International Journal of Bioinformatics Research
ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 6, Issue 1, 2015, pp.-321-328.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BIJ0000001

DISTRIBUTION AND VARIABILITY OF CEREAL POLYPHENOL OXIDASE(S) GENE FAMILIES

JUKANTI A.K.* AND KALWANI P.

Central Arid Zone Research Institute, Jodhpur, Rajasthan, India.
*Corresponding Author: Email- jukanti5@yahoo.com

Received: May 21, 2015; Revised: June 11, 2015; Accepted: June 12, 2015

Abstract- Polyphenol oxidases (PPOs) are copper-binding enzymes that oxidize polyphenols to quinones. PPOs are ubiquitous in plants, but, knowledge on their evolution and diversity in cereals is explored to a limited extent. This study reports their distribution and diversity in maize, sorghum, barley and millet. We have identified additional PPO proteins in the four crops than those previously reported. In all, 27 PPO were studied and an overall sequence identity across the species ranged between 30-99%. In addition to several variants of 'HxxYC' motif, a novel motif (HRxYxxFxER) that begins with third conserved histidine residue of 'Copper A' is reported. Another tri-peptide motif, 'AGS' was found to be 100% conserved. Twin-tyrosine (YxY) motif was substituted by 'FTY' or 'YRF' motif in four millet and one barley sequences. Among the 27 PPOs, 7 proteins were predicted to be synthesized via secretory pathway. PROSITE Scan analysis identified several domains including Zinc finger C2H2-type, Immunoglobulin [Ig]-like, and 'TAT' signal. Phylogenetic analysis revealed two major clades and also indicated a non-species specific diversification of the cereal PPOs investigated. Overall, our study analysed the diversity among the four cereal PPOs, it is observed that their number and distribution is consistent with their implications in different roles.

Keywords- polyphenol oxidase, diversity, cereals, millets, phylogenetic analysis

Citation: Jukanti A.K. and Kalwani P. (2015) Distribution and Variability of Cereal Polyphenol Oxidase(s) Gene Families. International Journal of Bioinformatics Research, ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 6, Issue 1, pp.-321-328.

Introduction

Polyphenol oxidases (PPOs) are copper-containing enzymes with diverse phylogenetic distribution in plants, animals, bacteria and fungi [1]. PPOs are nuclear-encoded proteins that are synthesized as a precursor cytosolic protein, processed into mature proteins and transferred to plastids [2,3]. PPOs are of three types: (i) cresolases (monophenol oxidase) (ii) *ortho*-diphenol: oxygen oxidoreductase (catecholase) - mostly occur as multi-gene families and (iii) laccase-like multi copper oxidases [4]. PPOs hydroxylate monophenols to *ortho*-diphenols followed by dehydrogenation to *ortho*-quinones. The quinones either undergo self-polymerization or react with nucleophiles to produce dark colored pigments that are mostly undesirable [5].

PPOs are expressed in diverse tissues and conditions suggesting their involvement in varied processes [6]. PPOs have been implicated in several processes including: (i) time-dependent darkening of cereal-based food products [7] (ii) defense against pests and pathogens [8,9] (iii) in latex coagulation [10], phenyl propanoid [11] and betalain [12] biosynthesis. PPOs limit the pathogen translocation by forming a physical barrier formed due to conversion of phenolic compounds by PPOs. Further, the *ortho*-quinones produced can bind proteins, reducing their digestibility and nutritive value to herbivores [13], thereby decreasing the pest incidence. Undesirable darkening or browning of food products reduces their nutritional quality/appearance, consumer acceptance and causes significant economic impact.

PPOs usually contain an N-terminal chloroplast transit peptide (cTP), a dicopper center, and a C-terminal region [14]. Generally, the cTP is of ~80-100 amino acid length, it regulates the import of mature PPO into the thylakoid lumen via the 'twin-arginine' dependent (Tat) pathway [15]. Interestingly, a signal peptide involved in secretory pathway has also been identified and vacuolar localization of snapdragon/poplar PPOs was demonstrated [16,17]. The two copper binding domains (Cu A and Cu B) of PPOs are each bound by three conserved histidine residues and form the active site [18]. Among the two domains, CuA is more variable than CuB and this probably affects the substrate specificity [11]. The proteolytic cleavage of C-terminal fragment (seen in *Vitis vinifera and Vigna faba)* may facilitate activation of latent PPO [19]. The molecular weight of latent PPO is ~65-70 kDa, which is reduced to ≤60 kDa upon its import in to plastids [14,20].

Molecular characterization of PPOs in plants suggested the presence of multi-gene and mostly intronless PPO families (tomato/potato - seven and five single-exon PPOs respectively) [21,22]. But, presence of introns in PPOs has been reported in monocot species like wheat and pineapple [23]. Though PPO sequences have been reported in several crops, little is known about PPOs in cereal (maize/barley/sorghum/foxtail millet) crops. Further, diversity in PPO's occurrence and structure in these crops has never been fully captured or reported. Therefore, we have taken advantage of the

recent progress made in PPO research and genome sequencing efforts to study the multi-gene families of different cereal and millet PPOs. We compared and characterized 27 PPOs in four economically important cereal crops grown in different agro-ecological zones of the world. In the text of this manuscript foxtail millet will be referred to as 'millet'. The major objective of the present effort was to gain insights into the conserved domains/motifs present in PPOs that will aid in better understanding PPOs and the varied roles they are mostly implicated in.

Materials and Methods

Initially two wheat PPO proteins (AAT06523 & AEY79824) were utilized to search for PPO homologs from other cereals of interest including maize (*Zea mays*), sorghum (*Sorghum bicolor*), barley (*Hordeum vulgare*) and finger millet (*Setaria italica*). 'Basic Local Alignment Search Tool P (BLASTP)' of 'National Center for Biotechnology Information (NCBI - http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used for identifying PPO sequences from the 4 crops. BLASTP software tool was used to identify maximum number of PPO sequences. The obtained sequences were exported to 'San Diego SuperComputer Center (SDSC) Biology WorkBench (http://seqtool.sdsc.edu)' for preliminary sequence analysis. CLUTAL W tool was used to perform 'Multiple Sequence Alignment' of the identified PPOs [24]. After the preliminary sequence alignment, sequences <475 amino acids in length and those that were duplicated or identical to other sequences were not considered for further analysis.

