Molecular characterization and diversity of the *Pina* and *Pinb* genes in cultivated and wild diploid wheat

C. Guzmán*, L. Caballero[1], M.A. Martín and J.B. Alvarez

Departamento de Genética, Escuela Técnica Superior de Ingeniería Agronómica y de Montes, Edificio Gregor Mendel, Campus de Rabanales, Universidad de Córdoba, ES-14071 Córdoba, Spain.

[1] *Present address:* Megaseed S.A. Estrada 624 (5800) Río Cuarto. Provincia de Córdoba. República Argentina.

* Author for correspondence (Phone: +34-957212575; fax: +34-957218503; E-mail: ge2gugac@uco.es)

For Molecular Breeding

Abstract

Grain hardness is one of the most important characteristics of wheat quality. Soft endosperm is associated with the presence of two proteins in the wild form: Puroindoline a and b. The puroindoline genes and their derived proteins are present in the putative wheat diploid ancestors which are thought to be the donors of the A, B and D genomes in common and durum wheat. In this study, we investigated the variability of grain hardness in einkorn, along with the nucleotide diversity of *Pina* and *Pinb* genes in a collection of einkorn wheat and *T. urartu*, in addition to the study of the neutrality and linkage disequilibrium of these genes. Various alleles were detected for *Pina* and *Pinb* genes including three novel alleles for the *Pinb* locus: *Pinb-A^m1i*, *Pinb-A^m1j* and *Pinb-A^m1k*. Some differences were found in grain hardness between the different genotypes. The neutrality test showed a different pattern of variation between the two *Pin* genes. The genetic analysis of a diploid wheat collection has demonstrated that these species are a potential source of novel puroindoline variants. Our data suggest that, although further studies must be carried out, these variants could be used to expand the range of grain texture in durum and common wheat, which would permit the development of new materials adapted to novel uses in the baking and pasta industry.

Keywords: einkorn wheat, grain hardness, kernel texture, puroindolines.

Introduction

The texture of the kernel is one of the most important characteristics of quality in wheat. Common wheat (*Triticum aestivum* L. ssp. *aestivum*; 2n = 6× = 42, **AABBDD**) is classified as hard or soft according to this trait and this determines the end use of the grain and affects both milling and bread-making (Pomeranz and Williams 1990). In general, hard wheat is used for making bread whereas soft wheat is used in the manufacture of cookies, cakes and pastries (Morris and Rose, 1996). The genetics of this characteristic has been widely studied (Bhave and Morris, 2008a, b; Morris and Bhave 2008) and has been found to be under the control of the *Ha* locus which is located at the extreme distal end of chromosome 5D (Law et al. 1978). This locus has a complex structure within a region of about 82,000 bp (Chantret et al. 2005), which includes the puroindoline genes (*Pina* and *Pinb*) together with the "grain softness protein" gene (*Gsp*). Puroindolines are basic cysteine-rich proteins with a characteristic tryptophan-rich domain which makes them unique among plant proteins and the amino acid sequences of both puroindolines (PIN-a and PIN-b) show up to a 60% similarity in common wheat (Gautier et al. 1994). In common wheat the texture of the soft endosperm is associated with the wild-type allelic forms of *Pina* and *Pinb*, whereas in hard wheat the texture of the endosperm is associated with mutations in either of these two genes. These mutations which consist of *null* mutations or modifications of one or more amino acids, almost always lead to a harder phenotype. In addition, the puroindolines have been associated with plant resistance to pathogens (Dubreil et al. 1998; Krishnamurthy et al. 2001)

Diverse studies have suggested that both puroindolines are absent in durum wheat (*T. turgidum* L. ssp. *durum* Desf. em. Husn.; 2n = 4× = 28, **AABB**) and that this is the reason for their very hard endosperm phenotype. Various authors have suggested that these genes were deleted from chromosome 5A and 5B in this type of wheat after hybridization between the diploid ancestors (Gautier et al. 2000; Chantret et al. 2005). Conversely, puroindolines genes and their derived proteins are present in the putative diploid ancestors of wheat that are thought to be the donors of the A, B and D genomes in common and durum wheat.

