Recurrent Deletions of Puroindoline Genes at the Grain Hardness Locus in Four Independent Lineages of Polyploid Wheat

Wanlong Li*, Li Huang, Bikram S. Gill

Wheat Genetic and Genomic Resources Center
Department of Plant Pathology
4024 Throckmorton Hall
Kansas State University
Manhattan, KS 66506-5502

*Corresponding author.
E-MAIL wli@ksu.edu; FAX (785) 532-5692

Running title: Ha deletions during polyploid wheat speciation

Key words: recurrent deletion, polyploid speciation, Ha locus, wheat
ABSTRACT

Polyploidy is known to induce numerous genetic and epigenetic changes but little is known about their physiological bases. In wheat, grain texture is mainly determined by the hardness (Ha) locus consisting of genes Pina and Pinb. These genes are conserved in diploid progenitors but were deleted from the A and B genomes of tetraploid Triticum turgidum (AB). We now report the recurrent deletions of Pina-Pinb in other lineages of polyploid wheat. We analyzed the Ha haplotype structure in 90 diploid and 300 polyploid accessions of Triticum and Aegilops species. Pin genes were conserved in all diploid species and deletion haplotypes were detected in all polyploid Triticum and most of the polyploid Aegilops species. Two Pina-Pinb deletion haplotypes were found in hexaploid T. aestivum (ABD). Pina and Pinb were eliminated from the G genome, but maintained in the A genome of tetraploid T. timopheevii (AG). Subsequently, Pina and Pinb were deleted from the A genome but retained in the Am genome of hexaploid T. zhukovskyi (AmAG). Comparison of deletion breakpoints demonstrated that the Pina-Pinb deletion occurred independently and recurrently in the four polyploid wheat species. The implications of Pina-Pinb deletions for polyploid-driven evolution of gene and genome and its possible physiological significance are discussed.
Introduction

For over 100 years, wheat grain has been classified into hard and soft types. Grain hardness or texture is mainly determined by the Hardness (Ha) locus. This classification forms the fundamental basis for differentiating wheat grain worldwide (reviewed in Morris 2002). Wheat speciation has been molded by polyploidy. Diploid (Triticum monococcum L.), tetraploid (T. turgidum L.) and hexaploid (T. aestivum L.) wheat species have been known since the 1920s (reviewed in Gill and Friebe 2002). Of the 600 million metric tons of wheat produced in the world in 2005 (http://faostat.fao.org), over 90% comes from hexaploid wheat, also called common or bread wheat, and the remaining from tetraploid, also called macaroni or durum wheat (http://www.fas.usda.gov/pecad/highlights/2005/10/durum_27oct2005/). Bread wheat grain is either soft and used for pastries or hard and used for bread and noodles. Durum wheat grain is classified as extra hard and is used for pasta. Diploid wheat grain is soft and, although it was the first domesticated wheat, is now grown as a specialty crop in isolated areas. Recently, See et al. (2004) produced a super-soft hexaploid wheat genotype by introgressing softness genes from diploid species.

Because of the pivotal importance of grain texture in determining end use quality, this trait has been intensively studied by geneticists, cereal chemists and, more recently, by molecular biologists. In the 1970s, Mattern et al. (1973) mapped a gene with a major effect on grain texture on chromosome 5D. Later, Law et al. (1978) further localized the gene to the short arm of chromosome 5D and designated the trait as hardness with alleles Ha for soft and ha for hard. Greenwell and Schofield (1986) found that a 15-kDa protein called friabilin from water-washed starch from grain was associated with the hardness locus. Abundant friabilin was found in soft wheat with Ha alleles, small amounts in hard wheat with ha allele and none in extra hard durum wheat grains. Blochet et al. (1991, 1993) isolated and sequenced lipid-binding proteins and called them puroindolines because of the presence of a ‘tryptophan domain’. Gautier et al. (1994) isolated cDNA clones corresponding to genes puroindoline a (Pina) and b (Pinb). A large body of research has shown that PINA, PINB and the grain softness protein (GSP) constitute a
Li et al. Ha deletions during polyploid wheat speciation

major fraction of the friabilin. Functional copies of both Pin genes are required for soft grain texture in wheat (reviewed in Morris 2002).

Cloning of the puroindoline genes stimulated genomics research on the Ha locus. Gautier et al. (2000), using a PCR approach, showed that Pina and Pinb are highly conserved in diploid wheats and the Triticeae and closely related cereals such as rye, barley and oats but were absent in sorghum, maize and rice. They were reported absent in tetraploid wheat species T. turgidum and T. timopheevii and reintroduced into hexaploid wheat as it arose from the hybridization of tetraploid T. turgidum (lacking Pin genes) and Aegilops tauschii Coss. (Pin genes present). Null mutations of Pin genes in hexaploid wheat cultivars of more recent origin have been reported (reviewed in Morris 2002).

Sequencing the Ha loci of T. monococcum (Chantret et al. 2004) and Ae. tauschii (Chantret et al. 2005) showed that Gsp (Gene2), Pina (Gene4) and Pinb (Gene6) are located in an interval of ~70 kb. A hypothetical gene (Gene3) was found between Gsp and Pina and a function-unknown gene (Gene5) between Pina and Pinb. In the D genome of Ae. tauschii and T. aestivum, a 5’ untranslated region (UTR) and a 5’ coding sequence of Pinb were duplicated downstream of the functional Pinb. Upstream of Gsp, BGGP (Gene1), coding for a β-1,3-galactosyl-O-glycosyl-glycoprotein, delimits the 5’ boundary of Ha. Downstream of Pinb, a cluster of ATPase genes (Gene7) coding for AAA-type ATPase and an unknown gene (Gene8) delimit the 3’ boundary of Ha. Compared to Ha-D, deletions of the genomic block containing Gene3, Pina, Gene5 and Pinb occurred at Ha-A and Ha-B of T. turgidum and T. aestivum (Chantret et al. 2005).

During routine mapping of the tetraploid T. timopheevii genome, we detected one copy of the Pin genes in its genome. This was a surprising result in view of the previous report of Gautier et al. (2000) and could be due to polymorphism for this locus in this species. A large and more comprehensive polymorphism survey of Pin genes in polyploid wheat species was initiated. To more fully characterize the nature of deletions for this region in the polyploid wheats, we undertook sequencing of the Ha region of Ae. speltoides Tausch., S genome of which has highest affinity among the diploid species to the B and G genomes of polyploid wheat. Because two of the three genomes of polyploid wheat were donated by Aegilops species, we also analyzed haplotype structure at the Ha locus of a small number of diploid and polyploid Aegilops species. These results and their
implications for polyploidy-driven mechanisms of gene and genome evolution and speciation are reported and their possible physiological significance is discussed.