'Multiple Sequence Alignment' was done using 'Clustal W' function of MEGA 6.0 [25] using progressive alignment method. In this method most similar sequences with the highest alignment score are aligned first. Later, progressively farther group of sequences are aligned until a global alignment is obtained. Default parameter settings with a delay divergent cutoff of 30% was utilized for sequence alignment. PROSITE Scan was used to compare all 27 protein sequences with default parameter settings for searching patterns and profiles (www.ebi.ac.uk/Tools/pfa/ps_scan) [26,27]. Target and signal peptide prediction was done utilizing TargetP 1.1. TargetP 1.1 is a tool based on neural networks for prediction of sub cellular localization [28]. TargetP 1.1 predicts chloroplastic 'Transit Peptide' (cTP), mitochondrial 'Transit Peptide' (mTP), and secretory 'Signal Peptide' (SP).

The phylogeny was inferred by using the maximum likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model [29]. The initial tree for maximum likelihood was formed automatically using default neighbor joining matrixes. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the phylogeny of the sequences analyzed [30]. The 'Nearest-Neighbor-Interchange', a heuristic method was used to improve the likelihood of a tree. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. All positions containing gaps and missing data were eliminated [31]. A polyphenol oxidase from nitrifying bacteria *Nitrosomonas europaea* (GenBank accession CAD85152) was used as an out-group to construct a rooted tree.

Results

Significant developments have taken place in the recent past in the field of sequencing (genome/expressed sequence tags 'EST'/gene) in several important crops. Therefore, we have taken advantage of these and other developments in performing a detailed analysis of PPOs in four important cereal crops. This search resulted in identification of additional multiple genes in these four crops. The published sequences of wheat PPOs were used to identify PPOs from different crops and these identified (crop specific) were further utilized to unearth the additional PPOs among the respective crops. In total, 27 PPO protein sequences (sorghum-9; maize-8; millet-7; barley-3] were identified and overall sequence identity among these ranged between 30-99% over a minimum amino acid length of 480. Wide variation in within crop PPOs (barley-39-74%; sorghum/maize/millet - 30-99%) was also observed (data not reported).

Multiple sequence alignment of the PPOs using MEGA6.0 revealed important conserved domain in all the sequences. Both the Cu-binding domains (A & B) were identified by presence of at least three conserved histidine (H) residues. The 'HxxYC' motif contains the first conserved 'H' residue of Cu-A domain [Fig-1]. The region between 'HxxYC' motif and third conserved 'H' is highly variable [Fig-1]. A 'HCAYC' motif is the most common one and others include 'HEAYC', 'HQSYC', HQAYC, or 'HESYC '. 'HQAYC' motif is observed in three PPO sequences (maize 10; sorghum 4/2) which are grouped together [Fig-1],[Fig-2]. Similarly, two maize PPOs (maize 4/7) contain 'HQSYC' motif and these two are clustered together [Fig-1],[Fig-2]. Interestingly, sorghum PPO 3 and millet 2 contain 'HESYC' and 'HEAYC' motifs respectively; they are clustered separately [Fig-2]. In addition to this we report a novel motif that begins with the third conserved 'H' (HRxYxxFxER). Millet PPO sequence 2 contains a distinct 'HRMYLYFYER' motif which is not present in any of the analyzed sequences [Fig-1]. Two other major modifications of the novel motif include 'HRMYIYFYER' and 'HRAYLYFFER' [Fig-1]. The first two conserved histidine residues of the 'Cu B' domain form a 'HxxxH' motif. Between the second and third conserved histidine residues either tryptophan, phenylalanine or proline are completely conserved along with aspartic acid [Fig-3]. In the 'Cu-B' domain, 'HGPVH' is the most common motif and it present in 11 PPOs sequences [Fig-3]. It is interesting to note that 5 of the 7 millet PPOs and 50% of the sorghum PPOs have this particular motif in common [Fig-3].

PPO_DWL (Pfam12142) and PPO_KFDV (Pfam12143) domains of 50 and 140-150 amino acid length respectively constitute the C-terminal end of PPO [Fig-4],[Fig-5]. Further, a twin-tyrosine (YxY) motif is observed in all the sequences except five ('Millet 4/5/6/7' and 'Barley 4') [Fig-4]. In the four millet sequences, first tyrosine residue in the 'YxY' motif is substituted by another aromatic amino acid, phenyl alanine 'F' (FTY motif) [Fig-4]. Interestingly, in the 'Barley 4' sequence the last tyrosine residue in the 'YxY' motif is replaced by 'F' (YRF). A glutamic acid rich motif was observed upstream of PPO_KFDV domain in several PPOs (data not shown). Interestingly, an 'AGS' motif that is 100% conserved in all the PPOs analyzed was identified. In several sequences a histidine residue is conserved at the C-terminus of 'AGS' [Fig-5]. Four and three conserved histidine residues are detected immediately at C-terminus of 'AGS' motif in Maize 5/Sorghum 8' and 'Millet 1' respectively [Fig-5].

As stated above PPOs contain transit/signal peptides at the N-terminus of the sequence. Among the 27 cereal PPOs analyzed, 20 of them contained a chloroplast transit peptide (cTP) at the N-terminal [Table-1] and the other seven sequences were predicted by TargetP1.1 to be synthesized via secretory pathway [Table-2]. The cTP (except for 'Maize 3' and 'Barley 3') and signal peptide length mostly ranged between 26-52 and 19-41 amino acids respectively [Table-1],[Table-2].

