, Davey M, Schraudner M, et al. (1997). Catalase is a sink for H<sub>2</sub>O<sub>2</sub> and is indispensable for

- stress defence in C3 plants. *EMBO J.* 16: 4806-4816.
- Wu G, Wilen RW, Robertson AJ and Gusta LV (1996). The electronic plant genes register PGR 96-126 molecular cloning of two wheat (*Triticum aestivum*) genes (Accession Nos. U69536 (SOD11) and U69632 (SOD12) encoding chloroplast Cu/Zn superoxide dismutases. *Plant Physiol.* 112: 1736.
- Wu G, Robertson AJ, Wilen RW and Gusta LV (1997). Isolation and characterisation of two cDNAs (Accession Nos. U72212 and U73172) encoding mitochondrial manganese superoxide dismutases in wheat. *Plant Physiol.* 113: 664.
- Zamocky M, Jakopitsch C, Furtmuller PG, Dunand C, et al. (2008). The peroxidase-cyclooxygenase superfamily: Reconstructed evolution of critical enzymes of the innate immune system. *Proteins* 72: 589-605.
- Zvelebil MJ and Baum JO (2008). Understanding Bioinformatics. Garland Science, Taylor & Francis Group, New York and London.