Residue Frequencies and Conserved Phylogenetic Signatures in Amino Acid Sequences of Plant Glutathione Peroxidases, Indicates Habitat Specific Adaptation and Dictates Interactions with Key Ligands

Sayak Ganguli[1,*], Abhijit Datta[2]

[1]DBT Centre for Bioinformatics, Presidency University, Kolkata, India
[2]Department of Botany, Jhargram Raj College, West Bengal, India

Abstract   Glutathione peroxidases in plants present one of the most interesting evolutionary enigmas due to the fact that aquatic members encode glutathione peroxidases with selenocysteine while land plants favour the amino acid cysteine. This analysis was performed to identify residue frequencies and conserved motifs in plant glutathione peroxidase sequences with the objective to identify possible evolutionary insights into this interesting group of stress enzymes. Phylogenetic footprinting was combined with parsimony analyses to identify true evolutionary signatures of the group. Following the identification of conserved phylogenetic signatures two essential ligands of the enzyme class-glutathione and hydrogen peroxide was used to study the interaction of the enzymes and the role of the conserved stretches in the interactions. Results indicate that these conserved stretches are important for maintenance of the structural and functional stability of these enzymes as well as interaction with key ligands.

Keywords   Plant Glutathione Peroxidase (GpX), Selenocysteine, Phylogenetic Footprinting, Evolutionary Signatures

1. Introduction

Plants suffer from various kinds of oxidative stress as they are always under the influence of environmental factors [1, 7]. These factors contribute towards aerobic reactions which result in free radicals which serve as deleterious molecules in the cytoplasm. To counter these toxic compounds, plants have devised enzymatic defense mechanisms in the form of free radical scavenging enzymes such as superoxide dismutases, catalases, ascorbate peroxidases and Glutathione Peroxidases (GPx). Glutathione peroxidase (EC 1.11.1.9 and EC 1.11.1.12) family of enzymes are involved in catalyzing of the process of reduction of H2O2 or organic hydroperoxides to water or their alcoholic forms where reduced Glutathione (GSH) acts as an electron donor [10, 19]. At the molecular level, plant GPx genes are closely related to animal PHGPx and their corresponding proteins. Three widely conserved Cysteine residues, which are assumed to be essential for the enzymatic catalysis [17]. The most significant difference between mammalian and land plant glutathione peroxidases is the absence of selenosysteine in the catalytic centre of the later group [2, 3, 5, 6]. Moreover, the plant GPx gene family may comprise up to six members that are distributed in different subcellular compartments. Evolutionary analyses of plant GpX members have shown that their evolution was probably as a resultant of gene duplication and horizontal gene transfer events[16].

2. Materials and Method

2.1. Data Mining and Curation

All glutathione peroxidase sequences which were available were downloaded from the NCBI Genpept resource and were carefully curated. Only those sequences which had more than 100 amino acids were selected for the study since protein motifs tend to occur in clusters and thus a minimum sequence length was required. Following this they were analyzed for their conserved domains to check whether all the sequences corresponded to the plant GpX family. No sequences were rejected at this step. Thus the final set of sequences numbered to 440 and the entire analyses was performed using them. It is important to mention at this point that distant homologues were not considered for the analyses.

2.2. Identification of Conserved Stretches

The PERL (Practical Extraction and Reporting Language) script that was formulated produced a total of 175 positions from the multiple sequence alignment that was provided to it. This MSA was generated using Phylip and then validated using SEAVIEW. Once the positions were generated they were matched with the consensus sequence pattern that was obtained from the UGENE tool.

