

Polyphenol oxidase (PPO) in wheat and wild relatives: molecular evidence for a multigene family

Alicia N. Massa · Brian Beecher · Craig F. Morris

Received: 12 August 2006 / Accepted: 22 January 2007 / Published online: 14 February 2007
© Springer-Verlag 2007

Abstract Wheat polyphenol oxidase (PPO) is the major cause of browning reactions that discolor Asian noodles and other wheat products. It has been hypothesized that genes encoding wheat PPOs may have evolved by gene duplication into a multigene family. Here we characterized PPO genomic sequences from diploid (*Triticum monococcum*, *T. urartu*, *Aegilops tauschii*, and *Ae. speltoides*), tetraploid (*T. turgidum*, subspecies *dicoccoides* and *durum*) and hexaploid (*T. aestivum* cultivars Klasic and ID377s) wheat species to gain a better understanding of the structure and organization of PPO genes. DNA fragments were amplified from a highly polymorphic and phylogenetic informative region of the gene. As a result, we obtained highly discriminative sequences. Three distinct PPOs, obtained from the A genome of *T. monococcum*, provided evidence for gene duplication events

(paralogous loci). Furthermore, the number of sequences obtained for bread and durum wheat was higher than the expected number of orthologous loci. Sequence comparison revealed nucleotide and structural diversity, and detected five sequence intron types, all with a common insertion position. This was hypothesized to be homologous to that of intron 2 of previously reported wheat PPOs. A MITE of the *Stowaway* family accounted for the major difference between the five intervening sequences, and was unique to *T. aestivum* cv. Klasic. Nucleotide and structural diversity, together with well-resolved phylogenetic trees, provided molecular evidence to support the hypothesis of a PPO multigene family structure and organization.

Introduction

Polyphenol oxidase (PPO) is the major cause of enzymatic discoloration in Asian noodles and other wheat-based end products (Baik et al. 1994; Morris et al. 2000). This enzyme catalyzes the oxidation of phenols to form quinones in the presence of molecular oxygen (Mayer and Harel 1979). The resulting quinones can then polymerize to form high-molecular weight black or brown pigments. Darkening and discoloration affect consumer acceptance of wheat products, especially yellow alkaline and white salted noodles (Morris et al. 2000).

The structure and organization of genes encoding PPOs have been intensively characterized in a diverse range of higher plants (Sherman et al. 1995). Most angiosperm PPOs appear to be arranged in multigene families, although the structure of the PPO genes seems to differ between monocots and dicots. Dicotyledonous species such as tomato (Newman et al. 1993), potato

Communicated by A. Kilian.

Mention of trademark or proprietary products does not constitute a guarantee or warranty of a product by the US Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable. This article is in the public domain and not copyrightable.

A. N. Massa
Department of Crop and Soil Sciences (affiliated with the USDA ARS Western Wheat Quality Laboratory),
Washington State University, Pullman,
WA 99164-6394, USA

B. Beecher · C. F. Morris (✉)
USDA ARS Western Wheat Quality Laboratory,
E-202 Food Science and Human Nutrition Facility East,
Washington State University, Pullman,
WA 99164-6394, USA
e-mail: morrisc@wsu.edu

(Thygesen et al. 1995), and red clover (Sullivan et al. 2004) contain intronless genes, while monocots like pineapple (Zhou et al. 2003) and banana (Gooding et al. 2001) carry genes with intervening sequences. In wheat, Sun et al. (2005) reported wheat PPO sequences from genomic DNA harboring two introns each. Thus, it appears that insertion of introns into PPO genes occurred after the divergence of monocots and dicots.

It has been hypothesized that genes encoding wheat PPOs may have evolved by gene duplication into a multigene family (Jukanti et al. 2004). However, complex genomes like allopolyploid wheat species make the detection of paralogous loci more difficult. Efforts have been made to identify genes related to PPO activity in both durum (Simeone et al. 2002; Watanabe et al. 2004) and bread wheat (Anderson and Morris 2003; Demeke et al. 2001; Demeke and Morris 2002; Jimenez and Dubcovsky 1999). Southern blot analyses have suggested that wheat contains more than one PPO sequence (Demeke and Morris 2002). Nevertheless, most of the genes identified so far have been associated with orthologs, which are located on homoeologous chromosomes, rather than paralogs (Jimenez and Dubcovsky 1999; Demeke and Morris 2002). A recent study based on sequence data suggested that wheat PPO genes are arranged in a multigene family consisting of two distinct phylogenetic groups with three members each (Jukanti et al. 2004). Genes from one group are expressed in developing kernels while genes of the other group are expressed in plant tissues other than seeds (Jukanti et al. 2006). Sequence similarity, as well as functional diversity, supports the hypothesis that these two groups represent paralogous loci. However, there is no molecular genetic or phylogenetic evidence to support the hypothesis that members within either the seed or non-seed gene clusters arose by gene duplication. Most genes in hexaploid wheat exist as triplicate sets of single-copy homoeologs from the A, B, and D genomes, and therefore represent orthologous rather than paralogous loci.

The objective of this study was to identify and describe novel PPO sequences from diploid, tetraploid, and hexaploid wheat species to gain a better understanding of the PPO gene family structure and organization. We were particularly interested in the identification and characterization of PPO genes associated with PPOs found in developing kernels.