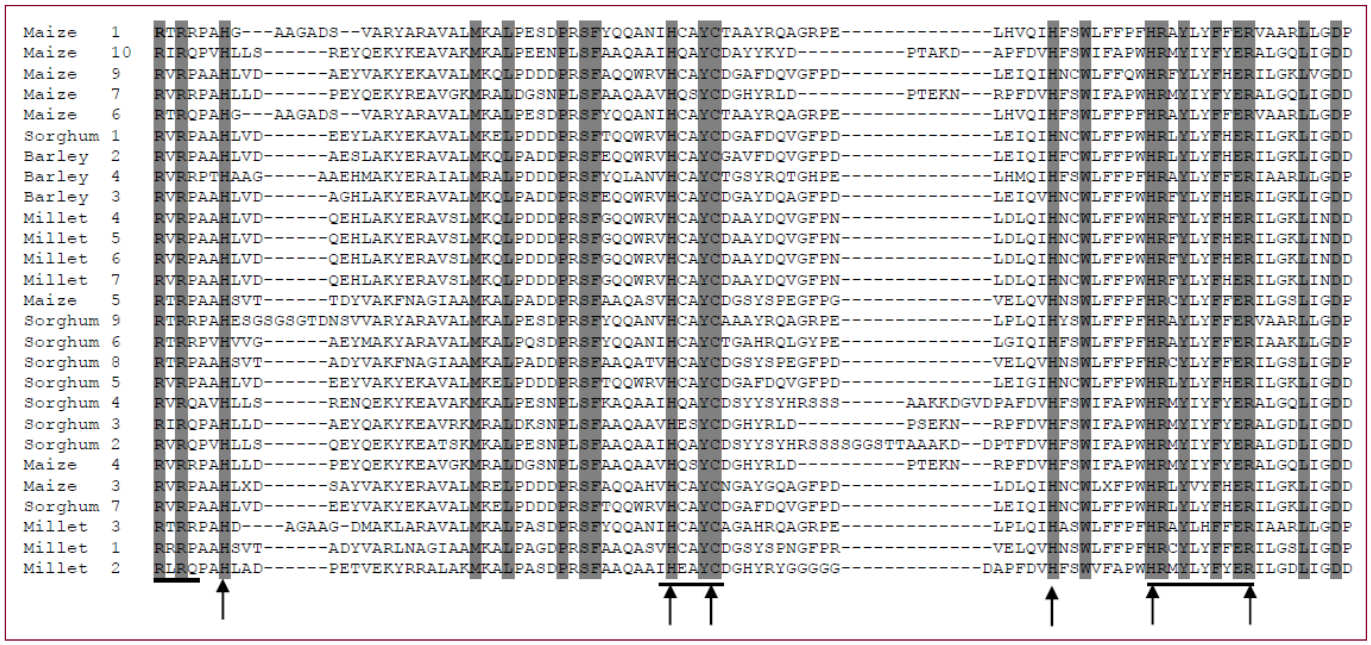

Fig. 1- Multiple sequence alignment of conserved CuA domain of different PPOs as identified using 'Molecular Evolutionary Genetics Analysis (MEGA 6.0)'.

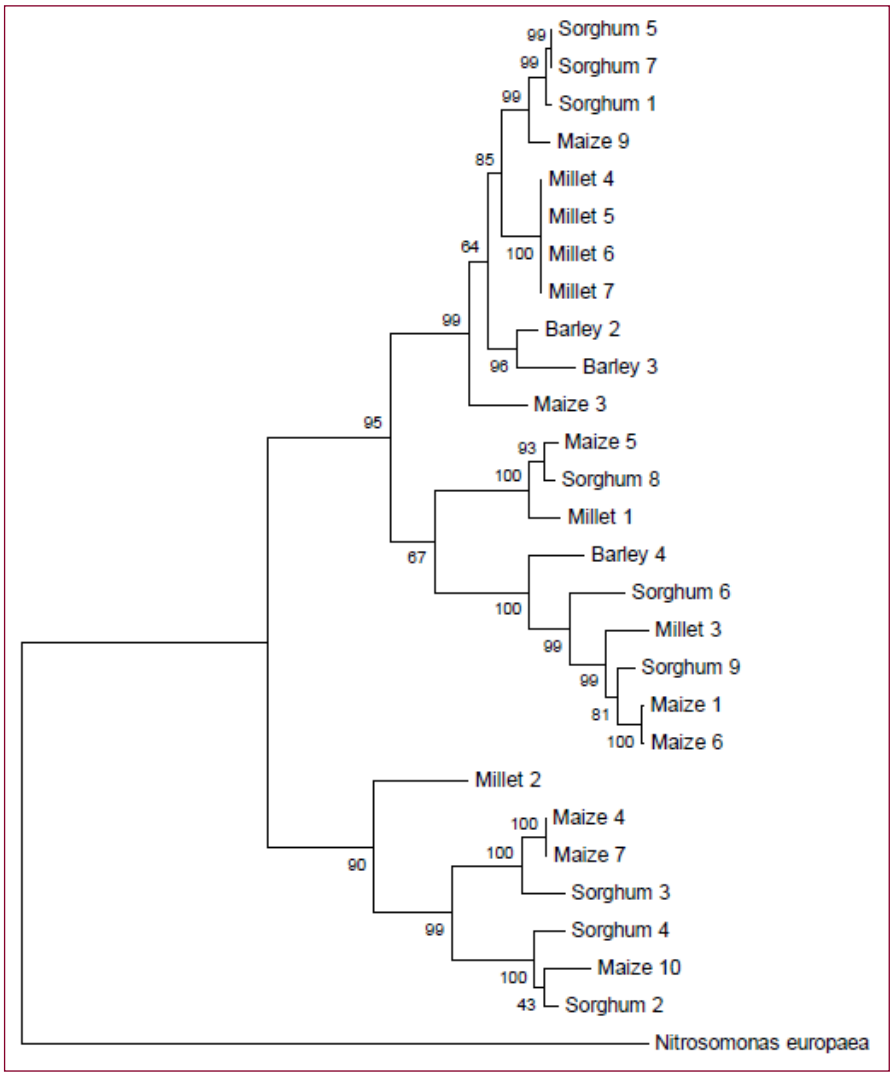

Fig. 2- Maximum likelihood tree showing phylogenetic relationships among different cereal PPOs.

PROSITE Scan' analysis of the 27 sequences identified several additional (to those described above) domains/profiles [Table-3]. Zinc finger C2H2-type domain signature (PS00028) was observed in three sequences: 'Maize 1, 6' and 'Barley 4'. Similarly, 'Maize 4, 7' contained an 'Immunoglobulin (Ig)-like domain' (PS50835). Further, a 'TAT' signal (PS51318) was detected in several sequences including 'Millet 1, 5, 7' and 'Sorghum 1, 5, 7'. Insect "Larval Storage Protein (LSP; PS00210)' signature was found in 'Maize 4, 7', 'Millet 2', and 'Sorghum 2, 3'.