2.3. Phylogenetic Footprinting

Phylogenetic footprinting [4] has been one of the most trusted methods to identify transcription factor binding sites in the upstream region of various genes since it was first used by Tagle *et al.* (1988) [24]. However, the authors decided to extend the repertoire of phylogenetic footprinting to amino acid sequences by combining maximum parsimony based tree generation methods along with basic multiple sequence alignment [9]. The basic steps of generating the footprint were as follows:

- Create a cluster of homologous sequences by applying homology search tools such as BLAST [18].
- Since the cluster was already prepared a basic multiple sequence alignment [23] was performed using Phylip and was validated using SEAVIEW program. The multiple sequence alignment was constructed using Kimura distances and maximum parsimony was used to generate the cladogram [25]
- A PERL program was designed which accepted the sequences in aligned format and calculated the residue frequencies for each position of the alignment. A host of frequency hits were retrieved: >20-90%

- From the multiple sequence alignment a consensus sequence was generated using the UGENE [22] tool and the conserved regions represented in the alignment were analyzed

2.4. Molecular Dynamic Simulations and Docking

Molecular dynamic simulations were performed with CABS FLEX server and docking was done using FLEX-X using an aquatic and a land plant glutathione peroxidase as the receptor and hydrogen peroxide and glutathione as ligand. Flexible docking was performed after specifying the grid for interaction at the catalytic centre for both enzyme structures. The following diagram summarizes the flowchart of the work (**Fig. 1**).

3. Results and Discussion

3.1. Amino Acid Content

Amino acid content and occurrence has been reported to be conserved from species to species. This conservation may be attributed to protein stability and expressivity of the proteome [11, 13, 14]. In case of individual protein families the occurrence and content govern the formation and maintenance of various motifs and domains. Recent studies have indicated that the mutational landscape of proteins can be correlated to the composition of amino acids. In this study Valine was found to be the most frequently occurring amino acid in the sequences under study. It was also observed that it was occurring in the conserved positions more often. The next most frequent amino acids were found to be lysine and Serine (**Fig. 2 and Table 1**).
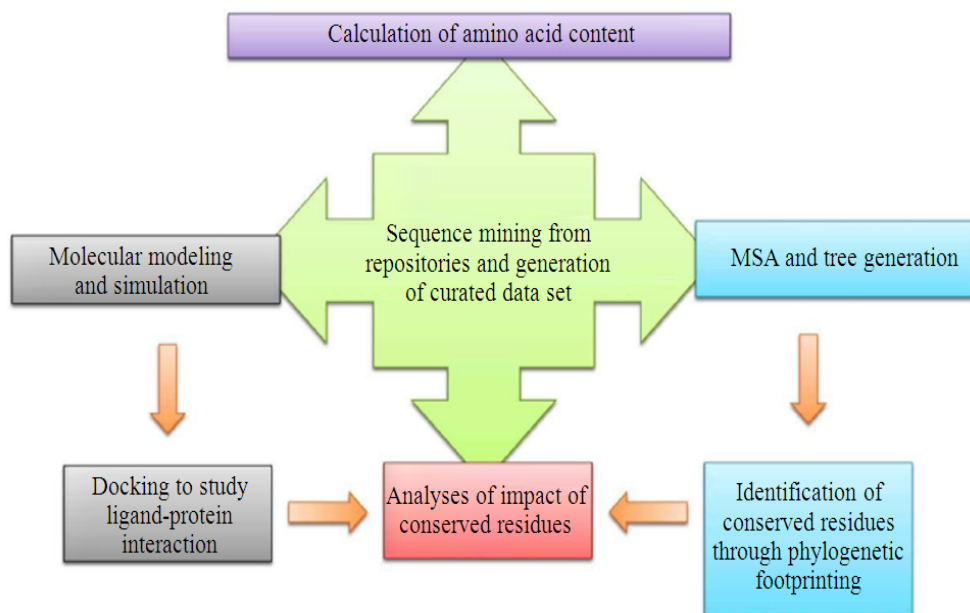

