ELSEVIER

Journal of Cereal Science 44 (2006) 21–33

Journal of CEREAL SCIENCE

www.elsevier.com/locate/jnlabr/yjcrs

Conserved regulatory elements identified from a comparative puroindoline gene sequence survey of *Triticum* and *Aegilops* diploid taxa

Marco C. Simeone [a,b,*], Kristene R. Gedye [c,d], Roberta Mason-Gamer [f], Bikram S. Gill [g], Craig F. Morris [e]

[a] *Department of Technologies, Engineering and Sciences of Forests and Environment, University of Tuscia, via S. Camillo de Lellis, 01100 Viterbo, Italy*
[b] *USDA-ARS Western Wheat Quality Laboratory, Pullman, WA, USA*
[c] *Department of Food Science and Human Nutrition, Western Wheat Quality Laboratory, Washington State University, Pullman, WA, USA*
[d] *Department of Plant Pathology, Washington State University, Pullman, WA, USA*
[e] *USDA-ARS Western Wheat Quality Laboratory, E-202 Food Science and Human Nutrition Facility East, P.O. Box 646394, Washington State University, Pullman, WA 99164-6394, USA*
[f] *Department Biological Sciences, University of Illinois at Chicago, Chicago, IL, USA*
[g] *Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA*

Received 11 October 2005; received in revised form 3 February 2006; accepted 11 February 2006

Abstract

Kernel texture ('hardness') is an important trait that determines end-use quality of wheat (*Triticum aestivum* L. and *Triticum turgidum* ssp. *durum* [Desf.] Husn.). Variation in texture is associated with the presence/absence or sequence polymorphism of two proteins, puroindoline a and puroindoline b. This work describes the flanking and coding region sequences of puroindoline genes from 25 accessions representing wild diploid taxa of the Triticeae related to the three genomes of *T. aestivum*. Analysis of variation at the nucleotide level included hard and soft *T. aestivum* wheat cultivars. Various degrees of insertions/deletions and point mutations were found, that did not affect the overall sequence structure identity. Nucleotide sequence comparisons and database searches facilitated the identification of the 5′ proximal regulating regions, revealing the presence of several putative control elements. An absolute conservation of some known regulatory elements for tissue specificity was observed, while different rates of conservation of reiterated motifs with possible enhancer functions, and the exclusive presence of some elements either in puroindoline a or puroindoline b were also found. A total of 24 new puroindoline alleles (unique sequences) were identified. Despite some primary structure variation, the main features of puroindolines, i.e. the signal peptide, the cysteine backbone, the tryptophan-rich domain, the hydrophobicity and basic identity of the proteins were all conserved.
Published by Elsevier Ltd.

Keywords: Puroindolines; Kernel texture; Grain hardness; Gene sequence; Promoter; Triticeae; Wheat; Evolution

1. Introduction

Puroindolines are small (ca. 13 kDa) basic, cysteine-rich proteins found in the caryopses of many taxa of the Triticeae tribe. They belong to a protein super-family (Douliez et al., 2000) that includes *alpha*-amylase/trypsin inhibitors, non-specific lipid binding proteins and a mixture of puroindoline-like polypeptides (Grain Softness Proteins, GSPs). In recent years, puroindolines have gained a considerable interest among wheat geneticists and breeders, offering new perspectives in the genetic improvement of cereals, spanning from end-use applications (Morris, 2002) to disease resistance (Giroux et al., 2003). Puroindolines were first described as a mixture, i.e. 'friabilin', in the surface of water-washed starch from *Triticum aestivum* L. endosperm (Greenwell and Schofield, 1986). Subsequently, puroindolines were resolved individually from wheat flour by Triton X-114 detergent phase partitioning, named and sequenced (Blochet et al., 1991, 1993). Wheat puroindolines exist in two isoforms—termed puroindoline a and puroindoline b—sharing 60% amino acid sequence identity (Gautier et al., 1994), unique tryptophan-rich lipid binding domains and 10 cysteine residues. In addition to interacting with lipids and stabilizing foams (Douliez et al., 2000), puroindolines affect dough quality (Igresias et al., 2001), exhibit anti-microbial (Dubreil et al., 1998) and anti-fungal activity

* Corresponding author. Address: Department of Technologies, Engineering and Sciences of Forests and Environment, University of Tuscia, via S. Camillo de Lellis, 01100 Viterbo, Italy.
E-mail address: mcsimeone@unitus.it (M.C. Simeone).

0733-5210/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.jcs.2006.02.002

(Krishnamurthy et al., 2001), and form ion channels in biological membranes (Charnet et al., 2003).

Puroindolines confer the soft kernel texture phenotype in the hexaploid wheat *T. aestivum* (2*n* = 42, AABBDD), as alteration in puroindoline b sequence or the absence of puroindoline a are associated with the hard kernel phenotype (Morris, 2002). The genes encoding puroindolines reside on the short arm of *T. aestivum* chromosome 5D (together with GSP-1), are inseparably linked to the *Hardness* locus, and are absent from the homoeologous chromosomes 5A and 5B. The durum wheat *Triticum turgidum* ssp. *durum* (2*n* = 28, AABB) similarly lacks the puroindoline genes on 5A and 5B, and has very hard kernels. The genomic rearrangements that shaped the *Hardness* locus in wheat, under the form of multiple insertions, deletions, duplications and insertions of transposable elements, have recently been explained by a comparative analysis of BAC clone sequences from *Triticum monococcum*, *Aegilops tauschii* and *T. aestivum* (genome D) by Turnbull et al. (2003) and Chantret et al. (2004, 2005).

Due to the evolution of *T. aestivum*, which involved combining an AB tetraploid with a D-genome bearing *Ae. tauschii*, all puroindolines sequence variants in wheat are considered to have arisen since the polyploidation event some 7000–10,000 years ago. To date, only one unique sequence each of puroindoline a and b are found associated with soft kernels in *T. aestivum*, whereas all known alterations (currently 12 total) are associated with hard kernel phenotype. Most of the sequence variants involve puroindoline b, with four single nucleotide polymorphisms (SNPs) changing a single amino acid, three SNPs creating a premature 'stop' codon, and two single nucleotide deletions producing open reading frame shifts (Chen et al., 2005; Morris, 2002; Ram et al., 2005; Xia et al., 2005). The last three variants involve puroindoline a with two large deletions in the gene regions that leave puroindoline b intact (Cloutier, personal communication), or remove both (Tranquilli et al., 2002), and a single nucleotide deletion that produces an open reading frame shift (Gazza et al., 2005). The likely occurrence of additional puroindoline sequence variants in *T. aestivum* have been indicated by Corona et al. (2001), based on different mobility in A-PAGE fractionation of starch granule proteins.

Jolly (1991) detected puroindolines via Western blotting in *T. monococcum* (A^m), *Triticum urartu* (A^u), *Ae. tauschii* (D), *Aegilops uniaristata* (N), *Aegilops sharonensis* (S^{sh}), *Aegilops speltoides* (S^s), *Aegilops bicornis* (S^b), *Aegilops caudata* (C), several subspecies of *T. aestivum* (ABD) and *Secale cereale* (R); puroindolines were subsequently detected, albeit in low levels, in *Aegilops comosa* (M), *Aegilops umbellulata* (U), *Aegilops neglecta* (UM and UMN [syn. *Aegilops recta*]) and *Hordeum vulgare* (H). Gautier et al. (2000) used the polymerase chain reaction (PCR) and Southern blot analysis to show that both puroindoline a and b were present in *T. monococcum* ssp. *aegilopoides* (A^b) and *monococcum*, *Ae. speltoides*, *Ae. tauschii*, *Ae. comosa*, *Ae. caudata*, *Ae. umbellulata*, *H. vulgare*, *Avena sativa* and *T. aestivum*, including cv. Chinese Spring (all foregoing nomenclature revised according to van Slageren, 1994). In the same study, the tetraploid taxa *Triticum dicoccoides*, *Triticum dicoccum*, *Triticum durum* and *Triticum timopheevii* were found to be devoid of puroindoline sequences. Puroindoline a, b and GSP gene sequences were also isolated and characterized in rye (Gautier et al., 2000; Simeone and Lafiandra, 2005).

In addition to the simple presence or absence of puroindoline genes and their proximal relationships in the *Triticeae* taxa, the specific sequence of puroindoline a and b, and each of their immediate flanking regions are of interest. The promoter of a puroindoline b from wheat was first cloned and sequenced by Digeon et al. (1999). Subsequently, coding sequence and flanking regions of puroindoline b genes were resolved by Darlington et al. (2001) in barley, and by Lillemo et al. (2002), who described the promoters of puroindolines a and b in two accessions each of *T. monococcum* ssp. *monococcum*, *T. urartu*, *Ae. speltoides* and *Ae. tauschii*. In these studies, some conserved regions bearing putative regulatory elements were proposed which stimulated further studies to verify the extent of their significance. Indeed, the yet unknown regulation of puroindoline gene transcription, expression and the pathways of protein deposition, as well as the probable existence of mechanisms of different gene regulation, are an interesting matter of concern in the context of the biological functioning of puroindolines, and for the future exploitation of this important trait. Actually, some case-studies where puroindoline transcripts are present, but expressed at different levels, have been recently described in a large germplasm screening and await a molecular description (Capparelli et al., 2003). In this view, a wider inter- and intra-specific assessment of the puroindoline gene sequences may reveal the degree of conservation of previously identified—and possibly new—controlling elements, and new allelic forms, which might all relate to differences in gene regulation and protein activity.