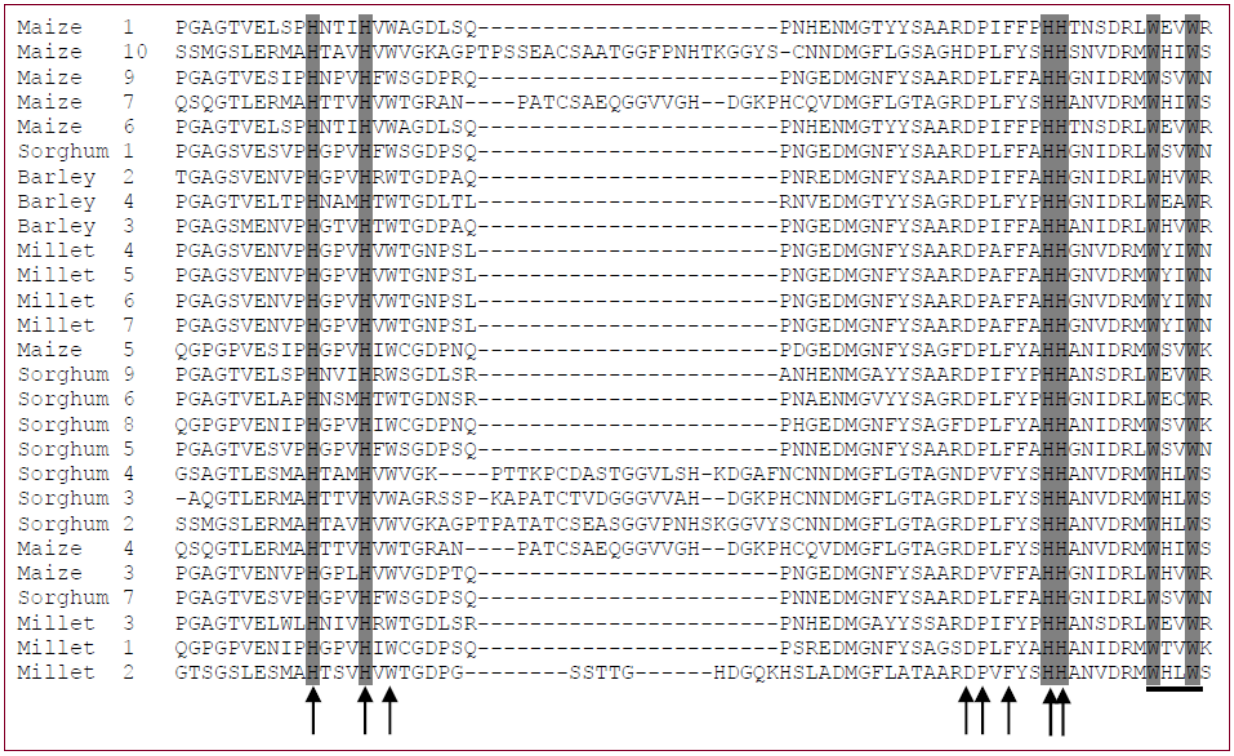

Fig. 3- Multiple sequence alignment of conserved CuB domain of different PPOs as identified using 'Molecular Evolutionary Genetics Analysis (MEGA 6.0)'.