Materials and methods

Genomic DNA was extracted from leaf tissue according to the method of Murray and Thompson

(1980) using a single plant per accession. Plant material consisted of four diploids: *Triticum urartu* Thüm. ex Gandilyan, *Triticum monococcum* L., *Aegilops speltoides* Tausch, and *Aegilops tauschii* Coss.; two tetraploids: *Triticum turgidum* ssp. *durum* (Körn. ex Asch. and Graebn.) Thell. cultivar Langdon and *T. turgidum* ssp. *dicoccoides* (Desf.) Husn.; and two hexaploid wheat cultivars: ID377s and Klasic (Table 1). The analyses also included publicly available DNA sequences of *T. aestivum* retrieved from GenBank (Table 1). PPO gene sequences were amplified by PCR, and products were cloned into a pCR 2.1 cloning vector (TA Cloning kit, Invitrogen, Carlsbad, CA, USA). PCR reactions consisted of 50 µl containing 5 µl of 10 × *Taq* buffer, 250 µM of each deoxynucleoside triphosphate, 1.5 U *Taq* DNA polymerase, 10 pmol of each primer, 40.55 µl of sterile water, and 2 µl of the DNA. The reactions were completed in 41 cycles of denaturation at 94°C for 30 s, primer annealing at 68°C for 60 s, and DNA elongation at 72°C for 120 s. Sequencing reactions were carried out with the Big Dye Terminator Version 3.1 Cycle Sequencing Kit and sequenced using an Applied Biosystems 3100 Genetic Analyzer (Perkin-Elmer Applied Biosystems Division, Foster City, CA, USA).

All sequences were aligned with ClustalW, version 1.83 (Thompson et al. 1997). The splicing sites of introns were identified by the alignment of genomic DNA to EST/cDNA sequences using BLAST search. Gene trees were generated by maximum likelihood and neighbor-joining algorithms as implemented in PAUP* 4.0 beta 10 (Swofford 2002) and bootstrap was performed with 1,000 replicates using the program MEGA version 2.1 (Kumar et al. 2004, <http://www.megasoftware.net/index.html>). A sequence variability profile of PPO was analyzed using the Sequence Variability Analysis Program (SVARAP, <http://ifr48.free.fr/recherche/labo/rickettsies/svarap/svarap.htm>). The variability along multiple nucleotide sequence alignments was calculated as the proportion of sequences that differ from the consensus sequence at a given site (Khamis et al. 2003; Colson et al. 2006). Transposable elements (TEs) were searched by comparison with the Triticeae Repeat Sequence Database (TREP, <http://wheat.pw.usda.gov/ITMI/Repeats/index.shtml>; Wicker et al. 2002). The secondary structure of TEs was predicted using the MFOLD software package (<http://www.bioinfo.rpi.edu/applications/mfold/dna/form1.cgi>; Zuker 2003). All sequences obtained here were deposited in GenBank under accession numbers DQ889690–DQ889710 (Table 1).

Table 1 Summary of plant material used in this study

Species	Genome	Accession ^a identifier	No. of clones sampled	Primer combinations	GenBank accession No.
<i>Triticum urartu</i>	A ^u	TA763	5	ES3/EA2, 759S1/1456A1, 841S1/1456A1	DQ889697
<i>Triticum monococcum</i>	A ^m	TA2026, cv. DV92	9	817S1/1488A1, 841S1/1456A1, ES1/EA2, ES2/EA2, ES3/EA2, 759S1/1456A1	DQ889694, DQ889695, DQ889701, DQ889710
<i>Aegilops tauschii</i>	D	TA2442	9	841S1/1456A1, 759S1/1456A1, 817S1/1456A1, ES1/EA2, ES2/EA2, ES3/EA2	DQ889696
<i>Ae. speltooides</i> var. <i>ligustica</i>	S	TA2779	6	ES3/EA2, 817S1/1456A1, 841S1/1456A1	DQ889703
<i>Triticum turgidum</i> ssp. <i>dicoccoides</i>	AB	TA84	8	817S1/1456A1, 817S1/1488A1, ES3/EA2, ES1/EA2	DQ889707
<i>Triticum turgidum</i> ssp. <i>durum</i>	AB	cv. Langdon, C1tr13165	10	817S1/1456A1, ES3/EA2	DQ889690, DQ889702, DQ889704
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	cv. Klasic	21	841S1/1488A1, ES3/EA2, 817S1/1488A1	DQ889691, DQ889698, DQ889699, DQ889706, DQ889708
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	cv. ID377s	18	817S1/1488A1, ES2/EA2, 759S1/1488A1, 841S1/1456A1	DQ889692, DQ889693, DQ889700, DQ889705, DQ889709
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	cv. Chinese Spring			AY596266
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	–			AY596267 ^b
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	–			AY596268 ^b
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	cv. Sumai3			AY596270 ^b
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	cv. Butte 86			AY515506 ^c
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	cv. Penawawa			AF507945 ^d
<i>Triticum aestivum</i> ssp. <i>aestivum</i>	ABD	–			BT009357 ^e

^a Accession identifiers: TA Wheat Genetics Resources Center, Kansas State University, C1tr National Plant Germplasm System, USDA-ARS

^b Jukanti et al. (2004)

^c Anderson and Morris (2003)

^d Demeke and Morris (2002)

^e Tingey SV, Wolters P, Powell W, Dolan M, Miao G-H, Caraher NR, Hanafey MK, Haimey CF (2003) Crop Genetics, E. I. DuPont de Nemours and Company, 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Results

First, we characterized the nucleotide variability of nearly the entire PPO coding region (~1,400 bp, excluding introns and gaps) on the basis of the alignment of five seed-specific PPO sequences (GenBank accessions AY515506 and BT009357, and B. Beecher, unpublished data). The mean variability for sliding windows of 50 nucleotides revealed two highly polymorphic areas with mean values up to 12.8% for the first region (~650–1,000 bp), and 17.2% for the second region (~1,050–1,650 bp). Moreover, they were flanked by conserved sequences with mean variability ranging from 0.8 to 2.8% (Fig. 1 a). Conserved areas were used further for primer design (Table 2).