In the study reported here, the nucleic acid gene sequences and a portion of the promoter regions of puroindoline a and b from several diploid *Triticum* and *Aegilops* taxa (genome groups A^u, A^b, A^m, S, S^b, S^l, S^s, S^{sh}, D) are described. The 5′ promoter domains of puroindolines are analyzed in the context of their sequence structure and identity. The occurrence of several putative regulatory motifs is investigated, and the relationships between puroindoline a and b, genome groups and cultivated wheat are presented and discussed. Extensive variation at the deduced amino acid level is also reported, and considered in relation to the soft phenotype of the studied germplasm.

2. Experimental

2.1. Germplasm

Seed of all diploid taxa and cv. Chinese Spring were obtained from the Wheat Genetics Resource Center, Kansas State University (Table 1). The *T. aestivum* cultivars Cheyenne (CItr 8885), Falcon (PI 292578), ID377s (PI 591045), Komar (CI 008004), Lewjain (CItr 17909) and Penawawa (PI 495916) were obtained from Dr Harold Bockelman, Curator, USDA-ARS National Small Grains Collection, Aberdeen, Idaho.

Table 1
Plant materials used to survey puroindoline gene sequence of Triticeae diploid taxa

Genus species	Subspecies	Genome	Accession or cultivar	Puroindoline a allele	NCBI accession	Puroindoline b allele	NCBI accession
T. urartu		A	TA763	*Pina-A1a*	AJ302094[a]	*Pinb-A1a*	AJ302103[a]
T. urartu		A	TA808	*Pina-A1a*	AJ302095[a]	*Pinb-A1b*	AJ302104[a]
T. urartu		A	TA828	*Pina-A1a*	AJ302094[b]	*Pinb-A1b*	AJ302103[b]
T. urartu		A	TA829	*Pina-A1a*	AJ302094[b]	*Pinb-A1b*	AJ302103[b]
T. monococcum	*monococcum*	A^m	TA2025	*Pina-A^m1b*	AY622786	*Pinb-A^m1c*	AY622797
T. monococcum	*monococcum*	A^m	TA2026	*Pina-A^m1b*	AY622786[b]	*Pinb-A^m1d*	AY622798
T. monococcum	*monococcum*	A^m	TA2037	*Pina-A^m1c*	AJ242715[c]	*Pinb-A^m1e*	AJ302102[c]
T. monococcum	*aegilopoides*	A^m	TA183	*Pina-A^m1b*	AY622786[c]	*Pinb-A^m1g*	AY622799
T. monococcum	*aegilopoides*	A^m	TA291	*Pina-A^m1b*	AY622786[c]	*Pinb-A^m1g*	AY622799[c]
T. monococcum	*aegilopoides*	A^m	TA546	*Pina-A^m1b*	AY622786[c]	*Pinb-A^m1g*	AY622799[c]
T. monococcum	*aegilopoides*	A^m	TA581	*Pina-A^m1b*	AY622786[c]	*Pinb-A^m1g*	AY622799[c]
Ae. speltoides	*speltoides*	S	TA2368	*Pina-S1c*	AY622787	*Pinb-S1c*	AY622801
Ae. speltoides	*speltoides*	S	TA1789	*Pina-S1d*	AY622788	*Pinb-S1d*	AY622802
Ae. speltoides	*ligustica*	S	TA1777	*Pina-S1e*	AY622789	*Pinb-S1e*	AY622803
Ae. longissima	*longissima*	S^l	TA1912	*Pina-S^l1a*	AY622790	*Pinb-S^l1a*	AY622800
Ae. longissima	*nova*	S^l	TA1921	*Pina-S^l1b*	AY622791	*Pinb-S^l1b*	AY622804
Ae. searsii		S^s	TA1837	*Pina-S^s1a*	AY622792	*Pinb-S^s1a*	AY622805
Ae. searsii		S^s	TA2355	*Pina-S^s1b*	AY622793	*Pinb-S^s1b*	AY622806
Ae. bicornis	*typica*	S^b	TA1954	*Pina-S^b1a*	AY622794	*Pinb-S^b1a*	AY622807
Ae. bicornis	*typica*	S^b	TA1942	*Pina-S^b1b*	AY622795	*Pinb-S^b1b*	AY622808
Ae. sharonensis	*sharonensis*	S^{sh}	TA1999	*Pina-S^{sh}1a*	AY622796	*Pinb-S^{sh}1a*	AY622809
Ae. tauschii	*tauschii*	D	TA1704	*Pina-D1d*	AY649744	*Pinb-D1i*	AY649747
Ae. tauschii	*meyeri*	D	TA1691	*Pina-D1a*	AY252013	*Pinb-D1j*	AY251964
Ae. tauschii	*anathera*	D	TA2381	*Pina-D1d*	AY649745	*Pinb-D1i*	AY649747[c]
Ae. tauschii	*strangulata*	D	TA10	*Pina-D1c*	AY649746	*Pinb-D1h*	AY649748
T. aestivum		AABBDD	Chinese Spring	*Pina-D1a*	DQ363911	*Pinb-D1a*	DQ363913
T. aestivum		AABBDD	Cheyenne	*Pina-D1a*	DQ363912	*Pinb-D1b*	DQ363914

Species classification according to van Slageren (1994).
[a] Sequences identical with NCBI accession numbers and diploid accession from Lillemo et al. (2002).
[b] Sequences identical with accession numbers indicated by Lillemo et al. (2002).
[c] Sequences identical to NCBI accession numbers, with corresponding allele.

The *T. aestivum* cultivars Butte 86, Express and Westbred 906 were obtained from Greg Vollmer, Washington State University Foundation Seed Program. Opata was provided by Prof. Calvin Qualset, University of California, Davis, CA.

2.2. DNA isolation and PCR amplification of puroindolines

DNA was extracted from leaf tissue (diploid taxa) or single kernels (*T. aestivum* cultivars) using a slightly modified procedure of Dellaporta et al. (1983). Amplification of puroindoline a and b coding sequences and related upstream and downstream regions were obtained as described in Lillemo et al. (2002). Primers were developed in the same regions, where possible, for better comparison of the amplified fragments. The following primer combinations were used: puroindoline a, PAA1/PAA2 for the A-genome taxa, PAD3/PAD2 for S- and D-genome taxa and *T. aestivum* cultivars; puroindoline b, PIPM2/PBA2 for the A- and D-genome taxa and *T. aestivum* cultivars, PIPM2/PBB2 for the S-genome taxa. Primer sequences are listed in Table 2. All PCR reactions were performed with the High Fidelity PCR enzyme (Boehringer Mannheim) and run as follows: the samples were denatured initially for 3 min at 94 °C, followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 58 °C for 1 min, and elongation at 72 °C for 1 min; a final elongation step of 5 min completed the cycle. Amplified fragments were separated by electrophoresis on 1.0–1.4% (w/v) agarose gels.

2.3. DNA sequencing and analysis

Unambiguous, single-banded fragments were recovered from the gels, purified using QIAquick Gel extraction kit (QIAGEN) and directly sequenced in both directions with the amplification primers. Forward and reverse sequence overlaps were obtained by use of internal primers determined upon sequencing. Cycle Sequencing and the BigDye Terminator Ready reaction kit (Applied Biosystems) were used. Data were collected on ABI Prism 377 automated gel reader at the laboratory of Biotechnology and Bioanalysis, Washington State University. Electropherograms were further checked by eye with the software CHROMAS 2.3 (www.technelysium.com.au), assembled and aligned for several standard descriptive parameters (including size, gap positioning, percentage pairwise distance, search for specific nucleotide motifs, protein translation, hydropathicity profiles and pI estimations) with Clustal W (Thompson et al., 1994) and with the Sequence Analysis Software DNAMAN-1999 (Lynnon Biosoft). Mutations detected in only one genotype were verified by a second PCR and sequencing analysis, to eliminate SNPs that might have been caused by DNA polymerase errors.

Table 2
Primers used for the amplification of the puroindoline a and b gene sequences from diploid Triticeae taxa and *T. aestivum*

Puroindoline a		Puroindoline b	
Strand	Antistrand	Strand	Antistrand
5′-CCGATGCATCTA-GATCCTCA-3′ (PAA1) 5′-CCGATGCATCTA-GATCCTCG-3′ (PAD3)	5′-TGGTATGTGACAGTTTAT-TAGCTAGTCAC-3′ (PAA2) 5′-TGGTATGTGACAGTTTAT-TAGCTAGTCAT-3′ (PAD2)	5′-CCACACACTACAAGGC-CAGTTC-3′ (PIPM2)	5′-CTCATACACAGGTATCTC-CAACACAA-3′ (PBA2) 5′-CTCCTACACAGATCAACA-TACAT-3′ (PBB2)

The sequences were derived from *T. monococcum* (PAA1/PAA2, Genbank acc. no. AJ302092), *Ae. speltoides* (PBA2/PBB2, AJ302105), *T. tauschii* (PAD3/PAD2, AJ302108) after Lillemo et al. (2002), and hexaploid wheat (PIPM2, AJ000548) after Digeon et al. (1999).

3. Results

3.1. Isolation of puroindoline genes

All primer combinations amplified a single band from each of the genome groups. Sequencing and comparisons with known puroindoline sequences identified the puroindoline coding and promoter regions. Twenty-four novel puroindoline sequences were isolated, characterized and assigned allelic designations following the convention of the catalogue of gene symbols in wheat (McIntosh et al., 2003). Of these, 11 novel sequences were found for puroindoline a and 13 for puroindoline b. Novel alleles were submitted to the NCBI database under accessions AY622786–AY622809; accession numbers and allele designations are detailed in Table 1. All attempts to amplify promoter sequences from the puroindoline a-null (*Pina-D1b*) hard wheat cultivars Falcon, ID377s, Komar, Butte 86, Express, and Westbred 906 were unsuccessful. The length of the puroindoline coding and flanking regions analysed is presented in Table 3.