RESULTS

Phylogeny of polyploid wheat and genetic nomenclature of loci

To interpret the Ha deletion haplotype survey results, it is important to briefly introduce the current understanding of the phylogeny of polyploid wheat. Two lineages of tetraploid wheat, emmer (T. turgidum L., 2n=4x=28, AB) and Timopheevi (T. timopheevii (Zhuk.) Zhuk., 2n=4x=28, AG), originated less than 0.5 million years ago (MYA) (Huang et al. 2002) from two separate hybridization events between Ae. speltoides (2n=2x=14, S) as the female parent and T. urartu Tumanian ex Gandilyan (2n=2x=14, A) as the male parent (Tsunewaki 1993, Dvorak 1998, Kilian et al. 2007). More recent hybridizations with two additional diploid species gave rise to hexaploid wheat lineages. Common wheat (T. aestivum, ABD) originated ~8,000 years ago (Nesbitt and Samuel 1996) from hybridization between T. turgidum and Ae. tauschii in cultivated fields and does not exist in the wild (Kihara 1944, McFadden and Sears 1946). T. zhukovskyi Menabde & Ericzjan (2n=6x=42, AA^mG) arose in cultivation from hybridization of T. timopheevii subsp. timopheevii with T. monococcum L. subsp. monococcum (2n=2x=14, A^m) (Upadhya and Swaminathan 1963, Dvorak et al. 1993). The A, B and D diploid donors of polyploid wheats diverged from a common ancestor ~3 MYA (Huang et al. 2002). In polyploid wheat, the Ha-A, Ha-B and Ha-D genetic nomenclature conveys both locus and genomic origin of Ha homoeoloci. Results summarizing the haplotype structure at the Ha locus in relation to the phylogeny of polyploid wheat are shown in Fig. 1.

Sequence analysis of the Ha-S genomic region of Ae. speltoides

Of the three diploid ancestors of polyploid wheat, BAC sequences of Ha genomic region of A- and D-genome ancestors were reported previously (Chantret et al. 2004 and 2005). We undertook the sequence analysis of a BAC from the S genome of Ae. speltoides. We
screened a BAC library of *Ae. speltoides* (Akhunov et al. 2005) with *Gsp* and *Pina* probes, each of which identified three BACs. *Gsp*-containing BACs did not overlap with *Pina*-containing BACs. We estimate that the *Ha-S* genomic region in *Ae. speltoides* is three times the size of *Ha-A* in *T. monococcum* and of *Ha-D* in *Ae. tauschii* and five times the size of *Ha-D* of *T. aestivum*. A *Pina* BAC, 197O23, was shotgun-sequenced at 8x coverage and assembled into 13 contigs after prefinishing, totaling 212,510 bp. Four non-transposable element (TE) protein-coding genes, *Pina*, *Pinb* and two *ATPase*, were found in this BAC, are located in the contig at the 3’ end in the same orientation and span 28,848 bp (Fig. 2). *Gene5*, previously reported to be present in the collinear region between *Pina* and *Pinb* of *T. monococcum* and *Ae. tauschii*, was not found in *Ae. speltoides*. Based on sequence homology and colinearity, the two *ATPase* genes in BAC 197O23 correspond to *ATPase-4* and *ATPase-5* at the *Ha-D* locus and are orthologous to two truncated *ATPase* genes upstream of *Gene8* at the *Ha-B* locus (Chantret et al. 2005). The rest of BAC 197O23 is gene-free and mainly occupied by TEs and tandem repeats, a typical feature of large genomes, where genes are clustered into islands and separated by nested TEs (Wicker et al. 2001). *Gene8*, *ATPase-1*, *ATPase-2* and *ATPase-3* were located in a separate BAC.

**Survey of the haplotype structure at the Ha locus in Aegilops and Triticum**

We determined the haplotype structure at the *Ha* locus by Southern analysis of tester DNA digested with restriction enzymes *EcoRI*, *HindIII* or *BamHI* using *Pina* and *Pinb* gene probes. We estimated the copy number of *Pina* or *Pinb* genes in tester species by counting the number of fragments detected by Southern hybridization. The data was tabulated to determine if the haplotype structure was conserved or there were null haplotypes for either one or both the *Pin* genes at the *Ha* locus (Table 1 and Supplementary Table 1s). Null haplotypes were further characterized according to the size of the deletion either by Southern analysis using additional gene probes that mark the *Ha* locus (see Fig. 1.) or by sequencing as described below.
Diploid species: We randomly selected at least two accessions from each of the 12 diploid species of *Aegilops* and *Triticum* (a total of 90 accessions, Table 1 and Supplementary Tables 1s) for the haplotype survey. In all cases, Southern hybridization detected a single band or rarely multiple bands for *Pina* and *Pinb* gene probes, indicating that haplotype structure at the *Ha* locus is conserved in the diploid species. A single-copy of *Pina* and *Pinb* was detected in A- and D-genome donor species of polyploid wheat (Table 1). Five *Aegilops* species share the S genome and all except *Ae. speltoides* are self-pollinated. All self-pollinated S-genome species had one copy of *Pina* and *Pinb*. Most accessions of *Ae. speltoides* also carry one copy of the *Pin* genes and the observed multiple Southern hybridization fragments in some accessions (Supplementary Table 1s) may be due to either heterozygosity, because it is a cross-pollinated species, or rarely from the presence of intragenic restriction sites or gene duplication. All the C-, M-, U- and N-genome species also had one copy of the *Pin* genes except for one accession of *Ae. comosa* where Southern analysis indicated multiple gene copies.

Tetraploid species: The above-mentioned diploid species have contributed genomes to tetraploid *Triticum* and *Aegilops* species, and two copies each of *Pina* and *Pinb* genes are expected in the genomes of these tetraploid species (Table 1). The tetraploid wheat species *T. turgidum* (AB) and *T. timopheevii* (AG) form the A-genome cluster. We screened 92 accessions of *T. turgidum* including eight subspecies representing the range of wild and domesticated forms. All showed the null haplotype for the puroindoline genes (Table 1 and Supplementary Table 1s, Fig. 3) confirming that *Pina* and *Pinb* have been deleted in this species (Gautier et al. 2000, Dvorak et al. 2004).

We screened 65 accessions of *T. timopheevii* including two subspecies representing the range of wild and domesticated forms (Supplementary Table 1s). All carried only one copy of the *Pina* and *Pinb* genes, indicating null haplotype at the *Ha* locus for one of its genomes (Supplementary Table 1s). *Gene5*, which lies between *Pina* and *Pinb*, was, as expected, present in one copy. The *Gsp* probe detected two copies indicating that one of the breakpoints that produced the null haplotype is located between *Gsp* and *Pina* (Fig. 3).
Nine tetraploid *Aegilops* species are grouped into the U- and D-genome clusters. In the U-genome cluster, one accession each was analyzed for *Ae. columnaris* Zhuk. (U°M°) and *Ae. peregrina* (Hack. in J. Fraser) Marie & Weiller (US); two accessions for *Ae. biuncialis* Vis. (UM), and nine accessions for *Ae. kotschyi* Boiss. (US); all showed conserved haplotype structure for the *Pin* genes in both of their genomes (Table 1). The other three U-genome cluster species, *Ae. geniculata* Roth (U°M°), *Ae. neglecta* Req. ex Bertol. (UM) and *Ae. triuncialis* L. (U'C), where more than one accession was analyzed, were polymorphic in *Pina* copy number. One or more accessions in each species had a conserved haplotype structure for the *Pin* genes, and the other accessions for the same species showed unique haplotypes in one of their genomes where *Pina* gene was deleted but *Pinb* gene was retained. Therefore, it is possible that if a larger survey of the U-genome cluster species is undertaken, all may turn out to be polymorphic for *Ha* haplotype structure.