A total of nine primer combinations allowed us to amplify and sequence one of the highly variable regions, between nucleotide positions 700 and 1,200

(Fig. 1a). As a result, we obtained a total of 21 distinct PPO sequences of ~600–900 bp long from diploid (seven), tetraploid (four), and hexaploid (ten) wheat accessions (Table 1). The analysis of sequence variability based on partial PPOs revealed a pattern of nucleotide polymorphism very similar to that observed in the analysis of the five PPO sequences (Fig. 1a). Furthermore, 78% (77) of the variable sites were parsimony informative (non-singletons). Therefore, by comparing DNA segments from a highly polymorphic and phylogenetic informative region of the gene, we were able to infer phylogenetic trees that were very similar to those based on the entire coding sequence.

Structural diversity

In addition to nucleotide sequence polymorphism, the amplified region presented an intron presence–absence

Fig. 1 a Sequence variability of seed-specific PPO genes (excluding introns and gaps). The y-axis indicates the mean variability per window of 50 nucleotides and the x-axis specifies the positions of nucleotides. The mean variability across nearly the entire coding region was derived from five PPO sequences (~1,400 bp). Mean variability of partial PPOs (in *bold*) were derived from 21 sequences (~500 bp). Arrows indicate primer binding sites. **b** Structure of the partial PPO gene region analyzed in this study (*shaded*) showing the intron insertion site

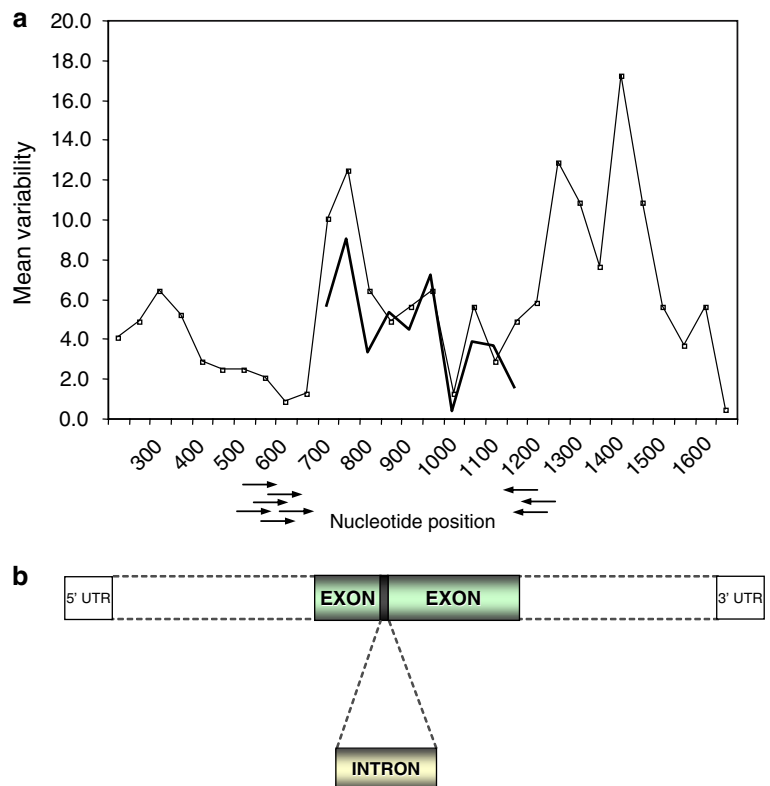


Table 2 Primers used in this study to amplify partial PPO sequences

ID	Sequence 5'-3'	T_m (°C)
ES1	AGCAGTGGCTCGTGCCTACTG	78
ES2	TCTCGGCAAGTCATCGGCGACGACAC	82
EA2	ACCAGGACGTGGCCTGCCGTGGCTTA	80
ES3	CGTGCCGGCGATCTACGCCAACAGGTC	81
759S1	CAAGTCATCGGCGACGACACCTT	72
817S1	GCATGACGCTGCCGGCGATCTAC	74
841S1	CCAACAGGTCGTCGCCGCTCTAC	74
1456A1	GCGGCCGGTTCGAGGCAGT	74
1488A1	CGGCAGGCCGACGTCCTGGT	72

polymorphism with partial PPOs consisting of two different structures: a single exon and an exon containing a single intron (intervening sequence). The alignment of sequences with an intron–exon organization revealed five intron sequence types of variable length (98–360 bp long) and nucleotide composition (Fig. 2). They were designated as *a* through *e* (Fig. 2). Intron sequences *a*, *b*, and *c* were unique, whereas introns *d* and *e* occurred in more than one taxon (Fig. 3). Interestingly, all introns showed a conserved insertion site within the aligned region. This was predicted at position 819/820 (the start codon as the first position) of a known mRNA (GenBank accession number BT009357) (Fig. 1b). Moreover, all introns fit into the class of non-canonical introns characterized by a GC donor dinucleotide (the two bases of the 5'-terminus) and an AG acceptor dinucleotide (last two bases of the 3'-terminus) (Fig. 2).