3.2. Analysis of 5′ and 3′ non coding sequences in puroindoline a

Primer set PAA1/PAA2 in combination with template DNA from all accessions of the A-genome taxonomic group produced a single PCR amplification product extending from 1336 (*T. monococcum* ssp. *monococcum*, 1363 bp in accession TA2037) to 1571 bp (*T. monococcum* ssp. *aegilopoides*). The four *T. urartu* accessions produced a band of 1363 bp. Comparison of the promoter sequences of the puroindoline a genes from the three A-genome taxa identified a 242-bp insertion at position −700 from the start codon in *T. monococcum* ssp. *aegilopoides*, two insertions/deletions of 32 and 9 bp shared by *T. urartu* and accession TA2037, and a few other minor insertions (1–6 bp).

Primer combination PAD3/PAD2 produced PCR amplification products in the S- and D-genome taxa. The amplified fragments from the S-genome taxa ranged in size from 1286 bp in *Ae. sharonensis* to 1396 bp in *Ae. searsii* (TA2355). This size range was due to various insertions/deletions in the promoter regions. The largest insertion, 83 bp at position −399, occurred in the *Ae. searsii* accessions.

Three of the four *Ae. tauschii* accessions (TA1704, TA2381 and TA10) produced a 1301 bp PCR product. The remaining *Ae. tauschii* accession (TA1691) produced a 1314 bp PCR product due to a 13-bp insertion at position −649.

Five *T. aestivum* cultivars (Cheyenne, Chinese Spring, Falcon, Lewjain and Penawawa) were used in a preliminary screening and all but Falcon produced the expected PCR product of 1339 bp (Lillemo et al., 2002). Although Cheyenne is a hard wheat cultivar, it carries the soft *Pina-D1a* allele; Chinese Spring, Lewjain and Penawawa are soft cultivars.

Table 3
Length (bp) of 5′, coding and 3′ downstream sequences determined for the puroindoline genes in every genotype

Species/subspecies	Puroindoline a			Puroindoline b		
	5′ Upstream region	Coding sequence	3′ Downstream region	5′ Upstream region	Coding sequence	3′ Downstream region
T. urartu	873	447	43	258	447	108
T. m. Ssp. *monococcum*	874	447	43	258	447	108
TA 2037	846	447	43	258	447	108
T. m. ssp. *aegilopoides*	1081	447	43	258	447	108
Ae. speltoides	821	447	43	272	444	108
TA 1777	827	447	43	272	444	108
Ae. longissima	799	447	43	272	447	108
TA 1921	813	447	43	426	447	108
Ae. searsii	906	447	43	272	444	108
Ae. bicornis	810	447	43	426	447	108
TA 1942	819	447	43	426	447	110
Ae. sharonensis	796	447	43	272	447	108
Ae. taschii ssp. *tauschii*	811	447	43	276	447	108
ssp. *meyeri*	824	447	43	273	447	108
ssp. *anathera*	811	447	43	276	447	108
ssp. *strangulata*	811	447	43	276	447	108
T. aestivum	849	447	43	273	447	108

Falcon is known to be a puroindoline a-null genotype (*Pina-D1b*) (Giroux and Morris, 1997) and may share the same large deletion which prevents gene transcription as found in cultivar Glenlea (Cloutier, personal communication). All other puroindoline a-null cultivars examined in this study (ID377s, Komar, Butte 86, Express, Westbred 906 and Opata) also failed to produce puroindoline a PCR amplification product. Puroindoline a PCR products from the *T. aestivum* cultivars Chinese Spring and Cheyenne were sequenced and found to be identical in every gene region.

The identity percentages of puroindoline a promoter sequences is presented in Table 4.

The overall identity was 70.2%. The A-genome taxa shared 83.4% identity, the S-genome taxa 88.0%, and the D-genome diploid accessions 98.1%.

The four *T. urartu* accessions showed very high sequence identity (99.9%), with only a single nucleotide change (at position −42 from the start codon) in accession TA763. All four accessions of *T. monococcum* ssp. *aegilopoides* showed complete sequence identity in the promoter region and along the entire fragment. The promoter of *T. monococcum* ssp. *monococcum* accession TA2037 differed from the other two accessions of this taxon (TA2025 and TA2026) that were identical. Sequence insertions and point mutations make the promoter of accession TA2037 sharing higher sequence identity with *T. urartu* accessions (99.3%) rather than to the other *T. monococcum* ssp. *monococcum* accessions (98.7%).

The promoter sequence identity in the S-genome taxa ranged from 98.0% in *Ae. speltoides* to 99.1% in *Ae. bicornis*. Two accesions of *Ae. speltoides* (TA1777 and TA2368) were 99.1% identical, whereas TA1789 diverged sensibly (2.5%). Remarkably, *Aegilops longissima* TA1912 shared highest identity with *Ae. sharonensis* (99.9%) than with the co-specific accession TA1921 (98.5%).

The promoters of the *Ae. tauschii* accessions revealed a very high level of sequence identity: *Ae. tauschii* ssp. *typica*, ssp. *anathera* and ssp. *strangulata* were more similar to each other (99.8–100%) than to ssp. *meyerii* (98.6–98.9%). Finally, the promoters of hexaploid wheats Chinese Spring and Cheyenne shared 100% identity with the promoter of cultivar Penawawa (Lillemo et al., 2002).

Across all taxa, the 3′ downstream regions of puroindoline a (43 bp) differed in only 1 bp (3′ end of primers PAA2 and PAD2).

3.3. Analysis of 5′ and 3′ non coding sequnces in puroindoline b

Primers PIPM2/PBA2 amplified puroindoline b from all diploid accessions of the A- and D-genome taxa, and from the *T. aestivum* cultivars. Of the hexaploid wheats only Chinese Spring and Cheyenne were further analyzed.

All of the A-genome accessions produced a PCR product 813 bp long. Compared to the D-genome taxa and *T. aestivum*, few minor insertons/deletions (1–3 bp) occurred in the 5′ upstream region of the A-genome group, with a larger one (16-bp) at position −120.

Primers PIPM2/PBB2 allowed the amplification of a fragment of 824–827 bp in all accessions of the S-genome group, except *Ae. longissima* ssp. *nova* and *Ae. bicornis* (both accessions) where additional 156 bp were amplified at the start of the promoter sequence. A 2-bp insertion was also found in the 3′ downstream region of accession TA1954.

Puroindoline b promoter sequences were highly variable across the entire sample, with an overall sequence identity of just 66.5%, compared to 92.8% identity among the puroindoline a promoters (Table 4), considering approximately the same range of length (280 bp).

Consistent with the results observed in the promoters of puroindoline a, every genome group showed very high puroindoline b promoter sequence identity, ranging from 99.2 (*Ae. speltoides*) to 100% (*T. urartu* and *T. monococcum* ssp. *aegilopoides*), with the exception of the *Ae. longissima* accessions (97.5%; with accession TA1912 sharing 98.9% with *Ae. sharonensis*). Also consistent were the results of accession TA2037 (*T. monococcum* ssp. *monococcum*), which shared 98.8% and 99.2% identity with the other two accessions

Table 4
Identity percentages of puroindoline a and b sequences in the diploid Triticeae taxonomic groups

Species or group	Puroindoline a				Puroindoline b		
	Promoter		Coding region	3′ Downstream region	Promoter (280 bp)	Coding region	3′ Downstream region
	Full length	280 bp					
T. urartu	99.9	98.5	100	n.d.	100	99.9	100
T. m. ssp. *monococcum*	94.5	97.9	99.6	n.d.	99.6	99.9	100
T. m. ssp. *aegilopoides*	100	100	100	n.d.	100	100	100
AA diploids	83.4	97.7	99.6	n.d.	99.6	99.6	100
Ae. speltoides	98.0	99.6	99.7	n.d.	99.2	98.6	98.3
Ae. longissima	98.5	97.5	99.8	n.d.	97.8	99.1	100
Ae. searsii	98.8	99.6	99.3	n.d.	99.3	99.6	100
Ae. bicornis	99.0	99.9	99.6	n.d.	99.8	99.6	99.5
SS diploids	88.0	97.1	99.3	n.d.	74.0	98.8	96.5
DD diploids	98.1	99.6	99.8	n.d.	98.6	99.0	99.0
All studied germ plasm	70.2	92.8	98.4	n.d.	66.5	96.2	92.1

Accessions grouped according to their species and genome. Identity was not estimated in the 3′ downstream regions of puroindoline a (43 bp), which differ in only 1 bp (3′ end of primers PA2 and PAD2) across all the analysed germplasm.

of ssp. *monococcum* (TA2025 and TA2026), 99.6% with the four *T. urartu* and the four *T. monococcum* ssp. *aegilopoides* accessions. The remaining two *T. monococcum* ssp. *monococcum* accessions (TA2025 and TA2026) shared 99.6% identity.

Sequence identity within the S-genome group as a whole (74.0%) was lower than that within the A-genome (99.6%) or the D-genome (98.6%) groups. In this respect, and similar to the puroindoline a promoters, *Ae. tauschii* ssp. *meyerii* appeared to be distinct from the other *Ae. tauschii* accessions (95.6 vs. 100.0% sequence identity), and shared the highest identity (99.3%) with the puroindoline b fragment from Chinese Spring. With regard to the hexaploid wheats, the promoters of Chinese Spring and Cheyenne shared 100% sequence identity with cultivar Penawawa and 99.3% (only two nucleotide substitutions at position −102 and −184) with cultivar Andain (Digeon et al., 1999).

The 3′ downstream regions were 92.1% identical. Intra-genome group values ranged from 96.5 (in the S-genome taxa) to 100% (in the A-genome taxa). The lowest intra-specific identity was recorded among the *Ae. speltoides* accessions.