The D-genome cluster species *Ae. crassa* Boiss. (DX), *Ae. cylindrica* Host (D'C°) and *Ae. ventricosa* Tausch (D'N°) showed a conserved haplotype structure for the *Pin* genes except one of the 12 accessions of *Ae. crassa* was null for *Pina* in one of its genomes.

**Hexaploid species:** Some of the tetraploid species of *Triticum* and *Aegilops* mentioned above have undergone a second round of hybridization with various diploid species to form hexaploid species. The A-genome cluster species *T. aestivum* (ABD) is expected to have one copy each of the *Pina* and *Pinb* genes contributed by *Ae. tauschii* (D). We surveyed 40 accessions of *T. aestivum* including all six subspecies and 18 historical hard wheat cultivars of *T. aestivum* subsp. *aestivum*. Of the 40 *T. aestivum* accessions, 37 showed a conserved haplotype structure for *Pina* and *Pinb* in the D genome (*Ha-D*) and, as expected, a null haplotype for the A and B genomes. Of the 18 hard wheat cultivars that do not express the PINA protein (Morris *et al.* 2001), Red Egyptian, Sea Island and Komar showed the null haplotype at *Ha-D*. Further probing with *BGGP* and *Gsp* detected three copies in Sea Island and Komar but only two in Red Egyptian (Fig. 4). The RFLP analysis of the Chinese Spring/Red Egyptian chromosome substitution lines 5A, 5B and 5D confirmed that *BGGP*, *Gsp*, *Pina* and *Pinb* were deleted from the *Ha-D* locus on
chromosome 5D of Red Egyptian. Therefore, at least two independent deletion events occurred at the Ha-D locus in common wheat, one with the distal breakpoints between Gsp and Pina in Sea Island and Komar and another haplotype with distal breakpoint beyond BGGP in Red Egyptian.

We surveyed three accessions of T. zhukovskyi (AA<sup>m</sup>G), the second A-genome cluster hexaploid species. It is expected to have two copies of Pin genes, one from T. timopheevii and the second from T. monococcum. However, only one copy of the Pin genes was detected, indicating the presence of a second null haplotype at the Ha locus in one of its genomes (Fig. 4). Because T. zhukovskyi is autoallohexaploid, the loss of a Ha locus could be due to either recombination between A and A<sup>m</sup> genome or a deletion event. This question can be resolved based on BGGP and Gsp hybridization patterns: 1) if each detects three bands, Pina and Pinb were deleted from one genome; 2) if two bands with similar intensity are observed, all BGGP, Gsp, Pina and Pinb were deleted from the A or A<sup>m</sup> genome; and 3) if the two bands differ significantly in intensity, A-A<sup>m</sup> recombination, instead of deletion, occurred. Our results support the second scenario (Fig. 4), i.e., BGGP, Gsp, Pina and Pinb were deleted from the A or A<sup>m</sup> genome, similar to the Ha-D haplotype in Red Egyptian.

We surveyed three D-genome cluster and one U-genome cluster hexaploid Aegilops species and all are expected to have two to three copies of Pina and three copies of Pinb, depending upon the genotype of the tetraploid parent (see above). Ae. crassa (DDX) and Ae vavilovii (Zhuk.) Chennav. (DXS<sup>+</sup>) were polymorphic; three accessions of Ae. crassa and two accessions of Ae. vavilovii had two copies of Pina and three copies of Pinb, and the rest had two copies of both Pina and Pinb. Ae. juvenalis (Thell.) Eig (DXU) had three copies of Pina and two copies of Pinb in their genomes.

The U-genome cluster hexaploid Ae. neglecta (U<sup>M</sup>M<sup>N</sup>N<sup>B</sup>) was polymorphic; six accessions had two copies of Pina and three of Pinb, four accessions had one copy of the Pina and three copies of the Pinb and the remaining three accessions had one copy of Pina and two copies of Pinb. Because tetraploid Ae. neglecta (UM) was polymorphic for Pina (one or two copies), the data suggest a recurrent deletion of Pina at tetraploid and hexaploid levels in this species. Compared to Pina, Pinb deletion was only detected at the hexaploid level in the polyploid Aegilops species.
Sequence analyses of unique haplotypes

A sequence analysis of deletion haplotypes detected in polyploid wheat species was used to further characterize and allocate their genomic origin. These results are summarized in Fig. 1.

The Ha-A haplotype of T. timopheevii: Detection of the conserved Ha haplotype in the Timopheevi lineage of polyploid wheat triggered an immediate effort to determine its genomic origin. We designed A- and G-genome specific primers based on Pina and Pinb sequences from diploid progenitors T. urartu and Ae. speltoides, respectively. The A-genome specific primers amplified strong fragments approximately 650 bp for Pina and 780 bp for Pinb. The G-genome specific primers produced very weak amplicons from T. timopheevii. Cloning and sequencing of the PCR products from six T. timopheevii accessions indicated that the A-genome amplicons showed highest homology to puroindoline genes from T. urartu at the DNA and protein sequence levels, but the G-genome amplicons showed no homology to the Pin genes but weak similarities to various TEs. This clearly indicated that the A-genome haplotype (Pina-A and Pinb-A) was maintained and the G-genome copy was deleted in T. timopheevii.

The promoters and coding regions of Pina and Pinb are highly conserved among the timopheevii accessions and between timopheevii and urartu except for a 1-bp insertion in the promoter region of Pina in one accession and an Ato-C transversion at position 181 of Pinb in another (Supplementary Fig. 1s and 2s). In the 3’UTR of Pinb, an 88-bp fragment, spanning the stop codon and polyadenylation signal, was found in triplicate in Pinb-A of T. timopheevii compared to its ancestor T. urartu. The repeat members are identical except for a single nucleotide polymorphism (SNP) (Supplementary Fig. 2s). A PCR assay showed that the 88-bp triple repeat is fixed at the species level (Fig. 5). A 7-bp sequence (CAACATG) was found at the beginning of each repeat member and immediately after the triple repeat in T. timopheevii and flanking the 88-bp sequence in T. urartu, suggesting that it originated by replication slippage after polyploidization (Supplementary Fig. 2s). The alignments of nucleotide sequences of the Pina and Pinb
genes and amino acid sequences of the PINA and PINB proteins between *T. timopheevii* and *T. urartu* are shown in Supplementary Figs. 1s, 2s, 3s and 4s.