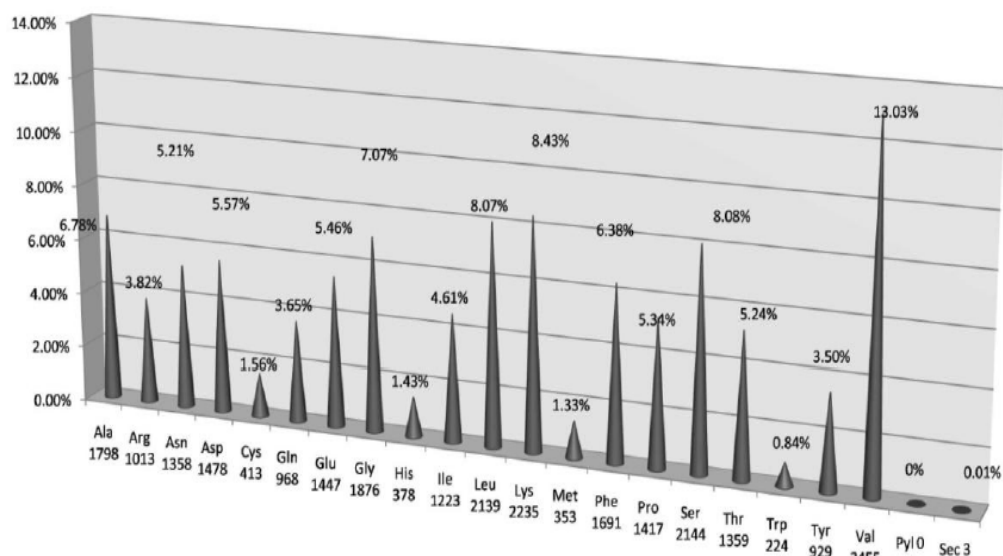**Figure 1.**   Flow chart of Analyses

Figure 2. Amino acid frequencies in conserved positions in glutathione peroxidase sequences

Table 1. Amino acid frequencies in conserved positions

Amino Acid	% in total sequences	Occurrence in conserved position
Ser (S)	8.08	3
Asp (D)	5.57	4
Phe (F)	6.38	5
Thr (T)	5.24	2
Val (V)	13.03	10
Lys (K)	8.43	8
Gly (G)	7.07	9
Leu (L)	8.07	5
Asn (N)	5.20	5
Cys (C)	1.50	2
Tyr (Y)	3.50	4
Glu (E)	5.40	2
Ile (I)	4.60	2
Pro (P)	5.34	5
Gln (Q)	3.65	2
Arg (R)	3.82	2
Trp (W)	0.84	1

3.2. Motif/Pattern Identification and Mapping of Conserved Positions

It was found that most of the sequence positions present in UGENE were present in the positions that were obtained using the in house program (Fig 3 and 5). However, only those residues which had >60% frequency in their respective positions were found to be common in both the result sets. The gaps that were present between the residues (>60% frequency) were replaced with X (symbol for any amino acid) and a minimum length of 6 amino acids were selected as the window size. Now using this window size a sliding window approach was initiated which resulted in the generation of 3420 motif patterns. Now each of these patterns were searched against Prosite and eight conserved consensus patterns were identified which serve as

phylogenetic signatures for plant glutathione peroxidases (**Table 2**). LYXKYK and EILAFPCNQFG patterns were the most dominant patterns with the largest number of hits (Fig. 3).
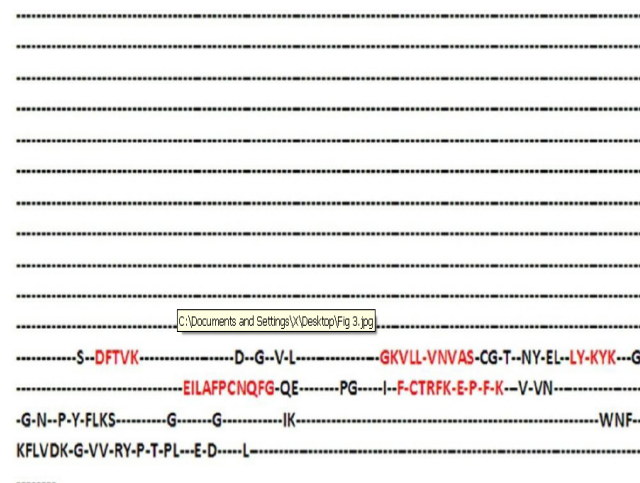