3.4. Identification of regulator elements in the non-coding sequences

Consistent with the results observed by Lillemo et al. (2002), the promoter sequences of both puroindoline a and puroindoline b had identical TATA motifs (TAAATAAA) located at the same position (−70/−62 from the start codon) in all the analyzed germplasm (see Fig. 1 to locate every investigated DNA motif). The only exception was found in the puroindoline b sequence of *Ae. searsii* (TA1837) in which the motif changed to TAAATTTA.

The putative CAAT-Box (CCAAT) was always present in an inverted orientation at position −140/−120 in every puroindoline a promoter, whereas CACAAT was found in the correct orientation in every puroindoline b promoter, at position -105/-95. The puroindoline b sequences of *Ae. tauschii* (TA1691), *Ae. searsii* (both accessions) and the puroindoline a of *Ae. bicornis* (TA1954) showed the CAAT-Box elements mutated to either NCAAT or CCAAAT. In these taxa, the eventual loss of functionality of the CAAT sequence would be compensated by an extra CCAAT element located further upstream (interestingly, this additional element is absent in the puroindoline promoters showing a common consensus sequence for the CAAT-Box). In the 3′ downstream region, the polyadenylation site (AATAAA) was found at position +25 in puroindoline a and at +24 and +83 in puroindoline b. This element was the only one recorded in the 3′ flanking region analysed in the puroindoline genes of the present work.

Other putative regulatory elements were identified in the 5′ upstream sequence by searches in the PLACE database (Higo et al., 1999) and are listed in Table 5. The prolamin-box and the $(CA)_n$-box are the only two motifs almost perfectly conserved in sequence and position in both puroindolines of every accession in the present study. The prolamin-box contains a consensus sequence, TGAGAAAAG, at position −223/−202 in both puroindoline a and b, exceptions were observed in puroindoline a of *T. monococcum* ssp. *monococcum* (TA2037) ($TGAG(A)_5G$), *Ae. speltoides* ($TGAC(A)_4G$), *Ae. sharonensis* and *Ae. longissima* ssp. *longissima* ($TGAGT(A)_3G$). Another variant of the same element (TGTAAAAT) is present ca. 430 bp further upstream in all puroindoline a promoters. The $(CA)_n$-box consensus sequence CAACAAC occurs at −180/−162 and −236/−226 in puroindoline a and −183/−161 in puroindoline b. The only exception occurs in puroindoline b of *Ae. longissima* ssp. *longissima* (TA1921), in which this motif changed to CAACCAC. However, in this accession the eventual loss of function might be potentially restored by the presence of another form of the $(CA)_n$-box with sequence AACACA. This additional $(CA)_n$-box is absent in the other puroindoline b promoters except *Ae. bicornis*, which shares the same 153-bp insertion with accession TA1921, containing the AACACA element. The presence of the (CA)-box in the 153-bp insertion in these two taxa is the only putative regulatory element observed in an insertion/deletion segment throughout all the analyzed germplasm.

Other putative regulatory elements showed specificity for only one puroindoline gene promoter. For example, the P-Binding site (AACAACC) at positions −630/−616 and −286 and the Amy-Box at position −680/−637 occurred only in puroindoline a; conversely, the NtBB1 motif (ACTTTA) was found exclusively in the promoters of puroindoline b at −146/−137.

Other motifs were more recurrent and/or more highly conserved in the promoters of puroindoline a rather than in those of puroindoline b. For example, the AACACA variant of the $(CA)_n$-box was located at −102 and −14 in every puroindoline a, while this motif was only found at −102 in puroindoline b of the A- and D-genome taxa. A second $(CA)_n$-box in puroindoline b, located at −401/−385, was found in only a few genotypes of the the A- and D-genome groups. Apparently, no putative regulatory elements were found to be more conserved in the promoters of puroindoline b than in the upstream sequences of puroindoline a.

Specific putative regulatory motifs were found to be conserved only among genotypes with the same genome composition. For example, the ACGT-Box is present at −576/−533 in every puroindoline a. A second ACGT-Box is only present at −487 in the diploid species bearing the D-genome, and at −195 only in the puroinoline b of hexaploid wheats and *Ae. tauschii* ssp. *meyeri*. The E-Box (CANNTG) is present (and at the same positions) twice in the A-genome group, four times in the D-, and three times in only some of the S-genome diploids, whereas it was found only once in the puroindoline b promoter of A- and D-genome groups. A similar pattern was observed for the $CAA(N)_{2-9}TTG$ element in the puroindoline a promoter, which occurred once in the *T. monococcum* germplasm (except TA2037), twice in the other A-genome group accessions, three times in the S-genome diploids, and four times in D-genome bearing diploid and hexaploid wheats. In the puroindoline b promoter, $CAA(N)_{2-9}$ TTG was found once in the A-, four times in the S-, three times in three D-genome genotypes, and four times in hexaploid wheat

and *T. tauschii* ssp. *meyerii*. The motif AACNNA is even more common and conserved over all the genotypes: this sequence occurred from five to nine times in the puroindoline a promoter (D-genome group and *T. monococcum* accessions, respectively), while in puroindoline b promoters the motif had a perfect identity in sequence and location in the A- and D-genome groups but was absent in the S-genome accessions. The DOF-CORE recognition sequence (AAAG) was frequently interspersed in every promoter sequence. Additional putative regulatory elements occurred less frequently. Among these, the third known variant for the $(CA)_n$-box, CNAACAC, was present only in accessions of the S-genome group. The RY REPEAT (CATGCA) was present only in the puroindoline a of *Ae. searsii*, and the G-Box was not found. All of the investigated putative regulatory elements were detected (at least once) in a region of approximately 300 bp preceding the start codon of both puroindoline promoters. The only elements present exclusively further upstream were the Amy-Box and the ACGT-Box in puroindoline a.

Finally, no conserved direct, inverted or mirror repeats were detected in any sequence, with the exception of a di-nucleotide run $(GA)_n$ located at around −500 in every puroindoline a. Interestingly, the length of the repeat showed an intriguing pattern of conservation: $(GA)_6$ in *T. urartu* and *T. monococcum* accession TA2037, $(GA)_9$ in *T. monococcum* ssp. *monococcum*, $(GA)_8$ in *T. monococcum* ssp. *aegilopoides*, $(GA)_{7–10}$

```
-849    CTTGTTAAAAAAAATTACTTCCTTGGAGACAAATACTATTTTGGAGAATAATGAAAATCA
-789    TTTAGAAACATTTGACATGAACGACATGATTCCAGAGACAATAAAAAACTTAAAAACAAA
-729    CGATGGATAAGAAGATGATGTCTTGATAAATCATTACTTGTCGAGTTGTCGTGTACTACT
-669    AGTCTGTAAAATACAGTCTCTAACAAATTTGGTACAATCTGAGATCCATCTTGAACAACC
-609    TGCACAATCCTACAAGTTTAGTTTCGCAAAAGAATAACAATATGAACCATGTGACTTTCT
-549    TGGTACGTACAAAACCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAGGGT
-489    CACACATCATTAGATAAGGTCTATCTTCACAAAGTTGCACATACATGTACAGTACAGGAA
-429    GCGACATGTATCTCAATACCACATGGTTCTAGATACTGGACGAAAAAGCAGTGGCTAGAA
-369    AGATGACGATATATAGATGCATTACTTATCATATACTACTACCTAGAAAAATACAATATC
-309    TAATTTCCTCTTGATCCTTCTTGAACAACCTGCACAACACTACAAGTTCAGTTTCACAAA
-249    AGCGTAAGTCTAAAGCTTTGGTACAACAACAACTTATGGTTTATTTTGAGAAAAGGTCAG
-189    ATTCAGTACACGGAACATCACATATCTCAACAACTTCCACCAGTTTTGTGTGCTTTCAAA
-129    GTAACTTTGATTGGTATCCAGCTATACAACACACAACCGCACACAGAAATCGTGCCACCT
 -69    CAATTATAAATAAAGGTGTGGCCTCATCTCATCTATTCATCTCCACCTGCACCAAAACAC
  -9    ACTGACAACATGAAGGCCCTCTTCCTCATAGGACTGCTTGCTCTGGTAGCGAGCACCGCC
                 M  K  A  L  F  L  I  G  L  L  A  L  V  A  S  T  A

  52    TTTGCGCAATATAGCGAAGTTGTTGGCAGTTACGATGTTGCTGGCGGGGGTGGTGCTCAA
   1     F  A  Q  Y  S  E  V  V  G  S  Y  D  V  A  G  G  G  G  A  Q
         ▲                                 ▲
 112    CAATGCCCTGTAGAGACAAAGCTAAATTCATGCAGGAATTACCTGCTAGATCGATGCTCA
  10     Q  C  P  V  E  T  K  L  N  S  C  R  N  Y  L  L  D  R  C  S

 172    ACGATGAAGGATTTCCCGGTCACCTGGCGTTGGTGGAAATGGTGGAAGGGAGGTTGTCAA
  30     T  M  K  D  F  P  V  T  W  R  W  W  K  W  W  K  G  G  C  Q

 232    GAGCTCCTTGGGGAGTGTTGCAGTCGGCTCGGCCAAATGCCACCGCAATGCCGCTGCAAC
  50     E  L  L  G  E  C  C  S  R  L  G  Q  M  P  P  Q  C  R  C  N

 292    ATCATCCAGGGGTCAATCCAAGGCGATCTCGGTGGCATCTTCGGATTTCAGCGTGATCGG
  70     I  I  Q  G  S  I  Q  G  D  L  G  G  I  F  G  F  Q  R  D  R

 352    GCAAGCAAAGTGATACAAGAAGCCAAGAACCTGCCGCCCAGGTGCAACCAGGGCCCTCCC
  90     A  S  K  V  I  Q  E  A  K  N  L  P  P  R  C  N  Q  G  P  P

 412    TGCAACATCCCCGGCACTATTGGCTATTACTGGTGATGTAGCTTCCATTTATGACTAGCT
 110     C  N  I  P  G  T  I  G  Y  Y  W  *
                              ▲

 472    AATAAACTGTCACATACCA
```

Fig. 1. (a, b) Nucleotide sequence of bread wheat cultivar Chinese Spring puroindoline a (a) and (b) coding sequence and upstream flanking region. The deduced amino acid sequence of the proteins is reported. Numeration of primary structure refers to the mature proteins (within arrow heads). Signal and the extra N- and C-terminal cleavable peptides as in Douliez et al. (2000). Putative regulatory elements are indicated (see text for details): bold (TATA-box), bold underlined (CAAT-element), underlined (Polyadenylation site), bold double underlined (Prolamin box), double underlined ($(CA)_n$ element), dashed underlined (dyad repeats), dotted underlined (Amy-box), wave underlined (NtBBF1), boxed (ACGT-element), shaded (E-box), shaded and white character (P-binding site).