Intron length polymorphisms resulted from several small length differences and one large insertion (206 bp) in the *T. aestivum* cv. Klasic intron *a* (GenBank accession DQ889708, see Fig. 3). BLAST searches using this fragment against the TREP database suggested two Miniature Inverted-repeat Transposable Elements (MITEs) from the *Stowaway* family.

It is interesting that one 115-bp long element appeared to be nested within a 101-bp element (Fig. 4). While, the outer element showed a 7-bp terminal inverted repeat (TIR, 5'CTCCCTC3') and the 2-bp long TA recognition target site characteristic of *Stowaway*-type elements. The inner element lacked the 3' TA target site, and the predicted DNA secondary structure showed three mismatches at the TIRs (Fig. 4). Nevertheless, it is predicted to fold into a double hairpin-like structure (Fig. 4). The remaining introns showed no evidence of *Stowaway* elements, however, those from wheat cultivars ID3775s (DQ889709) (intron *c*) and *T. monococcum* cv. DV92 (DQ889710) (intron *b*) harbored a 5-nt consensus sequence, 5'TCCCT3, which was identical to that of the 5'TIR of the *Stowaway* element observed in intron *a* (Fig. 2).

Gene duplication

Evidence of duplicate sequences within a single genome (paralogous loci) was found in *T. monococcum* cv. DV92. We amplified and sequenced three distinct PPO fragments from this accession, two of the fragments lacked introns (DQ889695, DQ889701) and the other one (DQ889710) had a 152-bp intron within

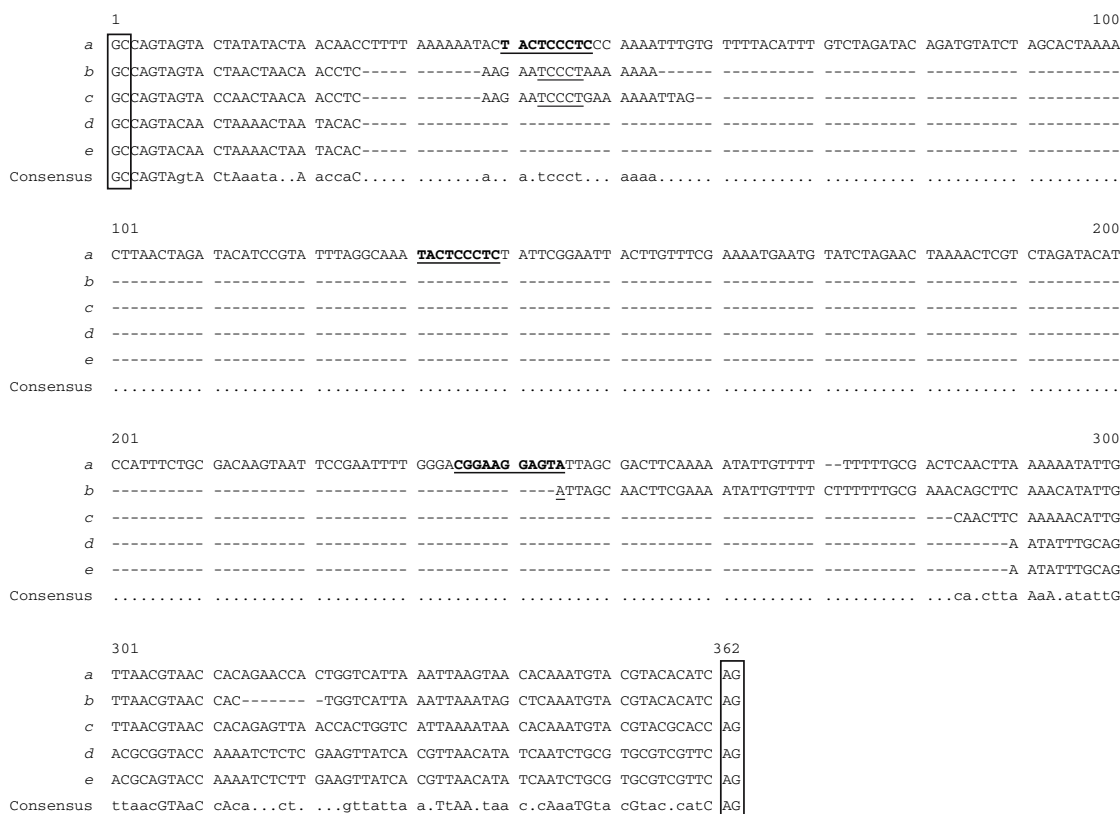


Fig. 2 Sequence alignment of introns *a* through *d*. The 5'GC and 3' AC dinucleotides at their termini are indicated with boxes. Terminal inverted repeats and TA recognition sites of the *Stowaway* element of intron sequence *a* are underlined and in bold