**The Ha-G haplotype of *T. timopheevii***: To determine the fine structure of Ha-G, particularly the deletion breakpoints, we isolated Ha-G from a fosmid library constructed from accession Tim01 of *T. timopheevii*. The Gsp-containing fosmid clone, 1E05, was sequenced at 8x coverage and assembled into five contigs after prefinishing, totaling 41,262 bp. The largest contig, 23,807 bp, contains three non-TE protein-coding genes within 10,919 bp, BGGP, Gsp and ATPase (Fig. 2). In addition, a 311-bp sequence, corresponding to the 5’ portion of Gene3 was found 1,806 bp downstream of Gsp and 2,749 bp upstream of ATPase, and a partial coding sequence of another copy of the ATPase gene was found 778 bp downstream of the intact ATPase at the 3’ end of the fosmid clone. At the nucleotide sequence level, Gsp-G is highly homologous to Gsp-S of *Ae. speltoides* and Gsp-B of *T. turgidum* and *T. aestivum*. ATPase-G showed highest identity (98%) to ATPase-1 (Gene7-1, pseudogene) from chromosome 5B of *T. turgidum* and *T. aestivum*. BGGP-G has high homology to BGGP-D of *Ae. tauschii* and *T. aestivum* and to BGGP-B of *T. turgidum* and *T. aestivum*. As expected, GSP-G is encoded by the plus strand, and BGGP-G and ATPase-G are encoded by the minus strand. The Ha-G haplotype lost Gene8 along with Pina, Pinb, ATPase-4 and ATPase-5, and has a distal breakpoint between Gene3 and Pina and a proximal breakpoint between Gene8 and ATPase-1. Compared to the Ha-S locus, more than 200 kb of sequence was deleted from the Ha-G and Ha-B loci during polyploid wheat evolution (Fig. 2).

**The Ha-A<sup>m</sup> haplotype of *T. zhukovskyi***: Because *T. zhukovskyi* originated from a cross between *T. timopheevii* (Ha-A, null Ha-G) and *T. monococcum* (Ha-A<sup>m</sup>), the conserved haplotype in this hexaploid wheat is either of Ha-A or Ha-A<sup>m</sup> origin. The A<sup>m</sup> genome of *T. monococcum* is highly homologous to the A genome of *T. urartu* due to their very recent divergence. Multiple sequence alignment of the puroindoline gene sequences available in NCBI database from these two species and sequences of *T. timopheevii* from this research were used to identify species-specific SNP to unambiguously assign the genome origin of Pina and Pinb in *T. zhukovskyi*. A monococcum-specific synonymous
point mutation was identified at position 384 within the open reading frame (ORF) where three accessions of *T. monococcum* carried an A, the mutant allele, and all accessions of *T. urartu, T. timopheevii* and the remaining *T. monococcum* accessions carried a G, the wild-type allele (Supplementary Fig. 5s). Similarly, a species-specific synonymous mutation was found at position 75 in the ORF of *Pinb* where all the *T. urartu* and *T. timopheevii* accessions carry a T and all the accessions of *T. monococcum* carry a C (Supplementary Fig. 6s). All three accessions of *T. zhukovskyi* carry *monococcum*-specific alleles indicating that the *Pina-A* and *Pinb-A* from *T. timopheevii* were deleted and are survived by *Pina-A* and *Pinb-A* from *T. monococcum* (Supplementary Figs. 5s and 6s). Furthermore, the 88-bp triple repeat was not detected in *Pinb* of *T. zhukovskyi* by either sequencing or PCR assay (Fig. 5) providing additional evidence that *Pinb-A* was deleted in this hexaploid wheat species.

**The *Ha* loci of synthetic tetraploids**

The fixation of *Ha* deletion patterns in tetraploid wheat suggested that puroindoline genes may have been eliminated immediately upon polyploidization similar to the low-copy sequences demonstrated by Feldman *et al.* (1997) and Liu *et al.* (1998). We tested this hypothesis on tetraploid AASS- and A<sup>m</sup>A<sup>m</sup>S<sup>sh</sup>S<sup>sh</sup>-genome amphiploids. The AASS amphiploid was produced in our laboratory by crossing diploid species *T. urartu* (A) with *Ae. speltoides* (S) followed by chromosome doubling. The A<sup>m</sup>A<sup>m</sup>S<sup>sh</sup>S<sup>sh</sup> was produced from *T. monococcum* (A<sup>m</sup>) and *Ae. sharonensis* (S<sup>sh</sup>) by Ozkan *et al.* (2001). We used genome-specific primers for both *Pina* and *Pinb* to produce amplicons for sequencing from the AASS and A<sup>m</sup>A<sup>m</sup>S<sup>sh</sup>S<sup>sh</sup> amphiploids and their diploid parents. The *Pina* and *Pinb* haplotypes were conserved in both amphiploids (Fig. 6). The *Pina* and *Pinb* sequences of the A and A<sup>m</sup> genomes between the amphiploids and their parents were identical (Supplementary Figs. 7s, 8s, 11s and 12s), but up to 1% variation was found in the *Pina* and *Pinb* sequences of the S and S<sup>sh</sup> genomes of the amphiploids compared to their *Aegilops* parents (Supplementary Tables 9s, 10s, 13s and 14s). More striking was a 21-bp deletion detected in *Pinb-S<sup>sh</sup>* of the amphiploid A<sup>m</sup>A<sup>m</sup>S<sup>sh</sup>S<sup>sh</sup> compared to its *Ae. sharonensis* parent. Six clones from that ligation were sequenced, and the wild-type allele was not found. *Pinb* sequences of the S-genome species including *Ae. bicornis* (Forssk.)
Jaub. & Spach (S^b), Ae. longissima Schweinf. & Muschl. (S^1), Ae. searsii Feldman & Kislev ex Hammer (S^S^S), Ae. sharonensis Eig (S^b) and Ae. speltoides (S) contain two 11-bp direct repeats (GAAGTTGGCGG) separated by a 10-bp spacer (CTGGTACAAT). This 21-bp deletion was obviously caused by unequal crossing over between the 11-bp direct repeats (Supplementary Fig. 10s) and led to a loss of seven amino acids (WYNEVGG) in the PINB protein. The A^m A^m S^sh S^sh amphiploid used is from the S2 generation, the unequal crossover either occurred during female meiosis of the Ae. sharonensis parent TMB02 or happened and was rapidly fixed after polyploidization.