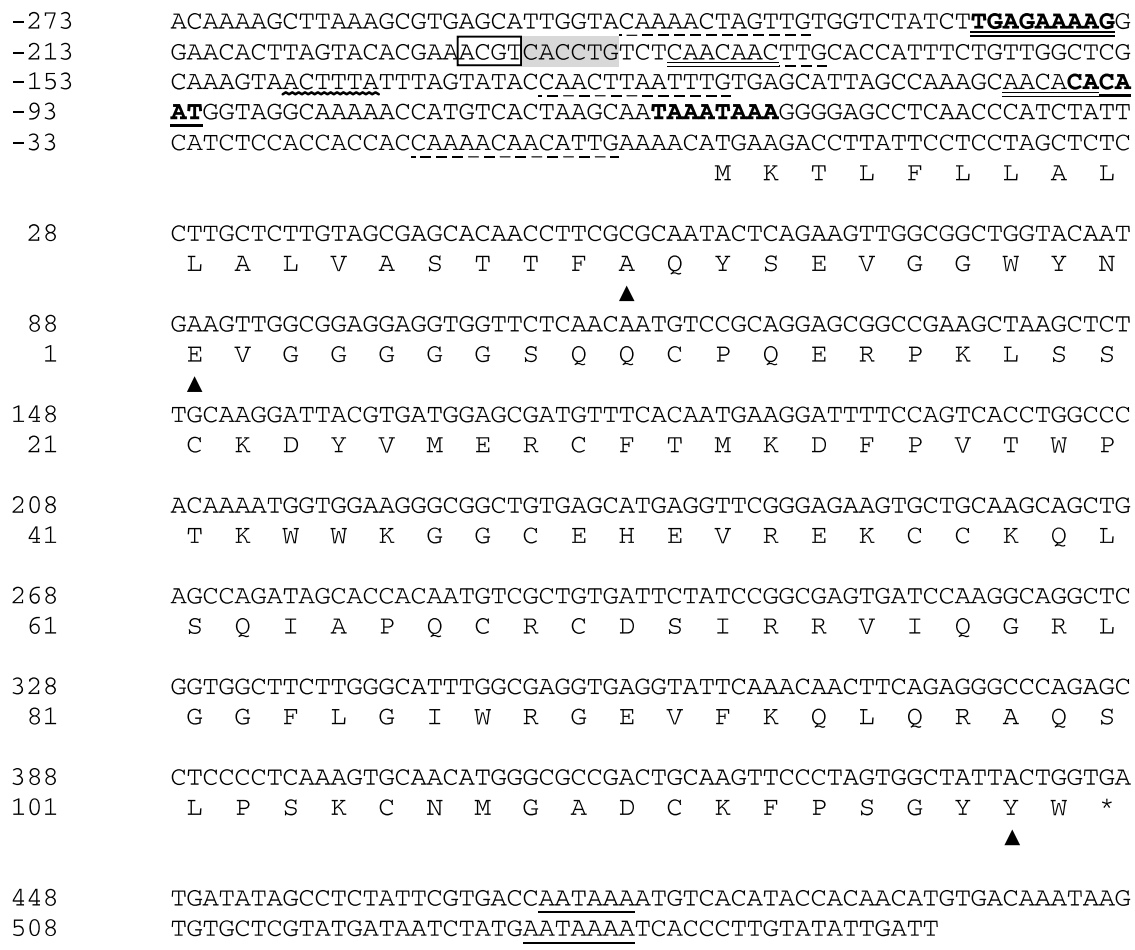

Fig. 1 (*continued*)

(with an intervening GT in all species except *Ae. bicornis*) in the S-genome group, $(GA)_6$ in the D-genome group and $(GA)_{18}$ in hexaploid wheat.

3.5. Coding sequences and deduced protein analysis

No insertions or deletions in the coding region of puroindoline a were observed among any of the taxa examined. However, the S-genome taxa possessed a shorter coding region (444 bp) due to a point mutation at position +447, changing the final codon (TGG) to a stop codon (TGA).

A variable number of single nucleotide polymorphisms (SNPs) was observed. For example, the *T. monococcum* ssp. *aegilopoides* accessions contained no variation, while various degrees of divergence occurred in the *Aegilops* species. Intra-specific sequence identity in the coding regions ranged from 99.3 to 100% (Table 4). Puroindoline a coding region from accession TA1691 shared complete sequence identity with hexaploid wheats. The overall coding sequence identity was quite high (98.4%). Conversely, intra-specific and overall sequence identities were both lower in puroindoline b (98.3–100 and 92.1%, respectively). *Ae. speltoides* ssp. *speltoides* and *Ae. searsii* sequences showed the only deletion found in the puroindoline coding regions, a triplet (AGG) at position +102 in puroindoline b.

All the deduced amino acid sequences used the same reading frame on the basis of the previously-published sequences (see Fig. 1). By means of sequence comparisons among all the 25 genotypes, 22 amino acid changes in puroindoline a and 28 in puroindoline b were identified, reflecting a potential difference in the mutation rate between the two genes. This is in agreement with previous findings that showed a distinct difference in sequence conservation and evolution between puroindoline a and b in a set of *Ae. tauschii* accessions (Massa et al., 2004). A multiple alignment of the puroindoline a and b allelic forms has been published previously (Morris et al., 2001).

Puroindoline a and b preprotein sequences were 148 residues long with the exception of the puroindoline a of the S-genome group, where the tryptophan (W) at position 148 was changed to a stop codon, and the puroindoline b of *Ae. searsii* (both accessions) and *Ae. speltoides* ssp. *speltoides* where either a glycine (G) or a serine (S), respectively, at position seven was absent. The two main features of the puroindoline isoforms, which are the cysteine backbone and the tryptophan-rich domain, were conserved among all the analyzed germplasm. Alignment of the 25 puroindoline a and

Table 5
List of the putative regulator elements investigated for presence and conservation in the promoter sequences of all the analysed Triticeae germplasm

Regulatory element	Consensus sequence	Position		Proposed function	Refs.
		Pin a	Pin b		
$(CA)_n$	CAACAAC	−236/−226 −180/−162	−183/−161[a]	Seed specificity	Morton et al. (1995) Goldberg (1986)
	ACACA	−102, −14	−102[b]		Ellerström et al. (1996)
	CNAACAC	Only in the S diploids	n.f.		
Prolamin box	TG(T/A/C)AAA(A/G)(G/T)	−660/−630 −223/−202	−223/−202	Endosperm specificity	Colot et al. (1987) Thomas and Flavell (1990) Vicente-Carbajosa et al. (1997)
	$CAA(N)_{2-9}TTG$	Recurrent 1–4 times[c]	Recurrent 1–4 times[c]	Endosperm and aleurone specificity	Leah et al. (1994)
NtBBF1	ACTTTA	n.f.	−146/−137	Tissue—specific expression, storage proteins regulation	Baumann et al. (1999) Mena et al. (1998)
Dof core recognition sequence	AAAG	Very frequent	Very frequent	Light regulation, tissue specific gene expression	Vicente-Carbajosa et al. (1997) Yanagisawa and Scheen (1998)
Ry repeat	CATGCA	Not conserved	n.f.	Storage proteins regulation	Morton et al. (1995)
G-box	CACGTG	n.f.	n.f.	Storage proteins regulation	Giuliano et al. (1988)
E-box	CANNTG	Recurrent 2–4 times[c]	−190[d]	Tissue-specificity, developmental control	Stålberg et al. (1996) Kawagoe et al. (1994)
ACGT-box	ACGT	−576/−533 −487[f]	−195[e]	Seed development	Foster et al. (1994) Toyofuku et al. (1998)
Amy-box	TAACA(G/A)A	−680/−637	n.f.	Response to gibberellin	Huang et al. (1990) Gubler et al. (1999)
P-binding site	AACAACC	−630/−616 −300/−280	n.f.	Recognition sequence for the myb-homologous P protein	Grotewold et al. (1994)
AACA-motif	AACNNA	Frequent (5–9 times)	Absent in the SS diploids	Recognition sequence for the rice myb protein	Suzuki et al. (1998)

[a] Not in *Ae. longissima* TA 1921.
[b] Only in the A-and D-genome taxa.
[c] The positions are conserved in every genome group.
[d] Only in the A-and D-genome taxa.
[e] Only in hexaploid wheat and TA 1691.
[f] Only in the D-genome taxa.

puroindoline b consensus sequences of the mature proteins could be accomplished by inserting only three gaps, two in puroindoline a and one in puroindoline b. The sequence of the signal peptides and the N-terminal cleavable peptides (Gautier et al., 1994) follow the consensus: $MKxLFL^{I}/_{L}{}^{G}/_{A}{}^{I}/_{L}$-$LALVAST^{A}/_{T}FA$ and $FAQYSEV^{V}/_{G}G^{W}/_{S}Y^{N}/_{Gap}$, respectively. All deduced amino acid sequences exhibited the typical characteristics of many eukaryotic signal peptides, mainly the lysine residue next to the initiating methionine and a hydrophobic-rich domain, while the additional 10–11 residues are more polar and contain one acidic amino acid.