Figure 3. A portion of the consensus sequence generated indicating the sequence positions which correspond to phylogenetic signatures

3.3. Phylogenetic Analyses and Comment on Evolution of Plant Glutathione Peroxidases

The phylogenetic analysis reveals that GPXs in plants exhibit two main groups (Fig. 4)-one group which belongs to the aquatic members while the second group belongs to terrestrial land plants. A further interesting fact is two Cysteine residues were found to occur in the EILAFPCNQFG and the FXCTRFK motifs which have been reported to be part of the interacting residues of this enzyme group. We have earlier reported that these cysteine residues are important moieties which hydrogen bond with the respective ligands of glutathione peroxidases [12]. This divide is justified as several reports and analysis of amino acids sequences reveal that GPXs of aquatic members

contain selenocysteine while those of land plants do not [8, 15]. Selenoproteins are absent in higher plants but green algae *Chlamydomonas* sp and *Ostreococcus* sp. possesses 12 and 26-29 selenoproteins respectively [20]. The lack of selenoproteins in land plants is a well documented phenomenon [15, 21]. The most accepted explanation states the conversion of anaerobic earth to aerobic earth through oxygenation and the selection of a more stable and less easily oxidizable forms of glutathione peroxidases. This analyses further strengthens the view with the observations that the identified phylogenetic signatures are important protein landmarks and their presence ubiquitously in case of both land and aquatic members indicate that possibly the evolution of glutathione peroxidases could have occurred along a modular assembly line.
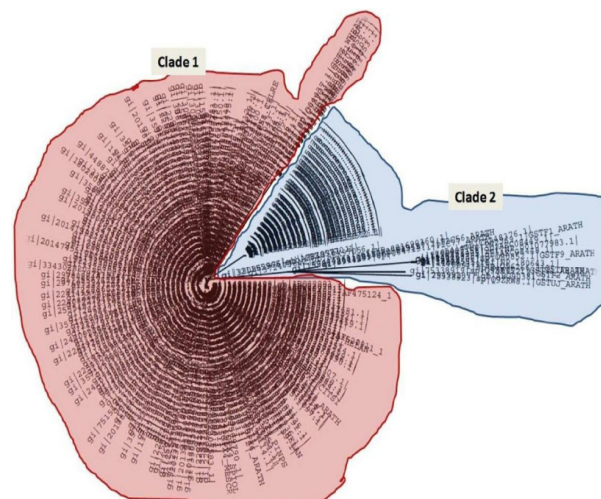