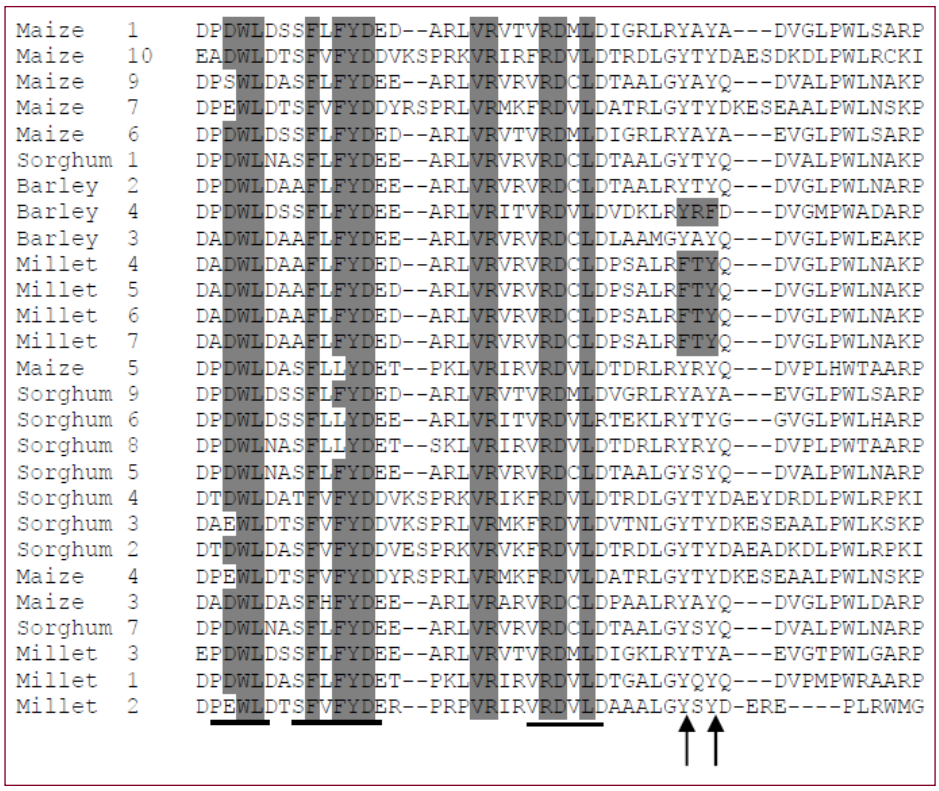

Fig. 4- Conserved DWL domain of different PPOs as identified using 'Molecular Evolutionary Genetics Analysis (MEGA 6.0)'.

```
Maize    1   EVLVVEGIKAD--GADFVKFDVYVNAV--DHEKIMP---GGREMAGSFVSLKHP--------GNMV-VQSSMRVALNELLEDLGAEG-DDSVTVTLVEVEG--RVRIGGLRIVYMAE
Maize    10  QLLVIEGIEYDPQAN--NKFDVAINLPADKALQVGP---QYKEYAGSFAVVP-------GS-GAGKTRKVKLSLCITEVLFDIDADG-DKTVDVVIVERTNAKITLNARPTIKNRN-
Maize    9   EVLVVQGIEIADHSSRFVKFDVFVND---SQSGGGM-PAA-AQCAGSVALTPHAVRPGKGK----GTVKTVARFGICDLLDDIGADG-DKTIIVSLVERCAGDMVTVGGVGIEYVG-
Maize    7   LVLLIEGIEYDPQIN--NKFDVVINVAREDAARVGP---KDSEYAGSFSAVP-------SSNAAGGTLVGKFTLALDGVLADLGLAG-ASAVDIVLVEHTEGEIKLYLPPTIENA--
Maize    6   EVLVVEGIKAD--GADFVKFDVYVNAV--DHEKIMP---GGREMAGSFVSLKHP--------GNMV-VQSSMRVALNELLEDLGAEG-DDSVTVTLVEVEG--RVRIGGLRIVYMAE
Sorghum  1   EVLVVEGIEVADHFSRFVKFDVFVNE---CQGGGGM-GAAVAQCAGSVAMTPHLVRPGKGR----GSVKTAARFGICDLLDSIGADG-DKTIVVSLVERCAGDMVTVGGVRIEYVK-
Barley   2   EVLVVEGIQIADHFNEFVKFDVLVNEP--ERAVGGTPATATGYCAGSFAYTPHMARSDEMRN---GPVKTVARFGVCDLMDDIGADG-DKTVVVSLVERCGCEMVTVGGVSISYLK-
Barley   4   EVLVVEGIETD--GTEMVKFDAYVNAM--EYEKVGP---SGRELAGSYLCLSHPST--DG-TGKGMTVETSMRMALNELLEDLGADG-DESVTVTLVERHG--KVKIGGLKIVYMME
Barley   3   EVLIIEGIEVADHFNKFIKFDVLVNEP--E--SGGD--GASGYCAGSVALTPHMVRTNEKK----GSVKTVARFGVCDLMDNIGADG-DKTVVVSLVERCGGELVTVGGVSVGYAK-
Millet   4   EVLCVEGIEVADHLK-FVKFDVFVNVP--SQSGAAT-PAA-AECAGSVALTPHAVRPGEGG----RAVRTAARFGICDLLDDIGADG-EKTIVVSLVERCAGDTVTVGGVRIQYVK-
Millet   5   EVLCVEGIEVADHLK-FVKFDVFVNVP--SQSGAAT-PAA-AECAGSVALTPHAVRPGEGG----RAVRTAARFGICDLLDDIGADG-EKTIVVSLVERCAGDTVTVGGVRIQYVK-
Millet   6   EVLCVEGIEVADHLK-FVKFDVFVNVP--SQSGAAT-PAA-AECAGSVALTPHAVRPGEGG----RAVRTAARFGICDLLDDIGADG-EKTIVVSLVERCAGDTVTVGGVRIQYVK-
Millet   7   EVLCVEGIEVADHLK-FVKFDVFVNVP--SQSGAAT-PAA-AECAGSVALTPHAVRPGEGG----RAVRTAARFGICDLLDDIGADG-EKTIVVSLVERCAVDTVTVGGVRIQYVK-
Maize    5   EVLVIDGIEVDRDVA--AKFDVFVNTE--DHGAVGS---GGRELAGSFVNVPHRHGHGHGHAKKGRGIKTKLRLALNEQLEDLEAEG-DESVVVTLVERQGKGRVKIGGVKIELMH-
Sorghum  9   EVLVVDGIEAD--GADFVKFDVYVNAV--DHEKVGP---GGREMAGSFVSLKHP--------GEVV-VQTSMRVALKEILEDLGAEG-DDSVTVTLVEVEG--RVRIGGLRIVYMTE
Sorghum  6   EVLVVEGIEAD--AGDFVRFDVYVNAR--DYHRVPS---GGREMAGSFVTLKHPG-------KEGTALRTRMTVALNELLEDLGAEG-DDSVTVTLVEVTG--QVTIGGLRIVYMDE
Sorghum  8   EVLVIDGIEVDRDVA--AKFDVFVNTE--DHAAVGS---GGRELAGSFVNVPHGHGHRHGHGKKGRGIKTKLRLALNEQLEDLKAEG-DESVVVTLVERQGKGKVKVGGVKIELMH-
Sorghum  5   EVLVVEGIEVADHFSRFVKFDVFVNE---CQGGGGM-GAAVAQCAGSVAMTPHLVRPGKGR----GSVKTAARFGICDLLDNIGADG-DKTIVVSLVERCAGDMVTVGGVRIEYVK-
Sorghum  4   QLLVIDGIEFDPQAN--NKFDVAINLPADKALQVGP---QYKEYAGSFAVVP-------GS-GGGETRIGKVTLCITDVLYDLDAED-DSSVDVVIVERTDAKVTINVRPIIKNRN-
Sorghum  3   LVLLIEGIEYDPQIN--NKFDVIINVPKEDAGKVGP---KDCEYAGSFTAVP-------SSNAAGGTLVGKVTLFIDGVLADLGLTN-ESAVDIVLVEHTDEEIKIYLPPTIENA--
Sorghum  2   QLLVIDGIEFDPQAN--NKFDVAINVPADKASQVGP---QYKEYAGSFAVVP-------GS-GGGGTRKGKVTLCITDVLYDLDAED-DSSVDVVIVERTDAKVTINARPTIKNRN-
Maize    4   LVLLIEGIEYDPQIN--NKFDVVINVAREDAARVGP---KDSEYAGSFSAVP-------SSNAAGGTLVGKFTLALDGVLADLGLAG-ASAVDIVLVEHTEGEIKLYLPPTIENA--
Maize    3   EVLVVDGIQVADHLR-YVKFDVFVN-----QCAGGDDSTAAAECAGSFVLTPHVIQQKEEGGG-GSPVKTAARFGITDLLDDIGADG-DVTIVVSLVERCAGDVVTVDGVSVTYVK-
Sorghum  7   EVLVVEGIEVADHFSRFVKFDVFVNE---CQGGGGM-GAAVAQCAGSVAMTPHLVRPGKGR----GSVKTAARFGICDLLDNIGADG-DKTIVVSLVERCAGDMVTVGGVRIEYVK-
Millet   3   EVLVVEGIEAH--GGDFVRFDVYVNAV--EHEKVSP---GAREMAGSFVSLKQPRM--EAAVGEVASVQTSMRVALDELLEDLGADG-DDSVTVTLVEVAG--RVRIGGLRIVYMVE
Millet   1   EVLVVDGIEVDRDVA--AKFDVFVNAE--DHGAVGS---GGRELAGSFVNVPHRHAHGHGRGKKGGGIKTTLRLALNEQLEDLEAED-DESVEVTLVERQGRGKVKVGSVRIELVR-
Millet   2   TILVFDGVEFEPGKG--GKFDVVINVPPEQAAGAGP---RHSEYAGSFATLPRG-----GSKKPGETVVVPFVLPLDEVLADIGVGDEDGAVNVVIVERTPG-IKIISPPRIEIRER
```

Fig. 5- Conserved KFDV domain of different PPOs as revealed by 'Molecular Evolutionary Genetics Analysis (MEGA 6.0)'.