The B and D genomes have been associated with species of the *Aegilops* genus, whereas the A genome is related to diploid wheat species (2n = 2× = 14, **AA**). The latter consist mainly of *T. monococcum* ssp. *aegilopoides* Link em. Thell. (syn. *T. boeoticum* Boiss.; $\mathbf{A^mA^m}$); einkorn (*T. monococcum* L. ssp. *monococcum*; $\mathbf{A^mA^m}$) and *T. urartu* Thum. *ex* Gandil ($\mathbf{A^uA^u}$), of which only einkorn was cultivated in antiquity while the others are wild types. Nowadays, the most accepted theory is that *T. urartu* was the donor of the A genome in polyploid wheat (Dvorak et al. 1993). These species have been shown to be an excellent source of useful traits for breeding programs designed to improve the characteristics of durum and common wheat (Sharma et al. 1981). With this in mind, our group has carried out diverse studies on aspects related to the bread-making quality of these species paying particular attention to the variation in the endosperm storage proteins (Alvarez et al. 2006; Caballero et al. 2008; Martín et al. 2008) and waxy proteins (Guzman et al. 2009). Other authors have evaluated the diversity of puroindolines in these species and other ancestors of diploid wheat (Massa et al. 2004, 2006; Simeone et al. 2006; Chen et al. 2009) in which various novel alleles have been identified, although their effect on kernel texture has not been tested.

The main aim of this study was to examine the variability of grain hardness in einkorn, along with the nucleotide diversity of the *Pina* and *Pinb* genes in a collection of einkorn wheat and *T. urartu*, in addition to investigating the neutrality and linkage disequilibrium of these genes.

Materials and Methods

Plant material

Twenty one lines of einkorn, together with seven lines of *T. urartu*, were used in this study (Table 1). These lines derived from an equal number of accessions obtained from three Germplasm Banks: the accessions coded as PI were of the National Small Grain Collection (Aberdeen, USA), the BGE accessions were the Centro de Recursos Fitogenéticos-INIA (Alcalá de Henares, Spain), and the TRI accession from the Institute for Plant Genetics and Crop Plant Research (Gatersleben, Germany). These materials were grown during 2009-10 using standard agronomic practices for the region (175 Kg/ha N, 90 Kg/ha P, and 90 Kg/ha K) at the CIFA-IFAPA experimental station in Córdoba, Spain.

Grain hardness

The grain hardness in einkorn lines was assessed by the Particle Size Index - PSI - (Williams et al. 1986). Three grams of flour were weighted exactly and transferred onto a 200 mesh sieve where the flour was shaken for three minutes. The PSI was calculated as the percentage of flour that passed through the sieve.

DNA extraction and PCR amplification

For DNA extraction, approximately 100 mg of young leaf tissue was excised, immediately frozen in liquid nitrogen and stored at -80°C. DNA was isolated using the DNAzol® method (Invitrogen, Carlsband, CA, USA).

The primers designed by Massa et al. (2004) were used for specific amplification of *Pina* or *Pinb* genes. Each 15-μl reaction included 50 ng DNA, 1.5mM $MgCl_2$, 0.2 μM of each primer, 0.2 mM dNTPs, 1.5 μl 10x PCR Buffer and 0.75U DNA polymerase (Promega, Madison, WI, USA). The PCR conditions included an initial denaturation step of 3 min at 94°C followed by 35 cycles as follows: 45 s at 94°C, 30 s at 64°C then 45s at 72°C. After the 35 cycles a final extension of 5 min at 72°C was done. The PCR products of both *Pina* and *Pinb* genes were digested with endonucleases restriction enzymes *DdeI*, *BsmAI* and *NciI*. Both amplification and digested products were fractionated in vertical PAGE gels (T: 12%, C: 1.28%) and the bands were visualized by ethidium bromide staining.

Cloning of PCR products and sequencing analysis

PCR products were excised from polyacrylamide gel and cloned into pGEM T-easy vector (Promega, Madison, WI, USA) for sequencing. Three positive clones for each PCR product were sequenced in both directions with universal primers using an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The sequences reported in the current study were deposited in the Genbank database under accessions number HQ696584 to HQ696594 and compared to other sequences available in the databases by BLAST analysis. Alignments between the sequences found and cv. Chinese Spring (X69914, X69912) were done using Geneious Pro ver. 4.7.6 software (Biomatters Ltd.).