**DISCUSSION**

The most remarkable observation on the structure and evolution of the Ha locus in wheat and the Triticeae is the absolute conservation of the locus in diploid species reported here and in previous papers (Gautier et al. 2000, Lillemo et al. 2002, Massa et al. 2004, Chen et al., 2005, Simeone et al. 2006) and recurrent and independent deletions in the polyploid Triticum and Aegilops species. To date, more than 200 accessions from the two diploid Triticum and ten diploid Aegilops species have been analyzed, and not a single case of deletion polymorphism at the Ha locus has been reported. Especially, no deletion polymorphisms have been detected in a diverse sample of more than 130 accessions of the A-, B- and D-genome donor species of polyploid wheat. This is in contrast to frequent deletion haplotype polymorphisms for a defense-gene cluster in D-genome diploid, Ae. tauschii (Brooks et al. 2006). Against this high rate of deletion polymorphism in polyploid species, not a single case of insertion-deletion polymorphism was documented in a sample of Pina and Pinb sequences from 50 accessions of diploid Ae. tauschii representing its geographical diversity (Massa et al. 2004). All polyploid wheats and most polyploid Aegilops species harbored deletion haplotypes of independent origin at the Ha locus. So how does a gene that is essential in a diploid suddenly become deleterious and must be deleted in a polyploid? To begin to answer this question, some discussion about the nature of the puroindoline genes, their function, the nature of gene action in polyploids and the mechanisms of polyploid genome evolution and speciation that promote expression and evolution of novel traits is needed.
Amino acid sequence analysis has shown that numerous storage proteins, including low-molecular-weight glutenin, alpha-/beta gliadins, lipid transfer proteins, chymotrypsin inhibitor WCI, alpha-amylase/trypsin inhibitor, GSP, PINA and PINB, belong to the alpha-amylase inhibitors (AAIs) and seed storage (SS) protein subfamily, because they have an AAI-SS domain. AAIs play an important role in the natural defense of plants against insects and pathogens mainly by inhibiting alpha-amylases and proteases. Puroindolines have bactericidal (Jing 2003) and fungicidal activities (Krishnamurthy et al. 2001). PINA and PINB proteins directly bind to the surface of starch granules in the endosperm cells and form a friabilin complex. Using isogenic lines, Swan et al. (2006) showed that puroindolines seem to protect starch from microbial digestion, and the increased expression of PIN proteins decreased the starch digestibility of wheat in the rumen by up to 30%. Alpha-amylase is an important enzyme in starch metabolism and induced in the aleurone by gibberellic acid from the embryo during germination. Conceivably, wheat starch can be protected from alpha-amylase digestion by AAI activity of the PIN proteins. Therefore, because of the important role of PIN proteins in plant defense and seed physiology, Pina and Pinb genes may be under strong selection pressure and are maintained in all the diploid species.

One of the consequences of polyploidy is doubling and tripling of gene copy number and, thus, the amount of proteins may be doubled or tripled for some of these genes. This dosage response has been demonstrated for the Pin genes and a super soft hexaploid wheat genotype has been created (See et al. 2004). Because wheat starch is protected from alpha-amylase digestion by AAI activity of the PIN proteins, we hypothesize that the sudden dosage-driven increase in expression levels of puroindoline genes in polyploid would impede the embryos from obtaining nutrition from the endosperm. The situation may be more severe when polyploid plants are under abiotic stress, such as heat and drought during grain filling, which adversely affects endosperm development. Point mutations in Pinb can liberate the PIN proteins from binding to the starch granule surface and cause significant difference in grain texture (Giroux and Morris 1997, 1998, Morris et al. 2001), however, the PIN proteins with the AAI activity remain in the endosperm cells. Therefore, deletion of Pina-Pinb genes provides the most efficient mechanism to reduce AAI activity and is the least detrimental, because they are structural genes. The
Li et al. *Ha deletions during polyploid wheat speciation*

Reduction in PIN proteins in turn enhances the exposure of starch granules to alpha-amylases during germination to provide sufficient nutrition for new polyploid seedling to compete in stand establishment with surrounding diploid populations. Thus the deleterious action of a single gene may determine the fate of a new polyploid species and, therefore, we consider it a bottleneck speciation event.

To test our hypothesis of deleterious effect of high dose of puroindoline genes on seed physiology, a preliminary germination experiment was performed using seeds from hexaploid Chinese Spring wheat plants with two (background control), four and six doses of the *Pina-Pinb* genes (See et al. 2004). The seeds from each genotype were placed in Petri dishes on wet filter paper at room temperature. After 24 hours, the check cultivar Chinese Spring (with two doses) sprouted and radicle and coleoptile were visible. The genotype with four doses showed sprouting activity but the radicle and coleoptile were invisible. The genotype with six doses did not sprout until 48 hours after imbibition and radicle and coleoptile were visible only after 96 hours of imbibition. The negative correlation between the *Pina-Pinb* dosages and the rate of germination and growth was consistently observed in the subsequent days. Although this experiment appears to demonstrate the deleterious effects of higher doses of *Pin* genes on seed germination and stand establishment, more rigorous experiments need to be conducted including assaying AAI activity of PIN proteins and evaluating germination, growth and vigor of seedlings of the isogenic lines with different doses of *Pin* genes under field conditions. But what are the chances of occurrence and fixation of such a rare deletion mutation in a small founder polyploid species population and what genes might compensate for this missing function?

Because at least one copy of the *Pin* genes is maintained in the vast majority of polyploid species, the loss and fixation of null haplotypes copies of both *Pin* genes in tetraploid *T. turgidum* requires explanation. As mentioned earlier, a number of proteins have AAI activity and higher doses of their expression may partially compensate for the loss of function of the *Pin* genes. Gsp was not involved in most deletion haplotype polymorphisms and may have some compensatory functional role in defending against the microbe attacks. Because extensive resetting of gene expression patterns follows polyploidy, it is possible that other genes may have been recruited for the same role. In
fact, QTLs for hardness phenotype have been mapped in other regions of the genome besides the 5DS locus (Breseghello et al. 2005, Narasimhamoorthy et al. 2006).

Accumulating evidence suggests that polyploidization is accompanied by significant genome restructuring and resetting of gene expression patterns (reviewed in Chen and Ni 2006 and Chen 2007). Numerous genetic and epigenetic changes have been observed in neopolyploids (amploids), including sequence elimination (Song et al. 1995, Levey and Feldman 2004), chromosome rearrangements (Pontes et al. 2004), changes in methylation (Shaked et al. 2001) and gene expression patterns (Adams et al. 2003, He et al. 2003), reactivation of TEs (Kashkush et al. 2003) and microRNA expression (Tian et al. 2005). The rate of DNA sequence divergence is much higher in polyploid wheat compared to diploid relatives (reviewed in Dubcovsky and Dvorak 2007). The great abundance of repetitive sequences, especially retroelements, in the wheat genome (Li et al. 2004, Devos et al. 2005) seem to promote gene deletion/duplication events. Chantret et al. (2005) reported that the loss of a block of genes, including Pina and Pinb gene from A and B genomes of tetraploid T. turgidum, occurred independently by illegitimate recombination among retroelements bordering the Ha locus. Such events appear to be common at the Ha locus, because Chantret et al. (2005) documented two additional such events in polyploid wheat genotypes. The loss of large gene block, including Ha-G in T. timopheevii as reported here, may have been caused by a similar mechanism. We also documented replication slippage as a mechanism leading to triple repeat of an 88-bp sequence at the Ha-A locus of T. timopheevii. Unequal crossing-over led to the deletion of 21-bp sequence in Pinb-Ssh gene in A^m^A^ns^SshSsh synthetic amphiploid. Although Pina and Pinb were conserved in a number of tetraploid Aegilops species, they may not be transcribed as documented by Chen et al. (2005). Others were polymorphic for the Pina or Pinb deletion haplotype and none of the hexaploid species tolerated three doses of Pin genes (Table 1). Although there is a high rate of deletion in polyploid Triticum and Aegilops species, polyploidy per se does not cause Pin gene deletions as indicated by our data on synthetic amphiploids. The high mutation rate at the Ha locus in polyploid species, coupled with high gametic transmission of deletion haplotypes due to polyploid buffering, and the high fitness cost of higher doses of Pin gene expression on seed
physiology led to the fixation of deletion haplotypes at the Ha locus in the founder populations during polyploid wheat speciation.