The puroindoline a amino acid sequences of all members of the A-genome group were virtually identical, the exception being a single amino acid substitution ($^{S}/_{G}$ at position 80 in the mature protein) in all *T. urartu* accessions and *T. monococcum* ssp. *monococcum* accession TA2037. The glycine residue at position 80 was present in the puroindoline a of every diploid examined in this study and in hexaploid wheats. All 11 members of the A-genome group had four conserved amino acid differences when compared to Chinese Spring. Three of the four *Ae. tauschii* accessions shared a single residue substitution ($^{Q}/_{R}$ at position 58), the exception being *Ae. tauschii* ssp. *meyerii* which contained the wild-type arginine, identical to Chinese Spring. The $^{Q}/_{R}$ substitution at position 58 was also observed in each member of the A-genome group. The S-genome group accessions contained the highest intra-group divergence (12 substitutions) and the most differences compared to Chinese Spring (18 substitutions). Each S-genome group accession had a unique puroindoline a sequence, with the exception of *Ae. sharonensis* and *Ae. longissima* ssp. *longissima*, which were identical and shared a single substitution ($^{K}/_{N}$ at position 18) with *Ae. longissima* ssp. *nova*. All members of the S-genome group contained the neutral $^{K}/_{R}$ substitution in the center of the tryptophan-rich domain. Within the S-genome

group, puroindoline a from *Ae. speltoides* ssp. *ligustica* and from the two *Ae. searsii* accessions were the least divergent from Chinese Spring and the most similar to the A-genome group.

Based on the chemical nature of the amino acid residues, the observed substitutions are six neutral and 16 non-neutral. Of the non-neutral substitutions, five are distributed in the signal peptide and cleavable N- and C-terminal peptides, seemingly without affecting their functional properties, and five are located very close to the mature protein termini ($^{G}/_{A}$ at position 7, $^{N}/_{D}$, $^{P}/_{R}$, $^{G}/_{S}$ and $^{I}/_{S}$ in the last 6 residues). The remaining non-neutral substitutions are concentrated in a stretch at positions 16-19 ($^{K}/_{E}L^{N}/_{K}^{S}/_{A}$), and at positions 49 ($^{L}/_{Q}$), 58 ($^{Q}/_{R}$) and 80 ($^{S}/_{G}$). Only one neutral replacement is present in a long amino acid stretch between positions 20 and 48, where the tryptophan rich domain is located (residues 38–47). Based on the primary structure of the mature proteins, the hydropathicity graphics of the different puroindoline a alleles did not show any appreciable variation when compared to the wild type. The protein pI was calculated between 9.93 (*Ae. bicornis*) and 10.46 (*Ae. longissima*), the total charge (at pH 7.00) was between 4.00 (A-and D- genome groups, except ssp. *meyerii*) and 6.00 (*Ae. longissima*). Soft wheat Chinese Spring showed values of 10.27 and 5.00, respectively.

Concerning puroindoline b, of the four *T. urartu* accessions, only TA763 differed from the others with a substitution ($^{R}/_{L}$ at position 68 in the mature protein). *T. monococcum* ssp. *aegilopoides* could be differentiated from the other A-genome group accessions by a single substitution $^{A}/_{V}$ at position 91. All A-genome group accessions shared 10 conserved amino acid substitutions when compared to Chinese Spring. These substitutions may be the more ancestral sequence as they are present in most, if not all, of the other diploid taxa in the present study. For example, the substitution $^{K}/_{R}$ at position 79 is shared by all the diploids and the substitution $^{I}/_{L}$ at position 95 is shared by 24, the exception being *Ae. speltoides* ssp. *speltoides* (TA1789). The deduced amino acid sequences of the *Ae. tauschii* accessions were very similar to one another, only ssp. *meyerii* differed from the other three accessions. Eight sequence differences were shared by all four lines compared to Chinese Spring. The three *Ae. tauschii* accessions (TA1704, TA2381 and TA10) containing identical puroindoline a further differed from the hexaploid wheat by six more substitutions. Compared to Chinese Spring, *Ae. tauschii* ssp. *meyerii* differed the least of the *Ae. tauschii* accessions, with a total of nine substitutions, of which six were conserved among all 15 A- and D-genome group accessions, two were conserved among the other *Ae. tauschii* accessions and only one, a $^{S}/_{R}$ substitution at position 19, was unique.

Compared with the others, the S-genome group showed numerous unique substitutions. For example, the substitution of $^{T}/_{S}$ and $^{A}/_{G}$ in the signal peptide, $^{G}/_{S}$ at position 19, $^{L}/_{M}$ at position 26, $^{LH}/_{IR}$ at positions 72–73 and $^{I}/_{S}$ at position 115. No amino acid changes were found in any diploid sequence that matched the hardness mutations that have been identified in cultivated *T. aestivum*. The 28 substitutions can be classified as 11 neutral and 17 non-neutral. One of the non-neutral substitutions was located in the signal peptide and three were very close to the N- and C-termini ($^{G}/_{A}$ at position 4, $^{G}/_{S}$ at position 7, and $^{S}/_{I}$ at position 115). The remaining substitutions were distributed throughout the mature protein, with the exception of the region between position 29 and 54 where the 'hard' mutations of *T. aestivum* occur. This is the region (residues 39–47) where the Trp-rich domain is located, with part of the neighboring helices. The hydrophobicity of the puroindoline b proteins was generally mantained across all taxa, with only *T. urartu* accession TA763 being sensibly more hydrophobic in a stretch comprised between residue 68 and 73. On average, the pI values of the proteins ranged between 10.2 and 10.3 and the total charge was between 6.09 and 7.09. Only *T. urartu* TA763 was sensibly different, with values of 10.7 and 5.09, respectively. Soft wheat Chinese Spring showed values of 10.5 and 8.09, and the most similar was *Ae. tauschii* ssp. *meyerii* with values of 10.4 and 8.09.

4. Discussion

In-depth studies of the puroindoline b promoter region have been conducted by transformation experiments in rice (Digeon et al., 1999). Differential expression patterns of the full-length promoter sequence (1068 bp) from a soft *T. aestivum* wheat cultivar and various portions of it, fused with *UidA* gene reporter system, suggested that the sequence between −388 and −210 bp upstream of the start codon contains *cis*-acting elements necessary for endosperm-specific expression, and the sequence between −210 and −124 bp is critical for the specific expression in the epithelium of the scutellum of the transformed developing rice kernels. Lillemo et al. (2002) analysed the promoter regions of both puroindolines from a small set of wild diploid wheats and identified some candidate elements for the regulation of these genes, most of which are located in the region indicated by Digeon et al. (1999). In our study, the occurrence of these and new candidate sequences for putative regulation of both puroindoline genes is verified in a large germplasm analysis, including soft and hard hexaploid wheats, and closely related wild diploid wheats. The high rate of conservation of some regulatory elements in many different taxa of the Triticeae tribe reduces the possibility that their occurrence is simply random.

Among the identified regulatory elements, the importance of the prolamin-box and the $(CA)_n$-box seems predominant, as both are highly conserved in position and sequence in all the analyzed germplasm. In wheat storage proteins, the prolamin-box is usually located around position −300 bp from the start codon of prolamin genes. However, no direct influence on gene regulation by the prolamin-box has been observed: the region most involved occurs between −170 and −130 bp from the start codon (Thomas and Flavell, 1990). Nevertheless, the strong conservation of the prolamin-box in the promoter region of all storage proteins of cereals suggests some particular function that requires further clarification. If the prolamin-box does play a role in the regulation of genes for endosperm proteins, it is likely to do so in combination with other elements. For example, many enhancer elements are composed

of multiple short sequence motifs that bind combinations of transcription factors to confer inducibility, tissue specificity and general enhancement on promoters. In this context, the prolamin-box may act synergistically with neighboring sequences, although not in an essential way. Alternatively, it might even represent a kind of a 'first' signal for the general regulation of the genes for endosperm-localized proteins. However, in the promoter of the puroindoline genes, the prolamin-box is located in the region that has been identified to be necessary for endosperm deposition, and several downstream motifs are recurrent and highly conserved.

The $(CA)_n$-box is known to promote strong seed specificity in many dicotyledon and monocotyledon plants, in association with additional cis-acting elements. The $(CA)_n$-box is usually present in different forms, CAACAAC (Morton et al., 1995), AACACA (Goldberg, 1986) and CNAACAC (Ellerström et al., 1996). All of these $(CA)_n$-box sequences partially match other core motifs, such as the P-binding site (Grotewold et al., 1994), the recognition sequence for the rice Myb protein (a tissue-specific transcription factor), and the AACA-motif of rice glutelin genes (Takaiwa et al., 1991; Zheng et al., 1993). The AACA-motif has been shown to be a negative regulator in non-seed tissues and to act in combination with the GCN4 element to direct endosperm-specific expression in the seed (Takaiwa et al., 1996). The AACA-motif has a core consensus sequence AACNNA but the sequence first proposed after an extensive analysis of rice glutelin genes is $^{T}/_{C}$AACAAAC$^{T}/_{A}$ $^{C}/_{A}$TAT, which in part resembles the Amy-box TAACAAA, known as activator of gibberellin-regulated genes in the aleurone cells of rice and barley. Of all the $(CA)_n$-box motifs, only the form CAACAAC is perfectly conserved in both puroindoline promoters. Several conserved locations for the AACA-motif were identified in this research, but they were limited to its proposed consensus sequence AACNNA. The GCN4 element is also absent and the P-binding box and the Amy-box are present and conserved only in puroindoline a.