Data Analysis

The PSI values were analysed by ANOVA, and means by *Pina*/*Pinb* composition were compared using least significant difference (LSD) test.

DNA sequences diversity analysis was conducted by DnaSp version 5.0 (Librado and Rozas 2009). Different parameters were calculated as the number of polymorphic sites (θ), and the average number of nucleotide differences per site between two sequences (π), both of them as a measure of nucleotide diversity. The Tajima´s (*D*) and Fu and Li´s (*D** and *F**) statistics were used to perform tests of neutrality (Tajima 1989; Fu and Li 1993). To test for intra- and interlocus linkage disequilibrium the degree of LD between segregating sites was estimated by the Fisher exact test (Sokal and Rohlf 1969) and Bonferroni procedure (Weir 1996) as implemented by DnaSP v 5.0.

Results and discussion

Amplification and sequencing of puroindoline genes

As a first approach to the analysis of the variability of both *Pin* genes, they were amplified using the gene-specific primers designed by Massa et al. (2004). A unique band of the expected size (516-bp for *Pina* and 529-bp for *Pinb*) was detected in all samples (Fig. 1a). In both cases, a certain degree of polymorphism was found. Some amplicons showed different electrophoretical mobility in the *Pina* (Fig. 1a, lanes 2, 8 10) and *Pinb* genes (Fig. 1b, lanes 5, 6, 9 and 10) that could be due to different DNA conformation. This suggests the presence of at least two alleles for each *Pin* gene. In order to detect internal variability, the amplicons of both *Pina* and *Pinb* genes were digested with endonuclease restriction enzymes *DdeI* and *BsmAI/NciI* respectively, (Fig. 1b). The choice of these enzymes was based on the fact that there is only one target sequence inside *Pin* genes as indicated in other published sequences of einkorn and *T. urartu* (AJ302094 for *Pina;* AJ302102 and AJ302104 for *Pinb)*. Two of the three larger amplicons in the *Pina* gene were not digested by *Dde*I, which indicated the presence of two different alleles. In the *Pinb* gene some of the amplicons of both sizes remained intact after digestion while other were digested into two smaller units, indicating the presence of two more alleles. These differences in the products of digestion provide an indicator of the presence of internal SNPs in the sequences.

In order to characterize and detect the presence of additional variability, the amplicons were excised from the gel, cloned and sequenced (Table 1). Four different alleles were found for the *Pina* gene and seven for the *Pinb* gene. All *Pina* alleles found

in this study had been previously described and catalogued by Simeone et al. (2006), except one described by Lillemo et al. (2000) which we tentatively named *Pina-A^{m}1d.* The *Pina-A^{m}1c* was the most frequent allele (39.3%) found while the alleles *Pina-A^{m}1b* and *Pina-A^{m}1d* were detected less frequently (17.9%). Of the seven *Pinb* genes detected, three (*Pinb-A^{u}1a, Pinb-A^{u}1b, Pinb-A^{m}1e)* had been described previously by Simeone et al. (2006). On the other hand, the *Pinb-A^{m}1h* allele which was catalogued by us presented a similar sequence to that published by Cenci et al. and stored in GenBank (EU307610; unpublished data). Finally, we were able to detect three novel alleles in our survey and catalogued them as *Pinb-A^{m}1i, Pinb-A^{m}1j* and *Pinb-A^{m}1k.*