The polyploidization leads to both additive and non-additive gene expression patterns (reviewed by Chen 2007). The documentation of recurrent deletions at the hardness locus suggests certain rules for the fate of loci that show dosage sensitive expression following polyploidization. If the additive gene expression has deleterious effect on the organism then it will be rapidly deleted. Conversely, if the additive gene expression has beneficial effect then it will be conserved. If the effect is neutral then the duplicate loci may undergo subfunctionalization or mutation to assume new functions (Adam et al. 2003). Comparative genomics is generating large databases of gene duplications and deletions following whole genome polyploidization. The challenge of the postgenomics era will be to determine the physiological bases of such duplication/deletion events.

METHODS

Plant materials

Plant materials are summarized in Table 1 regarding the species, ploidy levels, genome formula and number of accessions. The details of individual accessions are listed in Supplementary Table 1s with their Pina and Pinb scores. The accessions prefixed with TA are maintained by the Wheat Genetic and Genomic Resources Center (WGGRC), Kansas State University (Manhattan, KS); accessions prefixed with CItr, PI and PVP were obtained from Dr. Harold Bockelman at the USDA Small Grains Collection (Aberdeen, ID). T. timopheevii accession Tim01, the TH02/TMB02 amphiploid (A^mA^mSS), the Ae. sharonensis (TH02) and T. urartu (TMB02) were provided by Dr. Moshe Feldman, Weizmann Institute of Science (Rehovot, Israel). The amphiploid (TA3438, AASS) was derived from a cross made at the WGGRC between Ae. speltoides accession TA1785 and T. urartu accession TA831. All the lines were grown in the greenhouse. The ploidy levels of Ae. crassa and Ae. neglecta accessions were determined by Badaeva et al. (1998 and 2001) and this research. Critical accessions were identified by C-banding (Gill et al. 1991) when necessary.
Clones and primers
Wheat cDNA clones TMA9 (Pina) and TMA10 (Pinb) were provided by Dr. Marie-Françoise Gautier (Unité de Biochimie et Biologie Moléculaire des Céréales, INRA, France). EST BU100707 homologous to Gene1 was obtained from the Arizona Genomics Institute (Tucson, AZ). The BAC clone 197O3 of Ae. speltoides was provided by Dr. Jan Dvorak (University of California, Davis, CA). All PCR primers for amplification of Gsp, Pina, Gene5 and Pinb and annealing temperatures are given in Supplementary Table 2s.

RFLP analysis to detect deletions
The copy numbers of Pina and Pinb were determined based on the number of fragments detected in Southern blots. Approximately 100 mg of leaf tissue was collected from each accession, lyophilized in a 2-ml microcentrifuge tube and disrupted by shaking with metal beads. The procedures for DNA extraction, digestion, electrophoresis and the Southern hybridization generally followed those of Faris et al. (2000). For most accessions, EcoRI, HindIII or BamHI were used for digestion. The search of the coding sequences of Pina and Pinb deposited in the NCBI, did not reveal restriction sites BamHI and EcoRI, and HindIII restriction site was only found at position 24 of Pinb from the S-genome species. In some cases, more bands were observed than expected possibly due to heterozygosity, intragenic restriction site or duplication and they were ignored because the focus of the present work was on the detection of the Pin gene deletions.

Molecular cloning
The PCR products of genes Gsp, Pina, Gene5 and Pinb were separated by agarose gel electrophoresis, eluted from gel, ligated to T-easy vector (Promega, Madison, WI) and transformed into E. coli strain DH10B.

To clone the Ha locus from T. timopheevii, we constructed a fosmid library of accession Tim01. Briefly, total genomic DNA of Tim01 was sheared by 120 cycles of freezing in liquid nitrogen and thawing in a 65°C water bath, and separated by CHEF gel electrophoresis. Fragments of 30 to 50 kb were excised, eluted, end-repaired with an End-It kit (Epicentre Biotech, Madison, WI) and ligated to the CopyControl™ pCC1FOS™. The ligation was packaged with MaxPlax™ Lambda Packaging Extracts (Epicentre
Biotech, Madison, WI), diluted 100-fold and used to infect *E. coli* strain PE1300\textsuperscript{TM} following the manufacturer’s instructions. An aliquot of 70-µl infected bacteria (~70 clones) were distributed and maintained in 384-well plates. The library first was pooled by plate and screened by PCR using *Gsp*-S-specific primers. Positive plates were pooled by rows and columns. Once a positive well was identified, the culture from that well was spread onto LB agar plates containing chlorophenicol (12.5 µg/ml) and colonies were picked, arrayed in 96-well plates and screened individually by PCR. To isolate the *Ha* locus from the G-genome donor species, we hybridized the *Pina* and *Gsp* to the macroarray filters of an *Ae. speltoides* BAC library (Akhunov et al. 2005).

**Sequence analysis**

To design genome-specific primers, nucleotide sequences for *Gsp*, *Pina* and *Pinb* of diploid species of *Triticum* and *Aegilops* were retrieved from The National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) and subjected to multiple sequence alignment with ClustalW software at BCM (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) and formatted by BOXSHADE 3.21 at web site http://www.ch.embnet.org/software/BOX_form.html. For the *Pina* and *Pinb* genes cloned from *T. timopheevii*, *T. zhukovskyi*, amphiploids and their parents, eight clones from a ligation were sequenced, and the sequences were assembled with CAP3 program (Huang 1999) at web site http://pbil.univ-lyon1.fr/cap3.php, with all the parameters set default. The sequence contigs were used as queries for BLASTn searches against the NCBI nr database. A fosmid and a BAC was shotgun sequenced at 8x equivalents and assembled with the program Consed (Gordon et al. 1998). Protein-coding genes were predicted using the program FGENESH at the web site http://sun1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind with the organism set as monocot. Repeated sequences were identified by searching the TREP database (Wicker et al. 2002) by BLASTn and BLASTp at web site http://wheat.pw.usda.gov/ITMI/Repeats/blastrepeats3.html.