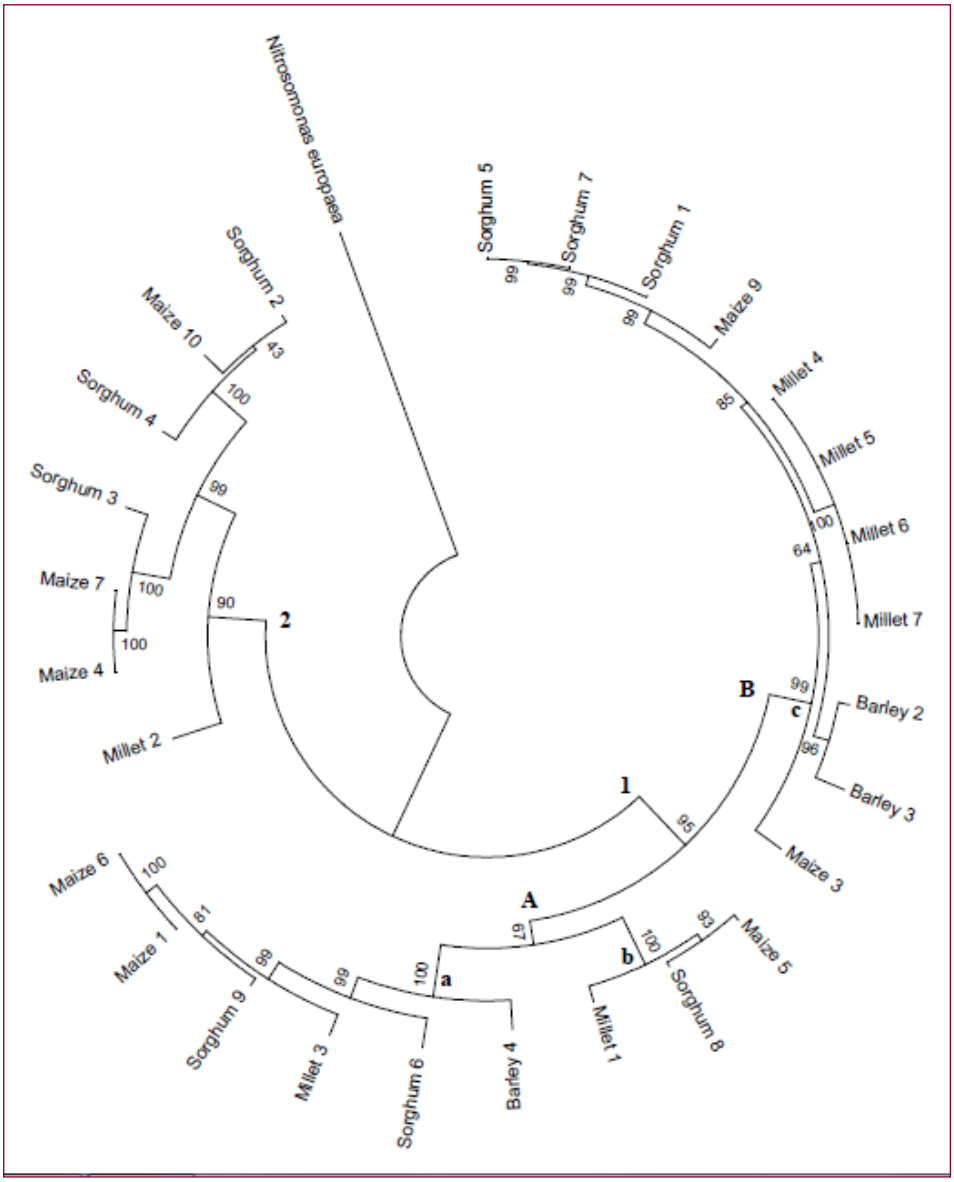

Fig. 6- Phylogenetic relationships of different cereal PPOs.

Table 1- Chloroplast transit peptide (cTP) identified among different cereal PPOs using ChloroP 1.1

Sequence	GenBank Accession	Length	cTP	cTP- length
Maize 3	NP_001149552	480	Y	19
Maize 4	NP_001142971	645	Y	46
Maize 5	DAA57297	597	Y	39
Maize 7	AFW61050	645	Y	46
Maize 9	AFW59173	546	Y	52
Maize 10	AFW56923	613	Y	36
Millet 1	XP_004972211	594	Y	37
Millet 2	XP_004974375	600	Y	40
Millet 4	BAP82254	571	Y	44
Millet 5	BAP82255	571	Y	44
Millet 6	BAP82256	571	Y	44
Millet 7	BAP82257	571	Y	44
Barley 3	BAK03848	581	Y	15
Sorghum 1	AHX26180	554	Y	26
Sorghum 2	EES14670	623	Y	38
Sorghum 3	EES13482	659	Y	50
Sorghum 4	EES13481	616	Y	41
Sorghum 5	EES11272	566	Y	38
Sorghum 7	XP_002446944	566	Y	38
Sorghum 8	EES03743	598	Y	39

Note: 'Y' and '-' refers to presence and absence of cTP (chloroplast transit peptide) Domain respectively

Phylogenetic analysis revealed two major clades [1,2] [Fig-2],[Fig-6]. The 1st larger clade is divided into two smaller clades (1A/1B), which are further sub-divided (a, b, c). Interestingly, the nodes at the base of both larger and smaller clades were mostly well supported and robust (boot strap values >80%). The sequences of smaller clade '1Aa' (Maize 1, 6; Sorghum 6, 9; Barley 4; and Millet 3) are all secretory in nature with no cTP [Fig-2],[Fig-6],[Table-2]. Most of the sequences (except 'Maize 10' and 'Sorghum 4') in clade '2' have LSP signature profile. Further, millet sequences in which the twin-tyrosine motif is not conserved (FTY instead of YxY; 'Millet 4, 5, 6, 7') are clustered in to a distinct clade [Fig-6]. It is observed that none of the analyzed cereal PPOs form a separate clade based on species specificity except for some sequences in millet (4, 5, 6, 7)and sorghum (1, 5, 7) [Fig-2]. But, several of the analysed PPO sequences from across species showed high percentage of sequence similarity. Some of the PPOs showed single amino acid change in the entire polypeptide sequence ex. sorghum PPOs 5/7, millet PPOs 4/5/6/7 and maize PPOs 4/7 (data not shown). Further, sorghum PPO 1 had a 13 amino acid difference with sorghum PPOs 5/7. Similarly, other PPO sequences like maize 1 and 6 had only 4 amino acids difference in the entire polypeptide length. Barley PPOs 2/3, sorghum 2 and maize PPO 3 are quite variable regions in the entire polypeptide length in comparison to most other sequences in the clade and this is also shown by clustering of sequences [Fig-2]. A 43 amino acid difference was observed between maize 5 and sorghum 8.