Grain hardness

The grain hardness of the evaluated lines was estimated by the PSI test (Table 2). Unfortunately, we were unable to use this measurement technique in four of the einkorn lines (PI 237659, PI 355532, PI 355522, and PI 277135) or in the *T. urartu* lines because there was insufficient grain available. Lines were grouped into four different genotypes according to their puroindoline composition. BGE-013624 (*Pina-A^{m}1c, Pinb-A^{m}1i*) was grouped with the *Pina-A^{m}1c, Pinb-A^{m}1e* lines because although *Pinb-A^{m}1e* and *Pinb-A^{m}1i* had different DNA sequences, the protein derived from these genes was the same in both cases. It was found that only the *Pina-A^{m}1b, Pinb-A^{m}1h* group was significantly different from the other groups. This may indicate that the *Pinb-A^{m}1h* allele was responsible for the soft phenotype whereas the *Pina-A^{m}1b, Pinb-A^{m}1k* group produced a harder texture. Another interpretation of this result is that the presence of *Pinb-A^{m}1k* may be related to a harder texture, as *Pinb-A^{m}1e* and *Pinb-A^{m}1i* code for the same protein. Nevertheless these data should be taken as preliminary since only a limited number of accessions was used in the current study.

Nucleotide sequence polymorphism, neutrality test and linkage disequilibrium (LD)

An alignment of the different sequences found together with its comparison to *Pina-D1a* and *Pinb-D1a* alleles from common wheat is showed in Fig. 2. The *Pina-A1* gene displayed seven polymorphic sites, six of them included in the 444 bp coding region (Table 3). Four of these SNPs were synonymous and two were no-synonymous. This level of polymorphism was similar to the described by Chen et al. (2009) that found five polymorphic sites in 56 accessions of einkorn, although very different of the

results showed with *Pina-D1* in *Aegilops tauschii* (Massa et al 2004).

The occurrence of single nucleotide polymorphisms in *Pinb-A1* was considerably higher than that observed in *Pina-A1*. Eleven polymorphic sites were detected within the entire *Pinb-A1* coding region, with an average number of 3.24 nucleotide differences between sequences (Table 3). All SNPs were transitions except one which was a transversion (T→G) at position 354. Eight of these SNPs were synonymous and three of them were non-synonymous. In this respect our results displayed lower levels of polymorphism than those of Massa et al. (2004) who found 33 polymorphic sites in the *Pinb-D1* gene.

The neutrality test was used to evaluate the degree of deviation from an equilibrium neutral model using different statistics parameters. In this way the number of segregating sites (θ) and average number of nucleotide differences per site in two sequences (π) could be calculated (Table 3). These two estimates are expected to give similar values under a drift-mutation balance; otherwise any form of natural selection may explain the maintenance of genetic variation. Two different patterns of variation were identified and the diversity of *Pinb-A1* was found to be slightly greater than that of *Pina-A1* (Table 3). The level and pattern of polymorphisms in the *Pinb-A1* gene was consistent with a neutral equilibrium according to the studies by Tajima (1989) and Fu and Li (1993). Conversely, nucleotide diversity in *Pina-A1* was significantly lower than expected under neutrality. These results do not concur with those obtained by Massa et al. (2004) who found that the *Pinb-D1* gene showed a significantly greater number of polymorphisms than expected under neutrality, while in our case *Pinb-A1* did not show any deviation from equilibrium. On the other hand, the results obtained by Massa et al. (2004) for the *Pina-D1* gene were consistent with a neutral equilibrium in contrast to our results reported above. This could be explained by the fact that the species (einkorn and *T. urartu* vs. *Aegilops* sp.) and the number of accessions were different in the two studies.

In order to study the linkage disequilibrium (LD), pairwise comparisons of nucleotide variants across *Pina-A1* and *Pinb-A1* were tested for a non-random association. For this purpose the two gene sequences were concatenated into a single sequence and arranged in the same order in which the genes are arranged along the chromosome (Tranquilli et al. 1999; Turnbull et al. 2003). All polymorphic sites (18; 7 from *Pina-A1* and 11 from *Pinb-A1)* were included in the analysis. A total of 153

pairwise comparisons were undertaken. According to Fisher exact test, 36.6 % (56) of the pairwise comparisons were significant ($p < 0.01$). After applying the Bonferroni procedure only 10.5% (16) remained significant. When both genes were analysed individually for LD, a higher level of LD was found in *Pina-A1* (76.2% according to Fisher and 33.3% after the Bonferroni procedure vs. 27.3% and 9.1% in *Pinb-A1*). These results support the hypothesis that both genes evolved under different selective constraints. This may be associated with the variation in their defence function *in vivo* as referred to by Massa and Morris (2006).