**Gene nomenclature**
Gene designations followed the rules of nomenclature as listed in the Wheat Gene Symbol Catalogue (McIntosh et al. 1998). The Ha locus is triplicated in common wheat and orthologous loci of A, B and D genome origin are designated as Ha-A, Ha-B and Ha-D, respectively. For hypothetical and function unknown genes at the Ha locus, the names designated by Chantret et al. (2005) are adopted to avoid any confusion.

All the DNA sequences are deposited in GenBank under the accession numbers________.

ACKNOWLEDGEMENTS

We thank Drs. Harold Bockelman and Moshe Feldman for providing seeds, Drs. Jan Dvorak, Marie-Françoise Gautier for providing cDNA clones, Dr. Bernd Friebe for C-banding identification of wheat accessions, and Jon Raupp and Duane Wilson for technical assistance. This paper is a contribution number 07-255-J from the Kansas Agricultural Experiment Station. The research was partially supported by grants from USDA and Kansas Wheat Commission.
FIGURE LEGENDS

Figure 1. A scheme for polyploid wheat phylogeny and changes in the Ha haplotype structure during their evolution. The solid arrows represent genes and their orientation, with symbols above. Species and their genome formulae are indicated underneath. Arrow size and spacing are not proportional to gene size or intergenic interval.

Figure 2. Comparison of haplotype structure at the hardness loci: Ha-D of Ae. tauschii and T. aestivum, Ha-S of Ae. speltoides, Ha-G of T. timopheevii and Ha-B of T. turgidum and T. aestivum. The solid arrows represent genes and their orientation with the gene symbols above them. Gene3 and Gene8 are hypothetical and Gene5 is coding for an unknown protein. The Ha-S haplotype spans three separate BACs as indicated by slashes. The open arrows in Ha-S and Ha-G were not sequenced they are deduced based on colinearity between the Ha-B and Ha-D loci. The lines connect the orthologs. The Ha-B and Ha-D haplotypes are after Chantret et al. (2005).

Figure 3. The autoradiogram of a Southern hybridization. Genes Gsp, Pina, Gene5 and Pinb are indicated at the left and species at the top. Gsp detected two major bands in both T. timopheevii and T. turgidum. Pina, Gene5 and Pinb detected a single fragment in T. timopheevii and none in T. turgidum.

Figure 4. The autoradiogram of Southern hybridization. Genes BBGP, Gsp, Pina and Pinb as indicated at the left and accession numbers at the top. Red Egyptian, Sea Island and Komar are hard red cultivars of common wheat (T. aestivum); accessions TA2610, PI355352 and PI355353 belong to T. zhukovskyi. BBGP and Gsp each detected two fragments in Red Egyptian and three in Sea Island and Komar, whereas Pina and Pinb did not detect any signal in all three wheat cultivars. In T. zhukovskyi, BBGP and Gsp detected two bands with similar intensity and Pina and Pinb detected a single fragment.

Figure 5. PCR assay of the 88-bp triple repeat at the 3’ end of Pinb-A. The 100-bp ladder is on either side and species are indicated at the top. Amplicons of T. urartu and T. zhukovskyi are 308 bp and those of T. timopheevii are 484 bp in length.

Figure 6. PCR assay of Pina and Pinb in amphiploids and their parents. The 100-bp ladder is on either side and the lowest band is 500 bp. The genome formula is indicated at the top and gene symbols are at the left to the picture.
REFERENCES

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100: 4649-4654

Akhunov ED, Akhunova AR, Dvorak J (2005) BAC libraries of *Triticum urartu, Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. Theor Appl Genet 111:1617-1622

Badaeva ED, Friebe B, Zoshchuk SA, Zelenin AV, Gill BS (1998) Molecular cytogenetic analysis of tetraploid and hexaploid *Aegilops crassa*. Chromosome Res 6: 629-637

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuk SA, Shostak NG, Chikida NN, Zelenin AV, Raupp WJ, Friebe B, Gill BS (2001) Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. Pl Syst Evol 246: 45-76

Blochet JE, Kabouloou A, Compoint JF and Marion D (1991) Amphiphilic proteins from wheat flour: specific extraction, structure and lipid binding properties. Gluten Proteins 1990 (Eds. W Bushuk and R Tkachuk), American Association of Cereal Chemists, St. Paul, MN. 314-325

Blochet JE, Chevalier C, Forest E, Pebay-Peyroula, Gautier MF, Joudrier P, Pezolet M, Marion D (1993) Complete amino acid sequence of puroindoline, a new basic and cystine rich protein with a unique tryptophan-rich domain, isolated from wheat endosperm by Triton X-114 phase partitioning. FEBS Lett 329: 336-340

Bresegelloa F, Finneyc PL, Gainesd C, Andrewsd L, Tanakab J, Pennere G, Sorrells ME (2005) Genetic loci related to kernel quality differences between a soft and a hard wheat cultivar. Crop Sci 45:1685-1695

Brooks SA, Huang L, Herbel MN, Gill BS, Brown-Guedira G, Fellers JP (2006) Structural variation and evolution of a defense-gene cluster in natural populations of *Aegilops tauschii*. Theor Appl Genet 112: 618-626

Chantret N, Cenci A, Sabot F, Anderson O, Dubcovsky J (2004) Sequencing of the *Triticum monococcum* hardness locus reveals good microcolinearity with rice. Mol Genet Genomics 271: 377-386
Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (\textit{Triticum} and \textit{Aegilops}). Plant Cell. 17: 1033-1045

Chen M, Wilkinson M, Tosi P, He G, Shewry P (2005) Novel puroindoline and grain softness protein alleles in Aegilops species with the C, D, S, M and U genomes. Theor Appl Genet 111: 1159-1166

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58: 377-406

Chen ZJ, Ni Z (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays 28: 240-252

Devos KM, Ma J, Pontaroli AC, Pratt LH, Bennetzen JL (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci U S A. 102: 19243-19248

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316: 1862-1866

Dvorak J (1998) Genome analysis in the \textit{Triticum-Aegilops} alliance. Proc. 9th Internl. Wheat Genet. Symp. (eds A. E. Slinkard) University Extension Press, Saskatoon, Canada 1: 8-11

Dvorak J, di Terlizzi P, Zhang H-B, Resta P (1993) The evolution of polyploid wheats: identification of A genome donor species. Genome 36:21-31

Dvorak J, Yang ZL, You FM, Luo MC (2004) Deletion polymorphism in wheat chromosome regions with contrasting recombination rates. Genetics 168:1665-1675

Faris JD, Haen KM, Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics 154: 823-835

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics 147: 1381-1387

Gautier MF Aleman ME, Guirao A, Marion D, Joudrier P (1994) \textit{Triticum aestivum} puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 25:43–57
Gautier MF, Cosson P, Guirao A, Alary R, Joudrier P (2000) Puroindoline genes are highly conserved in diploid ancestor wheats and related species but absent in tetraploid Triticum species. Plant Sci 153: 81–91

Gill BS, Friebe B (2002) Cytogenetics, phylogeny and evolution of cultivated wheats. In: Bread Wheat – Improvement and Production. Rajaram S., Curtis B.C., Gomez Macpherson H. (eds.), FAO Plant Production and Protection series No. 30, pp 567