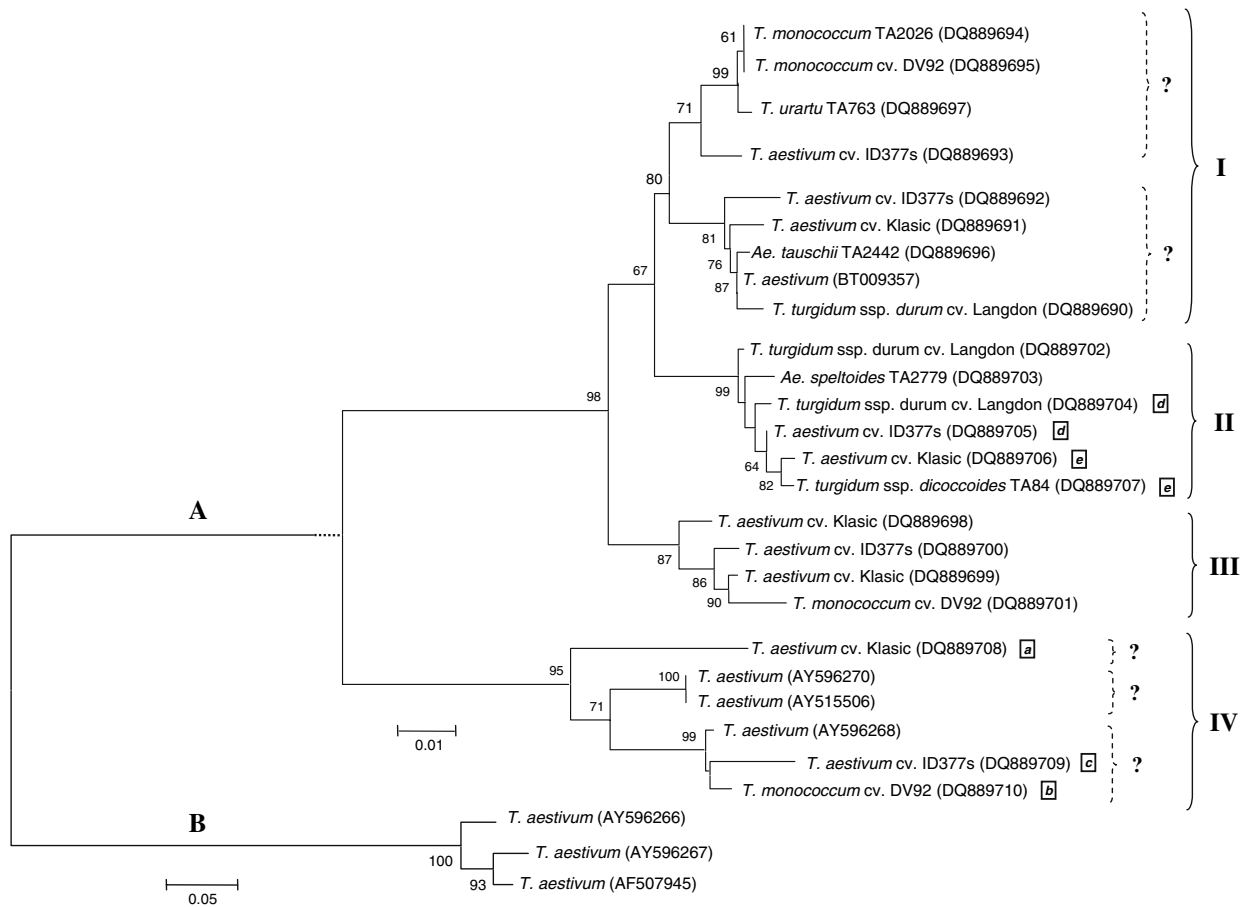


Fig. 3 Phylogenetic analysis of the PPO gene family. The tree was derived by neighbor-joining distance analysis using 28 sequences of about 440 nucleotides (excluding gaps and introns). Bootstrap values over 60% are indicated. The seed specific PPO subtree (**A**) was re-grafted to a different scale and appended to the main backbone tree as indicated by a broken horizontal line.

B Shows the branch of non-seed PPO sequences. Genes are labeled by species name, accession identifier, and GenBank accession number (in *parentheses*). Intron presence is indicated by the corresponding number in a square box. *Scale bar* indicates the number of nucleotide substitutions per site

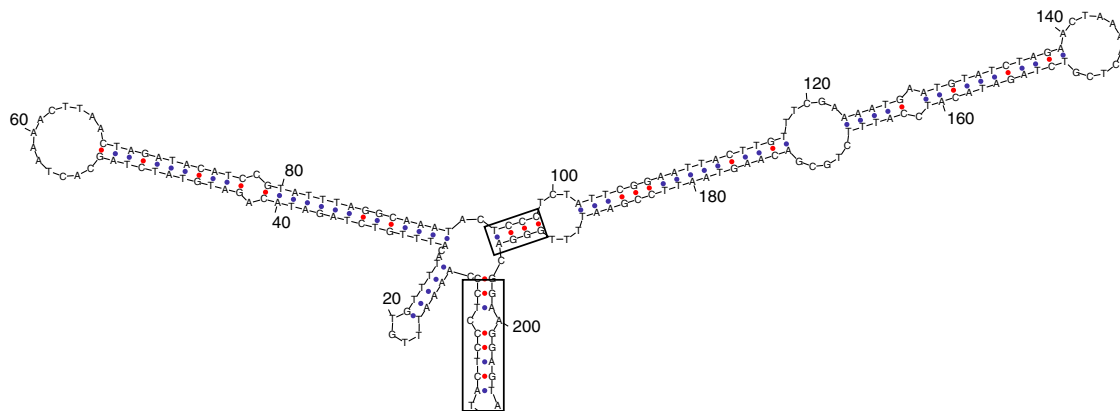


Fig. 4 Secondary structure of the double *Stowaway* element of intron sequence 1 from *T. aestivum* cv. Klasic with $\Delta G = -43.8$ kcal/mol. Terminal inverted repeats of the inner and outer element are marked by *boxes*

the target region (Fig. 3). In addition, several partial PPOs were amplified from *T. aestivum* cultivars ID377s (five) and Klasic (five), and *T. turgidum* cv. Langdon

(three). It is noteworthy that the number of copies per individual genotype was higher than the expected number of orthologs (homoeologs) based on their genome

constitution, that is, two orthologs from the tetraploid durum wheat cv. Langdon (AB) and three from the hexaploid bread wheat cultivars ID377s and Klasic (ABD), thus suggesting paralogous loci.