Puroindoline a and puroindoline b deduced proteins

All the alleles obtained in this study were putatively functional since the translations of the sequences yielded a continuous read from correct initiator to terminator with no premature termination codon. The deduced amino acid sequences of all 11 alleles were 148 residues and they shared a 99.6% identity with the *Pina* gene and 99.3% with the *Pinb* gene. The four *Pina* alleles and the seven *Pinb* alleles found were translated in three and four different proteins respectively (Figure 3). *Pina-A^{m}1b* and *Pina-A^{m}1c* encoded the same Pin-A protein while *Pinb-A^{u}1b, Pinb-A^{m}1e, Pinb-A^{m}1i and Pinb-A^{m}1k* encoded the same Pin-B protein. In each of the sequences there was a relatively high frequency of Cys (6.75%), which is one of the most characteristic features of puroindolines and forms the Cys backbone (Gautier et al. 1994; Lillemo et al. 2002). Due to the presence of this Cys skeleton, puroindolines have been related to other cysteine-rich proteins in wheat which have the same characteristics as some lipid transfer proteins and α-amylases. In addition, the tryptophan-rich domain which is particularly important in the structure and function of both PIN-A and PIN-B proteins was found to be highly conserved in all the sequences studied: all *Pina* sequences showed the WRWWKWWK domain while all the *Pinb* sequences were characterized byWPTKWWK.

These tryptophan residues are probably present in a surface loop which is stabilised by a disulphide bond and it has been suggested that they form the lipid-binding site (Koojiman et al. 1997). Mutations in these domains are unusual, although those such as Gly46→Ser and Trp44→Arg have been described in the *Pinb* gene of common wheat, and both such mutations lead to a hard phenotype (Lillemo et al. 2000). Other authors have described mutations in this region of the *Pina* gene in other species

(*Aegilops, Secale, Hordeum)* (Massa et al. 2006).

Conclusions

The genetic analysis of a diploid wheat collection has demonstrated that these species are a potential source of novel puroindoline variants. Our data suggest that, although further studies must be carried out, these variants could to be used to expand the range of grain texture in durum and common wheat, which would permit the development of new materials adapted to novel uses in the baking and pasta industry.

Acknowledgements

This research was supported by grant AGL2007-65685-C02-02 from the Spanish Ministry of Science and Innovation and the European Regional Development Fund (FEDER) from the European Union. The first author thanks to the Spanish Ministry of Education and Science (FPU programme) for a predoctoral fellowship. The technical assistance of Ana Moral is highly appreciated.

References

Alvarez JB, Moral A, Martín LM (2006) Polymorphism and genetic diversity for the seed storage proteins in Spanish cultivated einkorn wheat (*Triticum monococcum* L. ssp. *monococcum*). Genet Resour Crop Evol 53: 1061-1067.

Bhave M, Morris CF (2008a) Molecular genetics of puroindolines and related genes: allelic diversity in wheat and other grasses. Plant Mol Biol 66: 205-219.

Bhave M, Morris CF (2008b) Molecular genetics of puroindolines and related genes: regulation of expression, membrane binding properties and applications. Plant Mol Biol 66: 221-231.

Caballero L, Martín MA, Alvarez JB (2008) Allelic variation for the high- and low-molecular-weight glutenin subunits in wild diploid wheat (*Triticum urartu*) and its comparison with durum wheats. Aust J Agr Res 59: 906-910.

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, Wissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploidy wheat species (*Triticum* and *Aegilops*) Plant Cell 17: 1033-1045.

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

Chen Q, Qi PF, Wei YM, Wang JR, Zheng YL (2009) Molecular characterization of the *pina* gene in einkorn wheat. Biochem Genet 47: 384-396.

Dubreil L, Compoint JP, Marion D (1997) Interaction of puroindoline with wheat flour polarlipids determines their foaming properties. J Agric Food Chem 45: 108-116.

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

Fu YX, Li WH (1993) Statiscal tests of neutrality of mutations. Genetics 133: 693-709.

Gautier MF, Aleman ME, Guirao A, Marion D, Joudrier P (1994) *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.

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.