Table 2-Target peptides identified among different cereal PPOs using TargetP 1.1

Sequence	GenBank Accession	Length	Localization	TP length
Maize 1	ACG28948	569	S	26
Maize 6	AFW87458	569	S	26
Barley 2	BAJ10871	575	M	37
Barley 4	BAJ90144	580	S	22
Millet 3	XP_004965883	604	M	19
Sorghum 6	EES03688	591	S	25
Sorghum 9	EER90048	596	S	41

Note : TP length refers to Target Peptide length.
'S' and 'M' refers to Secretary and Mitochondrial Localization.

Table 3- Different domains/profiles identified among the different cereal PPOs using 'PROSITE Scan'

Sequence	GenBank Accession	Zinc Finger	TAT signal	Ig-like domain	CuA	CuB	Insect LSPs(*)	AA-tRNA synthetases
Maize 1	ACG28948	1	0	0	1	1	0	0
Maize 3	NP_001149552	0	0	0	0	1	0	0
Maize 4	NP_001142971	0	0	1	0	1	1	0
Maize 5	DAA57297	0	0	0	1	1	0	0
Maize 6	AFW87458	1	0	0	1	1	0	0
Maize 7	AFW61050	0	0	1	0	1	1	0
Maize 9	AFW59173	0	0	0	1	1	0	0
Maize 10	AFW56923	0	0	0	0	1	0	0
Millet 1	XP_004972211	0	1	0	1	1	0	0
Millet 2	XP_004974375	0	0	0	0	1	1	0
Millet 3	XP_004965883	0	0	0	1	1	0	0
Millet 4	BAP82254	0	0	0	1	1	0	0
Millet 5	BAP82255	0	1	0	1	1	0	0
Millet 6	BAP82256	0	0	0	1	1	0	0
Millet 7	BAP82257	0	1	0	1	1	0	0
Barley 2	BAJ10871	0	0	0	1	1	0	0
Barley 3	BAK03848	0	0	0	1	1	0	0
Barley 4	BAJ90144	1	0	0	1	1	0	0
Sorghum 1	AHX26180	0	1	0	1	1	0	0
Sorghum 2	EES14670	0	0	0	0	1	1	0
Sorghum 3	EES13482	0	0	0	0	1	1	0
Sorghum 4	EES13481	0	0	0	0	1	0	0
Sorghum 5	EES11272	0	1	0	1	1	0	0
Sorghum 6	EES03688	0	0	0	1	1	0	1
Sorghum 7	XP_002446944	0	1	0	1	1	0	0
Sorghum 8	EES03743	0	0	0	1	1	0	0
Sorghum 9	EER90048	0	0	0	1	1	0	0

Note: '0' and '1' - refers to absence or presence of domain as per PROSITE Scan respectively

Discussion

In our evaluation of PPOs from four important cereal crops (maize, sorghum, barley and foxtail millet), several previously uncharacterized PPOs were identified. Substantial diversity was observed within and among the different crops both in number and sequence identity. The divergence/similarity among the analyzed PPOs to some extent can be explained based on the modifications within in different domains: 'Cu-A and B', 'DWL' and KFDV; motifs: 'HxxYC' and 'HRxYxxFxER' and 'HxxxH'. It was observed that six PPOs (sorghum 2/3/4 and maize 4/7/10) containing 'HRMYIYFYER' motif were grouped into a separate clade [Fig-2],[Fig-6]. Similarly, 5 millet PPOs (1/4/5/6/7), 4 sorghum PPOs (1/5/8/7), barley 2 and maize 5 had a common 'HGPVH' motif, clustering them into same sub-clade (1Bc) [Fig-2],[Fig-6]. 'FTY' motif in 'DWL' domain and modification of 'KFDV' domain ('RFDV') was observed among four millet PPOs (4/5/6/7) and sorghum 6/millet 3 respectively [Fig-4],[Fig-5]. These differences/similarities coupled with random amino acid variations as explained in the results could be responsible for the divergence among the sequences and these could probably result in differences in functions. Interestingly, several PPOs were found to be secretory in nature [Table-2], which has been experimentally shown only in snapdragon and poplar [16,17]. But, predicted non-plastidic PPOs are found in several crops including rice, maize and blue columbine (*Aquilegia coerulea*) [11]. Several studies suggested the presence of two domains at the C-terminal end of PPOs: PPO_DWL and PPO_KFDV. Thought the functional importance of these two domains is yet to be ascertained, it is observed that C-terminal based-proteolytic processing (if it occurs) occurs in the PPO_DWL domain immediately after the twin tyrosine (YxY) motif [32]. This processing results in a loss of ~16-18 kDa polypeptide fragments including the PPO_KFDV domain [19].

As previously observed in other plants [33], multi-gene families were observed for all the four crops [Table-3]. Tran et al. [11] has reported that monocots typically contain 2-8 PPOs, our results presented here are in agreement [Table-3]. Additionally, several well-supported PPO clades (bootstrap values >70%) for several monocot PPOs coupled with at least three PPO genes in the common ancestor of modern grasses was observed. Interestingly, in eudicot species (*Glycine*, *Populus* and *Mimulus*) independent PPO gene diversification along with presence of several PPOs in their common ancestor was observed [11]. We observed that many of the PPOs across the four cereal species have PPOs intermingled on separate well-supported clades/branches. It is evident from our analysis that the ancestor of the four cereals analysed have multiple PPO proteins.

The number of PPOs reported here could be a conservative estimate as some shorter length PPO proteins were not considered for analysis. Despite the ambiguity regarding their exact number in different crops, the variability (in number) of PPOs observed here in different cereal crops is interesting as this variability is not observed in other oxidative enzymes. Further, PPO diversification in the four cereal crops may not be a consequence of species-specific gene duplication and divergence. This could perhaps be a reflection of different functions or clade-specific ecological/metabolic selection pressure. But, species possessing complex phenol-based secondary metabolism could contain duplicated PPO proteins. Among the cereals, sorghum is reported to contain high levels of phenolics and largest PPO family [34].

PPOs have been implicated in carrying out different functions in plants and this is demonstrated in their features including variation in numbers, structure, localization and lineage-specific diversification. PPOs could play diverse physiological roles owing to their diverse substrate specificity (tyrosine, catechol, phenol, and L-3,4-dihydroxyphenyl alanine (L-DOPA) [35] in addition to their role in cereal product quality. The primary cause of time-dependent discoloration of cereal-based products like white-salted and yellow alkaline wheat noodles are PPOs [7]. Over-expression or silencing of PPOs in tomato resulted either in decreased or increased susceptibility to pathogens respectively [36,37]. Further, PPOs are also reported to be in production of aurone and betalin pigments in Snapdragon (*Antirrhinum majus*) and order Caryophyllales respectively [38,39]. As evidenced from our data PPO family is significantly expanded in some crop species like sorghum, maize or millet but reduced in others like barley. Perhaps, this could be a strategy to match the distribution of secondary metabolites and different roles in plants, which also varies among the different plant species and appears to be governed by duplication and diversification of genes [40].