Evolutionary relationships

Phylogenetic analyses were based on a data set of 28 PPO sequences ~440-bp in length (excluding introns and gaps). This included seven sequences retrieved from GenBank. The latter group consisted of four seed specific PPO genes (BT009357, AY515506, AY596270, and AY596268) and three non-seed specific PPOs (AY596266, AY596267, and AF507945), which were used to root the tree (Fig. 3). Similar tree topologies resulted from maximum likelihood and neighbor joining analyses, although only the latter tree is presented (Fig. 3).

Phylogenetic inferences indicated four major clusters (bootstrap values $\geq 80\%$) within the seed specific PPO subclass (clusters I–IV; Fig. 3). Clusters I–III were more closely related to each other than they were to cluster IV. The analyses further suggested two minor groups within cluster I (α and β). Although paralogous–orthologous relationships between these two minor groups cannot be determined from the present data set, it is interesting that conserved sequences appear to characterize different genomes within α or β subgroups.

The paralogous sequences of *T. monococcum* cv. DV92 (DQ889695, DQ889701, and DQ889710) were allocated to three different clusters and suggested orthologous relationships within, and paralogous relationships between clusters I, III, and IV (Fig. 3). Interestingly, PPO fragments within cluster II, which were highly conserved (99% nucleotide similarity), had a different structure. That is, *T. turgidum* cv. Langdon (DQ889702) and *Ae. speltoides* TA2779 (DQ889703) had no introns, while the remaining four sequences carried a 98-bp intron with two highly similar sequence types (introns *d* and *e*, Figs. 2, 3).

Sequence divergence within Group IV, as indicated by branch length on the phylogenetic tree, together with intron nucleotide and structural diversity, further suggested three putative subgroups (α , β , and γ , Fig. 3). Yet, the evolutionary relationships among members of cluster IV are also supported by their introns (Fig. 5).

Discussion

In the present study, we amplified highly discriminative PPO sequences from diploid, tetraploid, and hexaploid wheat species, and provided molecular evidence to support the hypothesis of a PPO multigene family structure and organization. Nucleotide and structural diversity, together with well-resolved phylogenetic trees, allowed us to describe putative orthologous–paralogous relationships among members of this gene family.

Since most sequence data available to date is based on cDNA and ESTs, the structure of wheat PPO genes remains poorly characterized. Recently, Sun et al. (2005) amplified partial PPOs from genomic DNA and reported the occurrence of two introns in Chinese wheat PPOs. This study (Sun et al. 2005) detected intron size variation between genotypes with high and low PPO activity, but no evidence of intron presence–absence polymorphism was documented. Our results indicated structural diversity within the target region and identified two different gene structures: a single exon and an exon containing an intron. Preliminary results in our lab suggest that intronless fragments may represent single-intron genes (B. Becher, unpublished data). While, fragments carrying an intron most likely represent genes with two introns. All intervening sequences observed in the present study shared the same (putative homologous) insertion position, which was also the same as that of the second intron (intron 2) of Chinese wheat PPOs (Sun et al. 2005). Homology with intron 2 was also supported by sequence similarity and splice consensus dinucleotides. The intron 2 of

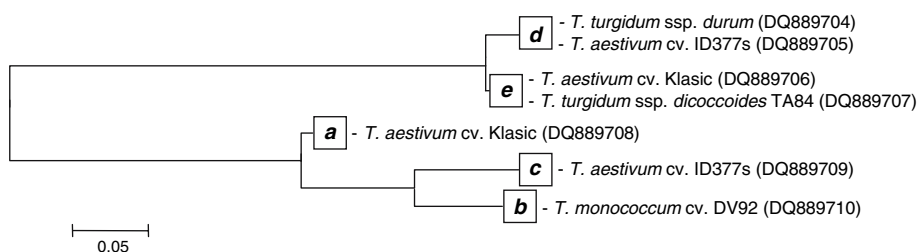


Fig. 5 Neighbor-joining tree of intron sequences (excluding gaps). Sequences are labeled as described in the text, followed by species name, accession identifier, and GenBank accession num-

ber (*in parentheses*). Scale bar indicates the number of nucleotide substitutions per site

Chinese wheat cultivars was identical to that of *T. aestivum* cv. ID377s (DQ889709) (intron *c*) and 88% similar to that of *T. monococcum* cv. DV92 (DQ889710) (intron *b*). Moreover, all introns from this study as well as intron 2 (Sun et al. 2005) fit into the GC-AG non-canonical introns.