Figure 4. Phylogenetic tree showing two distinct clades (Clade 1-Land plant GpXs and Clade 2-Aquatic GpX members)
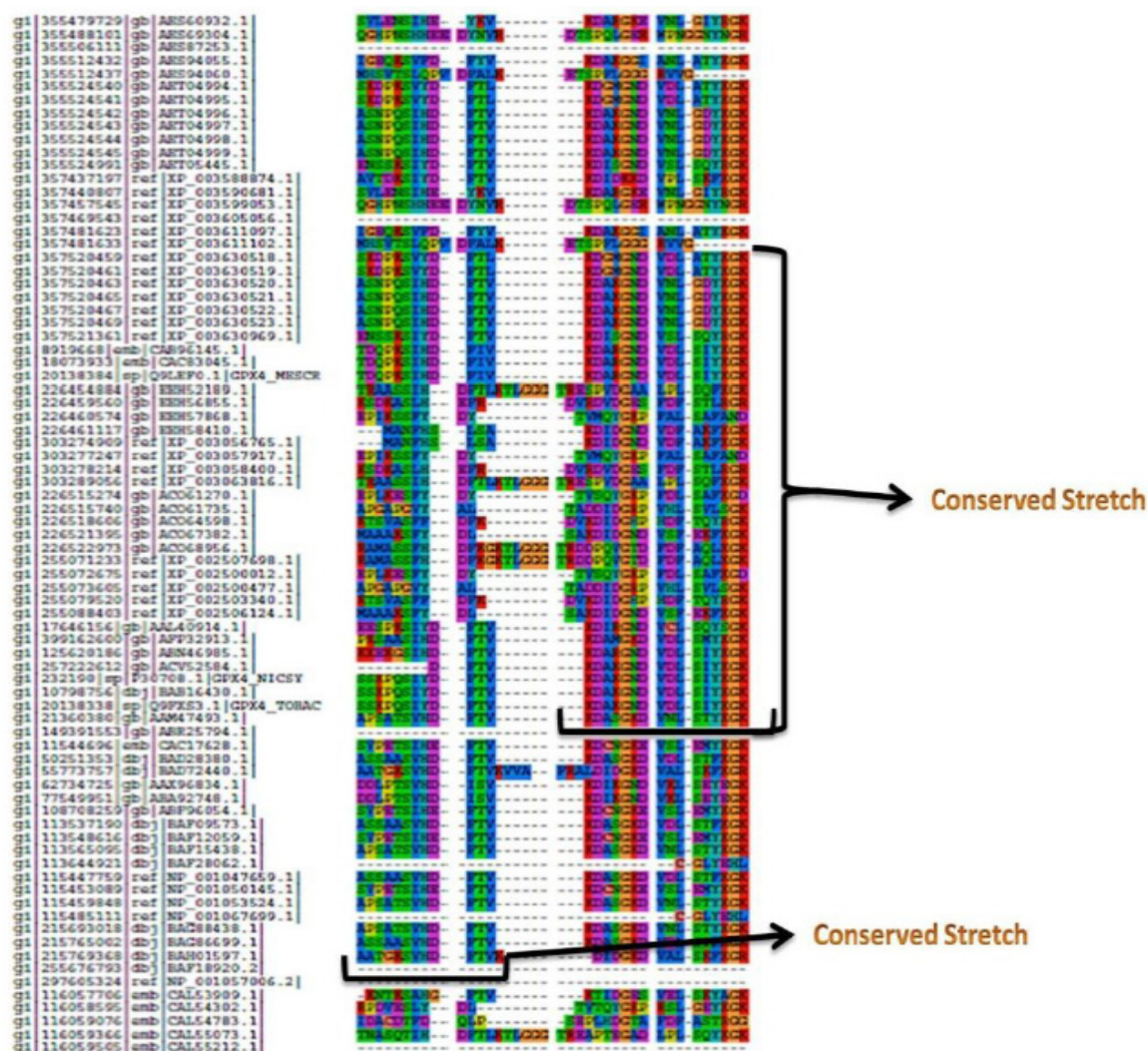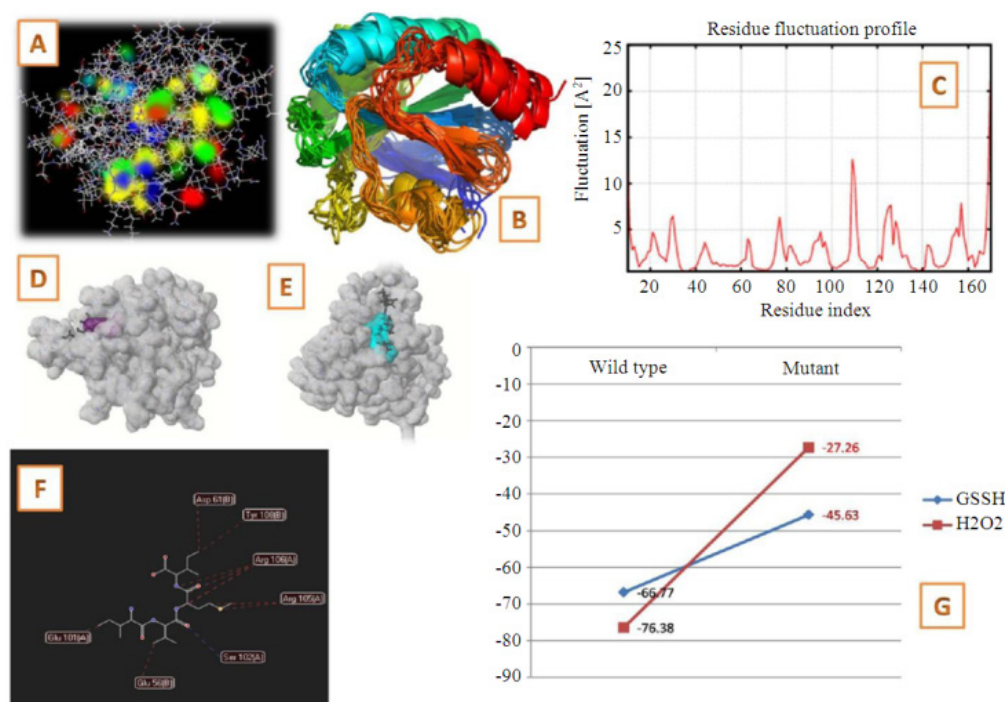