Gill BS, Friebe B, Endo TR (1991) Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (Triticum aestivum). Genome 34: 830-839

Giroux MJ, Morris CF (1997) A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. Theor Appl Genet 95: 857–864

Giroux MJ, Morris CF (1998) Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. Proc Natl Acad Sci USA 95: 6262–6266

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8: 195-202

Greenwell P, Schofield JD (1986) A starch granule protein associated with endosperm softness in wheat. Cereal Chem 63: 379-380

He P, Friebe BR, Gill BS, Zhou J-M (2003) Allopolyploidy alters gene expression in the highly stable hexaploid wheat. Plant Mol Biol 52:401-414

Huang X, Medan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9: 868-877

Huang S, Sirikhachornkit A, Su X, Faris J, Gill BS, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99: 8133-8138

Jing W, Demeoe AR, Vogel HJ (2003) Conformation of a bactericidal domain of puroindoline a: structure and mechanism of action of a 13-residue antimicrobial peptide. J Bacteriol 185: 4938-4947
Li et al. Ha deletions during polyploid wheat speciation

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet 33: 102-106

Kihara H (1944) Die Entdeckung des DD-Analysators beim Weizen Agric Hortic 19:889-890

Kilian B, Ozkan H, Deusch O, Effgen S, Brandolini A, Kohl J, Martin W, Salamini F (2007) Independent wheat B and G genome origins in outcrossing Aegilops progenitor haplotypes. Mol Biol Evol. 24: 217-227

Krishnamurthy K, Balconi C, Sherwood JE, Giroux MJ (2001) Wheat puroindoline enhance fungal disease resistance in transgenic rice. Mol Plant-Microbe Interact 14: 1255-1260

Law CN, Young CF, Brown JWS, Snape JW, Worland AJ (1978) The study of grain protein control in wheat using whole chromosomes substitution lines. In Seed protein improvement by nuclear techniques. Int. Atomic Energy Agency, Vienna, Austria. pp. 483–502

Levy AA, Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biol J Linn Soc 82:607-613

Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core Triticeae genome. Plant J 40: 500-511

Lillemo M, Simeone MC, Morris CF (2001) Analysis of puroindoline a and b sequences from Triticum aestivum cv. “Penawawa” and related diploid taxa. Euphytica 126: 321-331

Liu B, Vega JM, Feldman M (1998) Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. Genome 41: 535-542

Massa, AN, Morris CF, Gill BS (2004) Sequence Diversity of Puroindoline-a, Puroindoline-b, and the Grain Softness Protein Genes in Aegilops tauschii Coss. Crop Sci 44: 1808-1816

Mattern JM, Morris R, Schmidt JW, Johnson VA (1973) Location of genes for kernel properties in wheat cultivar “Cheyenne” using chromosome substitution lines. Proc. 4th Intern. Wheat Genet Symp. (eds ER Sears and LMS Sears) University of Missouri, Columbia, MO. 703-707

McFadden ES, Sears ER (1946) The origin of Triticum spelta and its free-threshing hexaploid relatives. J. Heredity 37: 81-89 and 107-116
Li et al. Ha deletions during polyploid wheat speciation

McIntosh RA, Hart GE, Devos KM, Gale MD, Rogers WJ (1998) Catalogue of gene symbols for wheat. Proc. 9th Intern. Wheat Genet. Symp. (eds A. E. Slinkard) University Extension Press, Saskatoon, Canada 5: 1-235

Morris CF (2002) Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Molecular Biology 48: 633-647

Morris CF, Greenblatt GA, Bettge AD, Malkawi I (1994) Isolation and characterization of multiple forms of friabilin. J Cereal Sci 20:167–174

Morris CF, Lillemo M, Simeone MC, Giroux MJ, Babb SL, Kidwell KK (2001) Prevalence of puroindoline grain hardness genotypes among historically significant north american spring and winter wheats. Crop Sci 41: 218-228

Narasimhamoorthy B, Gill BS, Fritz AK, Nelson JC, Brown-Guedira GL (2006) Advanced backcross QTL analysis of a hard winter wheat × synthetic wheat population. Theor Appl Genet 112: 787-796

Nesbitt M, Samuel D (1996) From staple crop to extinction? The archaeology and history of hulled wheats. In : Padulosi S, Hammer K, Heller J, Pacoli C (eds) Hulled wheats. Promoting the conservation and use of underutilized and neglected crops. Proc 1st Int Workshop Hulled Wheats. International Plant Genetic Resources Institute, Rome, Italy, pp 41-100

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (Aegilops-Triticum) group. Plant Cell 13:1735-1747

Pontes O, Neves N, Silva M, Lewis MS, Madlung A, Comai L, Viegas W, Pikaard CS (2004) Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid Arabidopsis suecica genome. Proc Natl Acad Sci USA. 101: 18240-18245

See DR, Giorou M, Gill BS (2004) Effect of Multiple copies of puroindoline gene on grain softness. Crop Sci 44: 1248-1253

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell. 13: 1749-1759

Simeone M, Gedye KR, Mason-Gamer R, Gill BS, Morris CF (2006) Conserved regulator elements identified from a comparative puroindoline gene sequence survey of Triticum and Aegilops diploid taxa. J Cereal Sci 44: 21-33
Li et al. Ha deletions during polyploid wheat speciation

Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc Natl Acad Sci USA 92: 7719-7723

Swan CG, Bowman JGP, Martin JM, MJ Giroux (2006) Increased puroindoline levels slow ruminal digestion of wheat (Triticum aestivum L.) starch by cattle. J Anim Sci 84: 641-650

Tian L, Chen M, Wang J, Wei NE, Sze S-H, Chen ZJ (2005) The role of microRNAs in non-additive gene regulation in Arabidopsis allopolyploids. Plant and Animal Genome W363

Tsunewaki K (1993) Genome-plasmon interaction in wheat. Jpn J Genet 68: 1-34

Upadhya MD, Swaminathan MS (1963) Genome analysis in Triticum zhukovskyi, a new hexaploid wheat. Chromosoma 14: 589-600

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution. Plant J 26: 307-316

Wicker T, Matthews DE, Keller B (2002) TREP: a database for Triticeae repetitive elements. Trends in Plant Science 7: 561-562

WEB SITE REFERENCES

http://faostat.fao.org: FAO Statistics.
http://pbil.univ-lyon1.fr/cap3.php: PBIL CAP3 Sequence Assembly Program.
http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html: BCM multiple sequence alignment server.
http://sun1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind: Softberry FGENESH.
http://wheat.pw.usda.gov/ITMI/Repeats/blastrepeats3.html: Triticeae Repeat (TREP) database.
http://www.ch.embnet.org/software/BOX_form.html: BOXSHADE server.
http://www.fas.usda.gov/pecad/highlights/2005/10/durum_27oct2005: durum wheat production predicted by USDA Foreign Agricultural Service.
http://www.ncbi.nlm.nih.gov: NCBI databases.