On the other hand, the dyad sequence CAA$(N)_{2-9}$TTG (Leah et al., 1994) was reiterated 1 to 4 times in the different genome groups but was always conserved in sequence and position. Similar dyad sequences are present in the promoters of other seed-specific proteins, such as wheat glutenins (located in the endosperm), barley chitinase and LTP-1 (located in the aleurone), and barley *alpha*-amylase inhibitor (located in both).

The prevalence and conservation of seed-specific regulatory elements in the promoter regions of the puroindolines is in agreement with the available immuno-staining data, although contrasting results locate both puroindolines in the starchy endosperm and only puroindoline b in the aleurone layer (Dubreil et al., 1994), whereas more recent data do not corroborate these assumptions by showing the two isoforms being instead abundant in both compartments (Capparelli et al., 2005). The expression levels of puroindolines during seed development were first studied by Gautier et al. (1994) who showed that mRNAs began accumulating between 8 and 12 days after anthesis (DAA), reached a peak between 26 and 36 DAA, and then declined rapidly. This body of information suggests that the two promoters must have most of their regulatory elements in common. Although the two promoter regions analyzed differ in length, the only difference between the regulatory elements of the puroindoline promoters observed in this research, relative to the motif sequence data available in the literature, concern their internal order, the frequency of their occurrence, the exclusive presence of the Amy-box and the P-Binding site in the upstream regions of puroindoline a, and of the NtBBF1 element in puroindoline b. It is worth considering, however, that the Amy-box and the P-Binding site do not occur in the full-length puroindoline b promoters of hexaploid wheat cultivar Andain (Digeon et al., 1999) and barley (Darlington et al., 2001).

The core motif of the Amy-box is a well-characterized gibberellin responsive element, and activated by the GAMYB transcription factors usually expressed in cereal aleurone cells (Gubler et al., 1999). The ACCCTA target sequence of the NtBBF1 DOF protein, a plant transcriptional factor active in the developing endosperm of barley and maize grains (Mena et al., 1998) was shown to be necessary for tissue-specific expression in transgenic tobacco, and capable to mediate auxin response (Baumann et al., 1999). However, a putative different hormonal control for the two genes is yet to be shown and it is unlikely to be the only factor responsible for the slightly different local distributions, if any, of the two puroindoline proteins in wheat grain.

Additional putative elements correlated with transcription, and possibly acting as enhancers (Muller et al., 1988) are obviously present; this research identified at least three. Among these, the most interesting is the DOF core recognition sequence, for which interactions have been supposed to occur in conjunction with the Prolamin-box and specific nuclear factors (Vicente-Carbajosa et al., 1997). For instance, a maize member of the large family of zinc-finger DNA binding proteins, named PBF, is able to interact with the basic leucine zipper protein Opaque2, a known transcriptional activator of zein genes whose target lies 20 bp downstream of the Prolamin-box in the zein promoter (Vicente-Carbajosa et al., 1997). Some other enhancer target sequences may be represented by the E-box, specifically for helix–loop–helix proteins and by the ACGT-box for b-zip proteins. However, little is known about the biological role of these latter elements in plants. It is remarkable, though, that we found the same pattern of elements conserved in sequence, position and redundancy, in the barley promoter regions described by Darlington et al. (2001), with special concern to the Prolamin box, the $(CA)_n$ elements, the CA$(N)_{2-9}$TTG motif, the NtBBF1 domain and the DOF core recognition sequence. Lillemo et al. (2002) highlighted the low probability for a multiple random occurrence of these regulatory motifs in a relatively short sequence; as such, the variation in the presence and frequency of the regulatory elements in the different genome groups might be indicative of different promoter strengths. As well, differences between puroindoline a and b promoters may account for the spatial and temporal regulation of the two isoforms, which are yet to be cleared. The promoters of the A-genome group contained every investigated putative element but with the lowest frequencies, relative to those

present at multiple locations. The promoters of the S-genome group were characterized by the highest variability in presence and frequency of the elements. The promoters of the D-genome group share with cultivated *T. aestivum* the same regulatory elements at the highest frequencies.

The promoters described in this work may help in elucidating the mechanism of specificity that characterizes the puroindoline proteins; such a source of natural variation could be used in transformation experiments to confirm the importance of sequence and location of these and other motifs in conferring spatial and temporal regulation and possibly modulating gene transcription. In this context, the specific promoters identified here could replace the commonly-used constitutive promoters, or the tissue-specific ones from gene systems other than puroindolines, in genetic transformation experiments aiming at modifying kernel texture of crops (Giroux et al., 2003). Finally, further promoter studies may help elucidate the *in vivo* function of these proteins, with a special regard to their involvement in defense (plant–pathogen interactions).

Concerning the new puroindoline alleles presented here, a large number of the mutations we found implied neutral residues substitutions. On the basis of the high conservation of every mutation, which is always a characteristic of entire genome groups, and the acknowledged soft phenotype of the wild diploid wheats (Williams, 1986), we can assume that the effects of the non-neutral variation are either balanced by other subsequent mutations or the soft phenotype is conferred by the maintained integrity of other protein domains. Clearly, expression studies of new alleles, from wild wheat relatives would greatly enlarge our knowledge on the functional behaviour of the puroindoline system. Indeed, the integrity of the Trp-rich domain is essential for the interaction with lipids. Other important features are the positively charged residues, the hydrophobic stretches and the residues involved in loops and turns of the mature proteins, some of which displayed some non-neutral variation. Thus, within the kernel 'softness' range, we might assume a 'gradient' of protein activity across the wheat-related diploid species. Indeed, if there appears to exist different puroindoline soft alleles providing quantifiable differences in the technological utilisation of wheat grain and flour (Gedye et al., 2004) it would be an interesting matter of concern for future wheat breeding programs.

Acknowledgements

The helpful assistance of Jon Raupp and Duane Wilson of the KSU WGRC, Morten Lillemo, currently with CIMMYT, and Camille Steber, USDA ARS, are thankfully acknowledged.

References

Baumann, K., De Paolis, A., Costantino, P., Gualberti, G., 1999. The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the *rolB* oncogene in plants. Plant Cell 11, 323–333.

Blochet, J.E., Kaboulou, A., Compoint, J.P., Marion, D., 1991. Amphiphillic proteins from wheat flour: specific extraction, structure and lipid binding properties. In: Bushuk, W., Tkachuk, R. (Eds.), Gluten Proteins 1990. American Association of Cereal Chemists, St Paul, MN, pp. 314–325.

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

Capparelli, R., Borriello, G., Giroux, M.J., Amoroso, M.G., 2003. Puroindoline-A gene expression is involved in association of puroindolines to starch. Theoretical and Applied Genetics 107, 1463–1468.

Capparelli, R., Amoroso, M.G., Palumbo, D., Iannaccone, M., Faleri, C., Cresti, M., 2005. Two plant puroindolines colocalize in wheat seed and in vitro synergistically fight against pathogens. Plant Molecular Biology 58, 857–867.

Chantret, N., Cenci, A., Sabot, F., Anderson, O., Dubcovsky, J., 2004. Sequencing of the *Triticum monococcum* Hardness locus reveals good microcolinearity with rice. Molecular Genetics and 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, M.F., Cattolico, L., Beckert, M., Aubourg, S., Weissembach, J., Caboche, M., Bernard, M., Leroy, P., Chalhoud, B., 2005. Molecular basis of evolutionary events that shaped the Hardness locus in diploid and poliploid wheat species (*Triticum* and *Aegilops*). Plant Cell 17, 1033–1045.

Charnet, P., Molle, G., Marion, D., Rousset, M., Lullien-Pellerin, V., 2003. Puroindolines form ion channels in biological membranes. Biophysical Journal 84, 2416–2426.

Chen, F., He, Z., Xia, X., Lillemo, M., Morris, C.F., 2005. A new puroindoline b mutation present in Chinese winter wheat cultivar Jingdong 11. Journal of Cereal Science 42, 267–269.

Colot, V., Robert, L.S., Kavanagh, T.A., Bevan, M.W., Thompson, R.D., 1987. Localization of sequences in wheat endosperm protein genes that confer tissue-specific expression in tobacco. EMBO Journal 6, 3559–3564.

Corona, V., Gazza, L., Boggini, G., Pogna, N.E., 2001. Variation in friabilin composition as determined by A-PAGE fractionation and PCR amplification, and its relationship to grain hardness in bread wheat. Journal of Cereal Science 34, 243–250.

Darlington, H.F., Rouster, J., Hoffmann, L., Halford, N.G., Shewry, P.R., Simpson, D.J., 2001. Identification and molecular characterisation of hordoindolines from barley grain. Plant Molecular Biology 47, 785–794.

Dellaporta, S.L., Wood, J., Hicks, J.B., 1983. A plant DNA minipreparation: version II. Plant Molecular Biology Reports 1, 19–21.

Digeon, J.-F., Guiderdoni, E., Alary, R., Micheaux-Ferriere, N., Joudrier, P., Gautier, M.-F., 1999. Cloning of a wheat puroindoline gene promoter by IPCR and analysis of promoter regions required for tissue-specific expression in transgenic rice seeds. Plant Molecular Biology 39, 1101–1112.

Douliez, J.-P., Michon, T., Elmorhani, K., Marion, D., 2000. Structure, biological and technological functions of lipid transfer proteins and indolines, the major lipid binding proteins from cereal kernels. Journal of Cereal Science 32, 1–20.

Dubreil, L., Quillien, L., Legoux, M.-A., Compoint, J.-P., Marion, D. 1994. Variability and localization of wheat kernel indolines and lipid transfer proteins. In: Wheat Kernel Proteins-Molecular and Functional Aspects. Università Degli Studi Della Tuscia, S. Martino Cimino, Viterbo (Italy), September 28–30, 1994, pp 331–333.