Figure 5. A portion of the alignment of glutathione peroxidase sequences with the highest levels of sequence conservation (possible phylogenetic signatures)

Table 2. List of identified motifs from PROSITE matching with the patterns identified from the motif identification program. ('X' indicates any amino acid here)

Motif	Hits in prosite	Prosite descriptions
1. Sxxdftvk	Found: 10 hits in 10 sequences	1. Probable glutathione peroxidase 2. Arabidopsis thaliana 2. Putative glutathione peroxidase 7, chloroplastic. Arabidopsis thaliana
2. Gkvllxvnvasxcgxt	Found: 8 hits in 8 sequences	1. Phospholipid hydroperoxide glutathione peroxidase, chloroplastic. Pisum sativum 2. Probable glutathione peroxidase 3, mitochondrial. Arabidopsis thaliana
3. Lyxkyk	Found: 233 hits in 233 sequences	1. Probable phospholipid hydroperoxide glutathione peroxidase. Solanum lycopersicum 2. Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial. Arabidopsis thaliana
4. Eilafpcnqfg	Found: 223 hits in 223 sequences	1. Glutathione peroxidase 1. Helianthus annuus 2. Probable phospholipid hydroperoxide glutathione peroxidase. Citrus sinensis
5. Fxctrfkxexpxfxk}	Found: 9 hits in 9 sequences	1. Probable phospholipid hydroperoxide glutathione peroxidase. Citrus sinensis 2. Probable phospholipid hydroperoxide glutathione peroxidase. Helianthus annuus
6. Pxyxflks	Found: 15 hits in 15 sequences	1. Phospholipid hydroperoxide glutathione peroxidase 1, chloroplastic. Arabidopsis thaliana 2. Glutathione peroxidase. Schizosaccharomyces pombe (strain 972 / ATCC 24843)
7. Kflvdk	Found: 21 hits in 21 sequences	1. Phospholipid hydroperoxide glutathione peroxidase, chloroplastic. Pisum sativum 2. Phospholipid hydroperoxide glutathione peroxidase, mitochondrial. Callithrix jacchus
8. Xryxpxtxpl thaliana	Found: 8 hits in 8 sequences	1. Probable glutathione peroxidase 2 Arabidopsis 2. Probable phospholipid hydroperoxide glutathione peroxidase. Citrus sinensis

Figure 6. Results of Structural and Interaction studies. A: Molecular model of glutathione peroxidase with its electron dense residues. B: Simulated structural alignment; C: Residue fluctuation profiles of the protein; D and E: Interactions of glutathione and hydrogen peroxide with the land plant and aquatic members respectively; F: Interacting residues following mutation; G: Changes in free energy of interaction of glutathione peroxidase with glutathione and hydrogen peroxide

Here the same enzyme structure was used for the analyses to represent the land plant lineage that was obtained from our phylogenetic study while the glutathione peroxidase structure of *Chlamydomonas reinhardtii* was used to represent the member of aquatic land plant lineage (Fig 6. A).