Conclusion

Our study of PPOs in four cereal crops identified significant diversity in gene family size and structure. The PPO protein diversity is not species-specific as evidenced by our phylogenetic analysis. Additionally, we have identified some interesting features among the 27 sequences, not reported earlier. Overall the dynamic nature of PPO gene family in the analyzed cereals is consistent with its implications in diverse potential roles in plants.

Conflicts of Interest: None declared.

References

[1] Mayer A.M. (1987) *Phytochemistry*, 26(1), 11-20.

[2] Sommer A., Ne'eman E., Steffens J.C., Mayer A.M. & Harel E. (1994) *Plant Physiology*, 105, 1301-1311.

[3] Koussevitzky S., Ne'eman E., Sommer A., Steffens J.C. & Harel E. (1998) *Journal of Biological Chemistry,* 273, 27064-27069.

[4] Shetty S.M., Chandrashekar A. & Venkatesh Y.P. (2011) *Phytochemistry,* 72, 2275-2287.

[5] Anderson J.A. & Morris C.F. (2001) *Crop Science*, 41, 1697-1705.

[6] Constabel C.P., Barbehenn R.V. (2008) *Defensive roles of polyphenol oxidase in plants*, Induced plant resistance to herbivory, Springer Verlag, New York, NY, USA, 253-269.

[7] Baik B.K., Czuchajowska Z. & Pomeranz Y. (1994) *Journal of Cereal Science,* 19, 291-296

[8] Thipyapong P., Hunt M.D. & Steffens J.C. (1995) *Phytochemistry,* 40, 673-676.

[9] Constabel C.P. & Ryan C.A. (1998) *Phytochemistry,* 47, 507-511.

[10] Wahler D., Gronover C.S., Richter C., Foucu F., Twyman R.M., Moerschbacher B.M., Fischer R., Muth J. & Prüfer D. (2009) *Plant Physiology,* 151, 334-346.

[11] Tran L.T., Taylor J.S. & Constabel C.P. (2012) *BMC Genomics*, 13, 395.

[12] Steiner U., Schliemann W., Böhm H. & Strack D. (1999) *Planta,* 208, 114-124.

[13]Ryan C.A. (2000) *Biochimica et Biophysica Acta* 1477, 112-121.

[14]van Gelder C.W.G., Flurkey W.H. & Wichers H.J. (1997) *Phytochemistry,* 45, 1309-1323.

[15]Koussevitzky S., Ne'eman E., Peleg S. & Harel E. (2008) *Physiologia Plantarum,* 133, 266-277.

[16]Ono E., Hatayama M., Isono Y., Sato T., Watanabe R., Yonekura-Sakakibara K., Fukuchi-Mizutani M., Tanaka Y., Kusumi T., Nishino T. & Nakayama T. (2006) *Plant Journal,* 45, 133-143,

[17]Tran L.T. & Constabel C.P. (2011) *Planta*, 234, 799-813,

[18]Klabunde T., Eicken C., Sacchettini J.C. & Krebs B. (1998) *Nature Structural & Molecular Biology,* 5, 1084-1090.

[19]Robinson S.P. & Dry I.B. (1992) *Plant Physiology,* 99, 317-323.

[20]Dry I.B. & Robinson S.P. (1994) *Plant Molecular Biology,* 26, 495-502.

[21]Newman S.M., Eannetta N.T., Yu H., Prince J.P., de Vicente M.C., Tanksley S.D. & Steffens J.C. (1993) *Plant Molecular Biology,* 21, 1035-1051.

[22]Thygesen P.W., Dry I.B. & Robinson S.P. (1995) *Plant Physiology,* 109, 525-531.

[23]Massa A.N., Beecher B. & Morris C.F. (2007) *Theoretical Applied Genetics,* 114, 1239-1247.

[24]Thompson J.D., Higgins D.G. & Gibson T.J. (1994) *Nucleic Acids Research,* 11(22), 4673-4680.

[25]Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. (2013) *Molecular Biology and Evolution,* 30, 2725-2729.

[26]Gattiker, A., Gasteiger E. & Bairoch A. (2002) *Applied Bioinformatics*, 1(2), 107-108.

[27]de Castro E., Sigrist C.J.A., Gattiker A., Bulliard V., Langendijk-Genevaux P.S., Gasteiger E., Bairoch A. & Hulo N. (2006) *Nucleic Acids Research,* 34, W362-W365.

[28]Emanuelsson O., Nielsen H., Brunak S. & von Heijn G. (2000) *Journal of Molecular Biology,* 300, 1005-1016.

[29]Jones D.T., Taylor W.R. & Thornton J.M. (1992) *Computer Applications in the Biosciences,* 8, 275-282.

[30]Felsenstein J. (1985) *Evolution,* 39, 783-791.

[31]Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) *Molecular Biology & Evolution,* 28, 2731-2739.

[32]Marusek C.M., Trobaugh N.M., Flurkey W.M. & Inlow J.K. (2006) *Journal of Inorganic Biochemistry,* 100, 108-123.

[33]Richter H., Lieberei R., Strnad M., Novák O., Gruz J., Rensing S.A. & von Schwartzenberg K. (2012) *Journal of Experimental Botany,* 63(14), 5121-5135.

[34]Awika J.M. & Rooney L.W. (2004) *Phytochemistry,* 65, 1199-1221.

[35]Nilthong S., Graybosch R. A. & Baenziger P.S. (2012) *Theoretical & Applied Genetics,* 125(8), 1705-1715.

[36]Li L. & Steffens J.C. (2002) *Planta,* 215, 239-247.

[37]Thipyapong P., Hunt M.D. & Steffens J.C. (2004) *Planta,* 220, 105-117.

[38]Nakayama T., Sato T., Fukui Y., Yonekura-Sakakibara K., Hayashi H., Tanaka Y., Kusumi T. & Nishino T. (2001). *FEBS Letters,* 499, 107-111.

[39]Gandía-Herrero F., Escribano J. & García-Carmona F. (2005) *Plant Physiology,* 138, 421-432.

[40]Ober D. (2010) *Plant Biology,* 12, 570-577.