Guzman C, Caballero L, Alvarez JB (2009) Variation in Spanish cultivated einkorn wheat (*Triticum monococcum* L. ssp. *monococcum*) as determined by morphological traits and waxy proteins. Genet Resour Crop Evol 56: 601-604.

Koojiman M, Orsel R, Hessing M, Hamer RJ, Bekkers ACAPA (1997) Spectroscopic characterization of the lipid-binding properties of wheat puroindolines. J Cereal Sci 26: 145-159.

Krishnamurthy K, Balconi C, Sherwood JE, Giroux MJ (2001) Wheat puroindolines 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 chromosome substitution lines. In: Seed Protein Improvement by Nuclear Techniques. International Atomic Energy Agency, Vienna, Austria, pp.483-502.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA

polymorphism data. Bioinformatics 25: 1451-1452.

Lillemo M, Morris CF (2000) A leucine to proline mutation in puroindoline b is frecuently present in hard wheats from Northern Europe. Theor Appl Genet 100: 1100-1107.

Lillemo M, Simeone MC, Morris CF (2002) Analysis of puroindoline a and b sequences from T*riticum aestivum* cv. "Penawawa" and related diploid taxa. Euphytica 126: 321-331

Martín MA, Martín LM, Alvarez JB (2008) Polymorphisms at the *Gli-A^{u}1* and *Gli-A^{u}2* loci in wild diploid wheat (*Triticum urartu*). Euphytica 163: 303-307.

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.

Massa AN, Morris CF (2006) Molecular evolution of the puroindoline-a, puroindoline-b, and grain softness protein-1 genes in the Tribe *Triticeae*. J Mol Evol 63: 526-536.

Morris CF, Rose SP (1996) Cereal Grain Quality (Chapman and Hall, New York), pp 3-54.

Morris CF, Bhave M (2008) Reconciliation of D-genome puroindoline allele designations with current DNA sequence data. J Cereal Sci 48:277-287.

Pomeranz Y, Williams PC (1990) Wheat hardness: its genetic, structural and biochemical background, measurement, and significance. Advances in Cereal Science and Technology 10, 471-546.

Sharma HC, Waines JG, Foster KW (1981) Variability in primitive and wild wheats for useful genetic characters. Crop Sci 21: 555-559.

Simeone MC, Gedye KR, Mason-Gamer R, Gill BS, Morris CF (2006) Conserved regulatory elements identified from a comparative puroindoline gene survey of *Triticum* and *Aegilops* diploid taxa. J Cereal Sci 44: 21-33.

Sokal RR, Rohlf FJ (1969) Biometry: the principles and practice of statitics in biological research. San Francisco: W H Freeman, 776p.

Tranquilli G, Heaton J, Chicaiza O, Dubcovsky J (1999) Genetic and physical characterization of grain texture-related loci in diploid wheat. Mol Gen Genet 262: 846-850.

Turnbull KM, Turner M, Mukai Y, Yamamoto M, Morell MK, Appels R, Rahman S

(2003) The organization of genes tighlty linked to the *Ha* locus in *Aegilops tauschii*, the D-genome donor to wheat. Genome 46: 330-338.

Tajima F (1989) Statiscal method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics: 123: 585-595.

Weir BS (1996). Genetic Data Analysis II. Sinauer associates. Sunderland, MA.

Willians, F; Jaby El-Haramein, F; Nakkoul, H; Rihawi, S. 1986. Crop quality evaluation methods and guidelines. International Center for Agriculture Research in the Dry Areas (ICARDA). Syria. 145 p.

Table 1. Plants used to survey puroindoline genes in einkorn wheat and *T. urartu*