The GC-AG splice site junctions have been previously reported in wheat *Wknox1* genes (Morimoto et al. 2005), *Arabidopsis thaliana* (Brown et al. 1996) and *Oryza sativa* (Sparks and Brendel 2005). This alternative isoform is a variant of the wild type GT-AG and accounts for only a small fraction (<2.6%) of all introns in *Arabidopsis* and rice genomes (Sparks and Brendel 2005). It has been suggested that the mechanism of splicing involving GC-AG introns is the same as that of GT-AG introns (Zhu et al. 2003). Nonetheless, it will be interesting to determine whether a T to a C mutation at the 5' dinucleotides (GT→GC) is relevant to splicing regulation of PPO genes.

Stowaway elements accounted for the major difference between the five introns. MITEs of the *Stowaway* family have been extensively characterized in plants and are ubiquitous in cereal grass genomes (Bureau and Wessler 1994a; Feschotte et al. 2002). However, elements with a double hairpin-like structure like that of *T. aestivum* cv. Klasic (DQ889708) (intron *a*) have only been described in *Heterantherium piliferum* of the tribe Triticeae (Petersen and Seberg 2000). While no further evidence of MITEs was found in the present study, the consensus sequence 5'TCCCT'3 observed in *T. monococcum* cv. DV92 (DQ889710) (intron *b*) and *T. aestivum* cv. ID377s (DQ889709) (intron *c*) (Fig. 2) suggest putative footprints left after excision of a *Stowaway*-type element. Based on evolutionary relationships, as indicated in Fig. 3, it is quite possible that the footprints observed in introns *b* and *c* are remnants from excision of the double *T. aestivum* cv. Klasic (intron *a*) element. MITE excision events appear to be rare. Nevertheless, phylogenetic evidence for excision of *Stowaway* MITEs has been previously documented in the tribe Triticeae (Petersen and Seberg 2000).

Although, *Stowaway* and other MITEs have been found to contain regulatory sequences involved in transcription initiation and polyadenylation of plant genes (Bureau and Wessler 1994a, b), their contribution to gene expression is uncertain. Based on the present results, it will be interesting to determine whether gain (or loss) of *Stowaway*-type elements within cluster IV sequences is associated with patterns of gene expression and ultimately, with differences in PPO activity.

Multigene families are thought to arise from duplication and divergence of an ancestral gene. Here we present molecular evidence of duplicate PPO

sequences (paralogous loci) from the A genome of *T. monococcum* cv. DV92 and provide support for a minimum of three family members within the PPOs expressed in developing kernels. Furthermore, these results revealed the presence of putative orthologs for each of these three paralogs indicating their occurrence in other *Triticum/Aegilops* genomes (Fig. 3). For instance, the two copies of *T. aestivum* cv. Klasic cluster III (putative orthologs to *T. monococcum* cv. DV92, accession DQ889701) must be PPO sequences from two different genomes of hexaploid wheat (Fig. 3). These results are consistent with previous assertions that PPO genes are located in homoeologous chromosomes of the A, B, and D genomes of wheat (Jimenez and Dubcovsky 1999; Demeke and Morris 2002; Watanabe et al. 2004).

In addition to the three paralogous loci of *T. monococcum* cv. DV92, phylogenetic inferences revealed a fourth group suggesting an additional (putative) paralogous PPO locus (cluster II). An intriguing characteristic of this group was the occurrence of intron absence–presence polymorphisms among its members. It may be possible that intronless sequences arose by recombination with a processed pseudo-gene or a reverse-transcribed cDNA copy of a processed mRNA (Frugoli et al. 1998). Yet, whether sequences within cluster II diverged by speciation (orthologous) or by gene duplication (paralogous) needs to be investigated further.

Sequence divergence together with structural polymorphism further suggested three minor groups within cluster IV. Subgroups α and β are probably paralogs as divergence between *T. aestivum* cv. Klasic (DQ889708) and *T. monococcum* cv. DV92 (DQ889710) is similar to that between the paralogs *T. monococcum* cv. DV92 (DQ889701) and *T. monococcum* cv. DV92 (DQ889695). Intron size variation as well as MITEs insertion (or excision) events within Cluster IV may also support this hypothesis. It is tempting to speculate that sequence divergence between this and the remaining gene clusters also reflects functional divergence.

A previous study reported three distinct wheat PPO sequences expressed in developing kernels (Jukanti et al. 2004), but no direct evidence of orthologous–paralogous relationships was documented. The present results suggest that the seed specific PPO genes of Jukanti et al. (2004) (GenBank accessions AY596268, AY596270, AY596269) stand for at least two putative paralogous loci. Genes AY596268 and AY596270 fell into cluster IV, while AY596269 was closely related (99% similarity, 370 bp) to sequences of cluster II (data not shown). Consistent with previous data (Sun et al. 2005), the sequence similarity between AY596268 and *T. monococcum* cv. DV92 (DQ889710)

(Fig. 3) suggest a common ancestral genome sequence. Using Chinese Spring nullisomic–tetrasomic lines, Sun et al. (2005) mapped the PPO gene AY596268 to the long arm of chromosome 2A.

The present study documents and describes novel genomic PPO sequences from diploid, tetraploid, and hexaploid wheat species, and provides a molecular genetic and evolutionary framework for these genes.