The binding pockets of both the lineages were identified and in case of aquatic lineage the key binding residue was identified to be selenocysteine while in case of land plants the key binding residue was cysteine as reported earlier (Fig 6, D, E and F). Interestingly in both the cases the phylogenetic signature stretches were found to be within a distance of 10 angstrom radii from the binding pockets. When these stretches were mutated with alanine then the interactions were severely affected as observed in the changes in binding energies (Fig 6. C and G)

4. Conclusions

Our analyses strengthens this hypotheses as the conserved residues identified in land plant lineages possess cysteine at the catalytic residue cluster, while in case of aquatic members it is replaced by selenocysteine. Both these residues are part of the conserved residues that were identified in the respective lineages thus showing that transition of selenocysteine to cysteine as a necessary adaptation for land plant transition. The involvement of the conserved residues in interactions also establishes their importance as functional phylogenetic signatures.

REFERENCES

[1]   Alscher, R.G., J.L. Donahue and C.L. Cramer, 1997. Reactive Oxygen species and antioxidants: Relationships in green cells. Physiol. Plant., 100: 224-223. DOI: 10.1111/j.1399-3054. 1997.tb04778.x

[2]   Berry, M.J., J.W. Harney, T. Ohama and D.L. Hatfield, 1994. Selenocysteine insertion or termination: Factors affecting UGA codon fate and complementary anticodon: Codon mutations. Nucl. Acids Res., 22: 3753-3759. DOI: 10.1093/nar/22.18.3753

[3]   Berry, M.J., R.M. Tujebajeva, P.R. Copeland, X.M. Xu and J.W. Harney *et al*., 2001. Selenocysteine incorporation directed from the 3′UTR: Characterization of eukaryotic EFsec and mechanistic implications. Biofactors, 14: 17-24. DOI: 10.1002/biof.5520140104.

[4]   Blanchette, M. and M. Tompa, 2002. Discovery of Regulatory Elements by a Computational Method for Phylogenetic Footprinting. Genome Res., 12: 739-748. DOI: 10.1101/gr.6902.

[5]   Böck, A., K. Forchhammer, J. Heider and C. Baron, 1991. Selenoprotein synthesis: An expansion of the genetic code. Trends Biochem. Sci., 16: 463-467. DOI: 10.1016/ 0968-0004(91)90180-4.

[6]   Castellano, S., S. Novoselov, G.V. Kryukov, A. Lescure and R. Guigo *et al*., 2004. Reconsidering the evolution of eukaryotic selenoproteins: A novel nonmammalian family with scattered phylogenetic distribution. EMBO Rep., 5: 71-77. DOI: 10.1038/sj.embor.7400036.

[7]   Chen, S., Z. Vaghchhipawala, W. Li, H. Asard and M.B. Dickman, 2004. Tomato phospholipid hydroperoxide glutathione peroxidase inhibits cell death induced by Bax and oxidative stresses in yeast and plants. Plant Physiol., 135: 1630-1641. DOI: 10.1104/pp.103.038091.

[8]   Copeland, P.R., 2005. Making sense of nonsense: the evolution of selenocysteine usage in proteins. Genome Biol., 6: 221-221. DOI: 10.1186/gb-2005-6-6-221.

[9]   Dayhoff, M.O. and R.M. Schwartz. 1978. Evolution of Prokaryotes and Eukaryotes Inferred from Sequences. In: Evolution of Protein Molecules, Matsubara, H. and T. Yamanaka (Eds.)., pp: 323-342.