Dubreil, L., Gaborit, T., Bouchet, B., Gallant, D.J., Broekaert, W.F., Quillien, L., Marion, D., 1998. Spatial and temporal distribution of the major isoforms of puroindolines (puroindoline a and puroindoline b) and the non-specific lipid binding protein (ns-LTP1e$_1$) of *Triticum aestivum* seeds. Relationships with their in vitro antifungal properties. Plant Science 138, 121–135.

Ellerström, M., Stålberg, K., Ezcurra, I., Rask, L., 1996. Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. Plant Molecular Biology 32, 1019–1027.

Foster, R., Izawa, T., Chua, N.-H., 1994. Plant bZIP proteins gather at ACGT-element. FASEB Journal 8, 192–200.

Gautier, M.-F., Aleman, M.-E., Guirao, A., Marion, D., Joudrier, P., 1994. *Triticum aestivum* puroindolines, two basic cysteine rich proteins: cDNA sequence analysis and developmental gene expression. Plant Molecular Biology 25, 43–57.

Gautier, M.-F., 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 Science 153, 81–91.

Gazza, L., Nocente, F., Ng, P.K.W., Pogna, N.E., 2005. Genetic and biochemical analysis of common wheat cultivars lacking puroindoline a. Theoretical and Applied Genetics 110, 470–478.

Gedye, K.R., Morris, C.F., Bettge, A.D., 2004. Determination and evaluation of the sequence and textural effects of puroindoline a and puroindoline b genes in a population of synthetic hexaploid wheat. Theoretical and Applied Genetics 109, 1597–1603.

Giroux, M.J., Morris, C.F., 1997. A glycine to serine change in puroindoline-b is associated with wheat grain hardness and low level of starch-surface friabilin. Theoretical and Applied Genetics 95, 857–864.

Giroux, M.J., Sripo, T., Gerhardt, S., Sherwood, J., 2003. Puroindolines: their role in grain hardness and plant defense. Biotechnology and Genetic Engineering Reviews 20, 277–290.

Giuliano, G., Pichersky, E., Malik, V.S., Timko, M.P., Scolnik, P.A., Casmore, A.R., 1988. An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. Proceedings of the National Academy of Science of USA 85, 7089–7093.

Goldberg, R.B., 1986. Regulation of plant gene expression. Philosophical Transactions of the Royal Society of London 314, 343–353.

Greenwell, P., Schofield, J.D., 1986. A starch granule protein associated with endosperm softness in wheat. Cereal Chemistry 63, 379–380.

Grotewold, E., Drummond, B.J., Bowen, B., Peterson, T., 1994. The myb-homologus P gene controls phobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell 76, 543–553.

Gubler, F., Raventos, D., Keys, M., Watts, R., Mundy, J., Jacobsen, J.V., 1999. Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant Journal 17, 1–9.

Higo, K., Ugawa, Y., Iwamoto, M., Korenaga, T., 1999. Plant cis-acting regulatory elements (PLACE) database: 1999. Nucleic Acids Research 27, 297–300.

Huang, N., Sutliff, T.D., Litts, J.C., Rodriguez, R.L., 1990. Classification and characterization of the rice α-amylase multigene family. Plant Molecular Biology 14, 655–668.

Igresias, G., Gaborit, T., Oury, F.X., Chiron, H., Marion, D., Branlard, G., 2001. Genetic and environmental effects on puroindoline a and puroindoline b content and their relationships to technological parameters in French bread wheats. Journal of Cereal Science 34, 37–47.

Jolly, C., 1991. The biochemistry and molecular genetics of grain softness and hardness in wheat, *Triticum aestivum*. PhD Dissertation, Macquarie University, Sydney, Australia.

Kawagoe, Y., Campbell, B.R., Murai, N., 1994. Synergism between CACGTG (G-BOX) and CACCTG Cis-elements is required for activation of the bean seed storage protein β-phaseolin gene. Plant Journal 5, 885–890.

Krishnamurthy, K., Balconi, C., Sherwood, J.E., Giroux, M.J., 2001. Wheat puroindolines enhance fungal disease resistance in transgenic rice. Molecular Plant and Microbe Interactions 14, 1255–1260.

Leah, R., Skriver, K., Knudsen, S., Ruud-Hansen, J., Raikel, N.V., Mundy, J., 1994. Identification of an enhancer/silencer sequence directing the aleurone-specific expression of a barley chitinase gene. Plant Journal 6, 579–589.

Lillemo, L., Simeone, M.C., Morris, C.F., 2002. Analysis of puroindoline a and b promoter sequences from *Triticum aestivum* cv. Penawawa and diploid wheat ancestors. Euphytica 126, 321–331.

Massa, A.N., Morris, C.F., Gill, B.S., 2004. Sequence diversity of puroindoline-a (*Pina-D1*), puroindoline-b (*Pinb-D1*) and the grain softness protein (*Gsp-1*) genes in *Aegilops tauschii* Coss. Crop Science 44, 1808–1816.

McIntosh, R.A., Yamazaki, Y., Devos, K.M., Dubcovsky, J., Rogers, W.J., Appels, R., 2003. Catalogue of gene symbols for wheat. Proceedings of the 10th International Wheat Genetics Symposium, Paestum, Italy, 1–6 September 2003.

Mena, M., Vicente-Carbajosa, J., Schmidt, R., Carbonero, P., 1998. An endosperm specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin box of a native B-Hordein promoter in barley endosperm. Plant Journal 16, 53–62.

Morris, C.F., 2002. Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Molecular Biology 48, 633–647.

Morris, C.F., Simeone, M.C., Gill, B.S., Mason-Gamer R.J., Lillemo, M., 2001. Comparison of puroindoline sequences from various diploid members of the Triticeae and modern cultivated hexaploid wheats. In: Wootton M., Batey I.L., Wrigley C.W. (Eds.) Cereals 2000. Proceedings of the 11th ICC Cereal and Bread Congress and 50th Australian Cereal Chemistry Conference, Surfers Paradise, Qld, Australia, 8–18 September 2000, R.A.C.I., North Melbourne, Australia, pp. 678–681.

Morton, R.L., Quiggin, D., Higgins, T.J.V., 1995. Regulation of seed storage proteins gene expression. In: Kigel, J., Galili, G., Dekker, M. (Eds.). Seed Development and Germination. New York, pp. 103–136.

Muller, M.M., Gerster, T., Schaffner, W., 1988. Enhancer sequences and the regulation of gene transcription. European Journal of Biochemistry 176, 485–495.

Ram, S., Jain, N., Shoran, J., Singh, R., 2005. New frame shift mutation in puroindoline b in Indian wheat cultivars Hyb 65 and NI5439. Journal of Plant Biochemistry and Biotechnology 14, 45–48.

Simeone, M.C., Lafiandra, D., 2005. Isolation and characterisation of friabilin genes in rye. Journal of Cereal Science 41, 115–122.

Stålberg, K., Ellerström, M., Ezcurra, I., Ablov, S., Rask, L., 1996. Disruption of an overlapping E-Box/ABRE motif abolished high transcription of the napA storage protein promoter in transgenic *Brassica napus* seeds. Planta 199, 515–519.

Suzuki, A., Wu, C.Y., Washida, H., Takaiwa, F., 1998. Rice MYB protein OSMYB5 specifically binds to the AACA motif conserved among promoters of genes for storage protein glutelin. Plant Cell Physiology 39, 555–559.

Takaiwa, F., Oono, K., Kato, A., 1991. Analysis of the 5′ flanking region responsible for the endosperm-specific expression of a rice glutelin chimeric gene in transgenic tobacco. Plant Molecular Biology 16, 49–58.

Takaiwa, F., Yamanouchi, U., Yoshihara, T., Washida, H., Tanabe, F., Kato, A., Yamada, K., 1996. Characterization of common *cis*-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Plant Molecular Biology 30, 1207–1221.

Thomas, M.S., Flavell, R.B., 1990. Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell 2, 1171–1180.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.

Toyofuku, K., Umemura, T., Yamaguchi, J., 1998. Promoter elements required for sugar repression of the *Ramy3D* gene for α-amilase in rice. FEBS Letters 428, 275–280.

Tranquilli, G., Heaton, J., Chicaiza, O., Dubcovsky, J., 2002. Substitutions and deletions of genes related to grain hardness in wheat and their effect on grain texture. Crop Science 42, 1812–1817.

Turnbull, K.-M., Turner, M., Mukai, Y., Yamamoto, M., Morrell, M.K., Appels, R., Rahman, S., 2003. The organization of genes tightly linked to the *Ha* locus in *Aegilops tauschii*, the D-genome donor to wheat. Genome 46, 330–338.

van Slageren, M.W., 1994. Wild Wheats: A Monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (Poaceae): A Revision of all Taxa Closely Related to Wheat, Excluding wild *Triticum* Species, with Notes on Other Genera and Tribe Triticeae, Especially *Triticum*. ICARDA/Wageningen Agricultural University, The Netherlands. Papers 94, pp. 512.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L., Schmidt, R.J., 1997. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. Proceedings of the National Academy of Science of USA 94 7685–7690.

Williams, P.C., 1986. The influence of chromosome number and species on wheat hardness. Cereal Chemistry 63, 56–57.

Xia, L., Chen, F., He, Z., Chen, X., Morris, C.F., 2005. Occurrence of puroindoline alleles in Chinese winter wheats. Cereal Chemistry 82, 38–43.

Yanagisawa, S., Sheen, J., 1998. Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. Plant Cell 10, 75–89.

Zheng, Z., Kawagoe, Y., Xiao, S., Li, Z., Okita, T., Hau, T., Lin, A., Murai, N., 1993. 5′ distal and proximal *cis*-acting regulator elements are required for developmental control of a rice seed storage protein glutelin gene. The Plant Journal 4, 357–366.