References

- Anderson JA, Morris CF (2003) Purification and analysis of wheat grain polyphenol oxidase (PPO) protein. *Cereal Chem* 80:135–143
- Baik BK, Czuchajowska Z, Pomeranz Y (1994) Comparison of polyphenol oxidase activities in wheats and flours from Australian and US cultivars. *J Cereal Sci* 19:291–296
- Brown JW, Smith P, Simpson CG (1996) Arabidopsis consensus intron sequences. *Plant Mol Biol* 32:531–535
- Bureau TE, Wessler SR (1994a) Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6:907–916
- Bureau TE, Wessler SR (1994b) Mobile inverted-repeat elements of the tourist family are associated with the genes of many cereal grasses. *Proc Natl Acad Sci USA* 91:1411–1415
- Colson P, Tamalet C, Raoult D (2006) SVARAP and aSVARAP: simple tools for quantitative analysis of nucleotide and amino acid variability and primer selection for clinical microbiology. *BMC Microbiol* 6:21
- Demeke T, Morris CF (2002) Molecular characterization of wheat polyphenol oxidase (PPO). *Theor Appl Genet* 104:813–818
- Demeke T, Morris CF, Campbell KG, King GE, Anderson JV, Chang HG (2001) Wheat polyphenol oxidase: distribution and genetic mapping in three inbred line populations. *Crop Sci* 41:1750–1757
- Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3:329–341
- Frugoli JA, McPeck MA, Thomas TL, McClung CR (1998) Intron loss and gain during evolution of the catalase gene family in Angiosperms. *Genetics* 149:355–365
- Gooding PS, Bird C, Robinson SP (2001) Molecular cloning and characterization of banana fruit polyphenol oxidase. *Planta* 213:748–757
- Jimenez M, Dubcovsky J (1999) Chromosome location of genes affecting polyphenol oxidase activity in seeds of common and durum wheat. *Plant Breed* 118:395–398
- Jukanti AK, Bruckner PL, Fischer AM (2004) Evaluation of wheat polyphenol oxidase genes. *Cereal Chem* 81:481–485
- Jukanti AK, Bruckner PL, Fischer AM (2006) Molecular and biochemical characterization of polyphenol oxidases in developing kernels and senescing leaves of wheat (*Triticum aestivum*). *Funct Plant Biol* 33:685–696
- Khamis A, Colson P, Raoult D, Scola BL (2003) Usefulness of *rpoB* gene sequencing for identification of *Afiopia* and *Bosea* species, including a strategy for choosing discriminative partial sequences. *Appl Environ Microbiol* 69:6740–6749
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Mayer AM, Harel E (1979) Polyphenol oxidases in plants. *Phytochemistry* 18:193–215
- Morris CF, Jeffers HC, Engle DA (2000) Effect of processing, formula and measurement variables on alkaline noodle color—toward an optimized laboratory system. *Cereal Chem* 77:77–85
- Morimoto R, Kosugi T, Nakamura C, Takumi S (2005) Intragenic diversity and functional conservation of the three homoeologous loci of the KN1-type homeobox gene *Wknox1* in common wheat. *Plant Mol Biol* 57:907–924
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular plant DNA. *Nucleic Acids Res* 8:4321–4325
- Newman SM, Eannetta NT, Yu H, Prince JP, de Vicente MC, Tanksley SD, Steffens JC (1993) Organization of the tomato polyphenol oxidase gene family. *Plant Mol Biol* 21:1035–1051
- Petersen G, Seberg O (2000) Phylogenetic evidence for excision of stowaway miniature inverted-repeat transposable elements in triticeae (Poaceae). *Mol Biol Evol* 17:1589–1596
- Sherman TD, Gardeur TL, Lax AR (1995) Implications of the phylogenetic distribution of polyphenol oxidase in plants. In: Lee CY, Whitaker JR (eds) *Enzymatic browning and its prevention*. American Chemical Society, Washington, DC, pp 103–119
- Simeone R, Pasqualone A, Clodoveo ML, Blanco A (2002) Genetic mapping of polyphenol oxidase in tetraploid wheat. *Cell Mol Biol Lett* 7:763–769
- Sparks ME, Brendel V (2005) Incorporation of splice site probability models for non-canonical introns improves gene structure prediction in plants. *Bioinformatics* 21:iii20–iii30
- Sullivan ML, Hatfield RD, Thoma SL, Samac DA (2004) Cloning and characterization of red clover polyphenol oxidase cDNAs and expression of active protein in *Escherichia coli* and transgenic alfalfa. *Plant Physiol* 136:3234–3244
- Sun DJ, He ZH, Xia XC, Zhang LP, Morris CF, Appels R, Ma WJ, Wang H (2005) A novel STS marker for polyphenol oxidase activity in bread wheat. *Mol Breed* 16:209–218
- Swofford DL (2002) PAUP: phylogenetic analysis using parsimony, version 4.0b 10. Sinauer Associates, Sunderland, MA
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Thygesen PW, Dry IB, Robinson SP (1995) Polyphenol oxidase in potato. *Plant Physiol* 109:525–531
- Watanabe N, Takeuchi A, Nakayama A (2004) Inheritance and chromosomal location of the homoeologous genes affecting phenol colour reaction of kernels in durum wheat. *Euphytica* 139:87–93
- Wicker T, Matthews DE, Keller B (2002) TREP: a database for triticeae repetitive elements. *Trends Plant Sci* 7:561–562
- Zhou Y, O'Hare TJ, Jobin-Décor m, Underhill SJR, Wills RBH, Graham MW (2003) Transcriptional regulation of a pineapple polyphenol oxidase gene and its relationship to blackheart. *Plant Biotechnol J* 1:463–478
- Zhu W, Schlueter SD, Brendel V (2003) Refined annotation of the *Arabidopsis* genome by complete expressed sequence tag mapping. *Plant Physiol* 132:469–484
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415