[10] Ganguli, S. and A. Datta, 2014. Prediction of Indels and SNP's in coding regions of glutathione peroxidases-an important enzyme in redox homeostasis of plants. Int. Lett. Nat. Sci., 2: 49-62.

[11] Graur, D., 1985. Amino acid composition and the evolutionary rates of protein-coding genes. J. Mol. Evol., 22: 53-62. DOI: 10.1007/BF02105805.

[12] Gupta, S., S. Saha P. Roy P. Basu and S. Ganguli, 2010. Structural and functional analysis of glutathione peroxidase from ricinuscommunis L. â€" a computational approach. Int. J. Bioinformat. Res., 2: 20-30. DOI: 10.9735/0975-3087. 2.1.20-30.

[13] Hormoz, S., 2013. Amino acid composition of proteins reduces deleterious impact of mutations. Sci. Rep., 3: 2919-2919. DOI: 10.1038/srep02919.

[14] Kuma, K. and T. Miyata, 1994. Mammalian phytogeny inferred from multiple protein data. Jpn. J. Genet., 69: 555-566. DOI: 10.1266/jjg.69.555.

[15] Lobanov, V.A., E.D. Fomenko, Y. Zhang, A. Sengupta and V.N. Gladyshev *et al*., 2007. Evolutionary dynamics of eukaryotic selenoproteomes: Large selenoproteomes may associate with aquatic life and small with terrestrial life. Genome Biol., 8: 198-198. DOI: 10.1186/gb-2007-8-9-r198.

[16] Margis, R., C. Dunand, T.K. Felipe and M.M. Pinheiro, 2008. Glutathione peroxidase family-an evolutionary overview. FEBS J., 275: 3959-3970. DOI: 10.1111/j.1742-4658. 2008.06542.x.

[17] Navrot, N., V. Collin, J. Gualberto, E. Gelhaye and N. Rouhier *et al*., 2006. Plant glutathione peroxidases are functional peroxiredoxins distributed in several subcellular compartments and regulated during biotic and abiotic stresses. Plant Physiol., 142: 1364-1379. DOI: 10.1104/pp.106.089458

[18] Needleman, S.B. and C.D. Wunsch, 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Molecular Biol., 48: 443-453. DOI: 10.1016/0022-2836(70)90057-4.

[19] Nicolas, N., C. Vale´rie, G. Jose´, G. Eric and and R. Nicolas, 2006. Plant glutathione peroxidases are functional peroxiredoxins distributed in several subcellular compartments and regulated during biotic and abiotic stresses. Plant Physiol., 142: 1364-1379.

[20] Novoselov, S.V., M. Rao, N.V. Onoshko, H. Zhi and V.N. Gladyshev et al., 2002. Selenoproteins and selenocysteine insertion system in the model plant cell system, Chlamydomonas reinhardtii. EMBO J., 21: 3681-3693. DOI: 10.1093/emboj/cdf372.

[21] Obata, T. and Y. Shiraiwa, 2005. A novel eukaryotic selenoprotein in the haptophyte alga Emiliania huxleyi. J. Biol. Chem., 280: 18462-18468. DOI: 10.1074/jbc. M501517200.

[22] Okonechnikov, K., O. Golosova and M. Fursov, 2012. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics, 28: 1166-1167. DOI: 10.1093/bioinformatics/bts091.

[23] Pearson, W.R. and D.J. Lipman, 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA., 85: 2444-2448. DOI: 10.1073/pnas.85.8.2444.

[24] Tagle, D.A., B.F. Koop, M. Goodman, J.L. Slightom and R.T. Jones et al., 1988. Embryonic ε and γ globin genes of a prosimian primate (Galago crassicaudatus) ☆: Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J. Mol. Biol., 203: 439-455. DOI: 10.1016/0022-2836(88)90011-3.

[25] Zhang, J. and M. Nei, 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood and distance methods. J. Mol. Evol., 44: 139-146. DOI: 10.1007/PL00000067.