Accession	Country	*Pina*	Sequence	*Pinb*	Sequence
Einkorn wheat (A^m)					
BGE 014269	Spain	*Pina-A^m1b*	HQ696585	*Pinb-A^m1h*	HQ696590
PI 352483	Spain	*Pina-A^m1b*	HQ696585	*Pinb-A^m1h*	HQ696590
PI 427927	Iraq	*Pina-A^m1b*	HQ696585	*Pinb-A^m1h*	HQ696590
PI 190945	Spain	*Pina-A^m1b*	HQ696585	*Pinb-A^m1k*	HQ696591
PI 190947	Spain	*Pina-A^m1b*	HQ696585	*Pinb-A^m1k*	HQ696591
BGE 013625	Spain	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
BGE 020470	Spain	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 221414	Serbia	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 237659	Kenya	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 272563	Hungary	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 306543	Romania	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 355532	Italy	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 428157	U.K.	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 591871	Georgia	*Pina-A^m1c*	HQ696586	*Pinb-A^m1e*	HQ696592
PI 355522	Balkans	*Pina-A^m1c*	HQ696586	*Pinb-A^m1h*	HQ696590
BGE 013624	Spain	*Pina-A^m1c*	HQ696586	*Pinb-A^m1i*	HQ696593
PI 277135	Albania	*Pina-A^m1d*	HQ696587	*Pinb-A^m1j*	HQ696594
PI 326317	Azerbaijan	*Pina-A^m1d*	HQ696587	*Pinb-A^m1j*	HQ696594
PI 355528	Minor Asia	*Pina-A^m1d*	HQ696587	*Pinb-A^m1j*	HQ696594
PI 428149	Sweden	*Pina-A^m1d*	HQ696587	*Pinb-A^m1j*	HQ696594
PI 518452	Spain	*Pina-A^m1d*	HQ696587	*Pinb-A^m1j*	HQ696594
T. urartu (A^u)					
PI 428335	Lebanon	*Pina-A^u1a*	HQ696584	*Pinb-A^u1a*	HQ696588
PI 428196	Turkey	*Pina-A^u1a*	HQ696584	*Pinb-A^u1b*	HQ696589
PI 428233	Turkey	*Pina-A^u1a*	HQ696584	*Pinb-A^u1b*	HQ696589
PI 428240	Turkey	*Pina-A^u1a*	HQ696584	*Pinb-A^u1b*	HQ696589
PI 428253	Iraq	*Pina-A^u1a*	HQ696584	*Pinb-A^u1b*	HQ696589
PI 428317	Iraq	*Pina-A^u1a*	HQ696584	*Pinb-A^u1b*	HQ696589
TRI-11563	URRS	*Pina-A^u1a*	HQ696584	*Pinb-A^u1b*	HQ696589

Table 2 Particle size index (PSI) and allelic composition of the einkorn lines used in grain hardness analysis

Puroindolines composition	PSI (%)	Accessions
Pina-A^m1b, Pinb-A^m1h	56.31a	BGE 014269; PI 352483; PI 427927
Pina-A^m1b, Pinb-A^m1k	48.26b	PI 190945; PI 190947
Pina-A^m1c, Pinb-A^m1e/i	49.34b	BGE 013624; BGE 013625; BGE 020470; PI 221414; PI 272563; PI 306543; PI 428157; PI 591871
Pina-A^m1d, Pinb-A^m1j	49.59b	PI 326317; PI 355528; PI 428149; PI 518452

Mean follow the same letter are not significant differences.

Table 3 Summary of DNA polymorphism and test statistics for selection in 28 accessions of einkorn wheat.

Gene	Length (bp)	*n*	*s*	*h*	θ	π	*D*	*D**	*F**
Pina	516	2.84	7	4	0.0035	0.0055	1.732	1.277	1.646
Pinb	529	3.24	11	7	0.0053	0.0061	0.473	0.457	0.540

n: average number of nucleotide differences; *s:* estimates of number of segregating sites; *h:* number of haplotypes; *θ:* number of segregating sites per site between two sequences; *π:* average number of nucleotide differences per site between two sequences; *D:* Tajima´s estimate *D*-test; *D*:* Fu and Li *D**-test; *F*:* Fu and Li *F**-test.

Caption figures

Fig. 1 **a** Amplification products of the Puroindoline a and Puroindoline b gene. **b** Pina and Pinb amplification products digested with DdeI and BsmAI/NciI respectively. Lanes: 1, PI 221414; 2, PI 190945; 3, BGE 013624; 4, PI 428335; 5, PI 428233; 6, PI 355522; 7, PI 272563; 8, PI 518452; 9, PI 428317; 10, PI 427917.

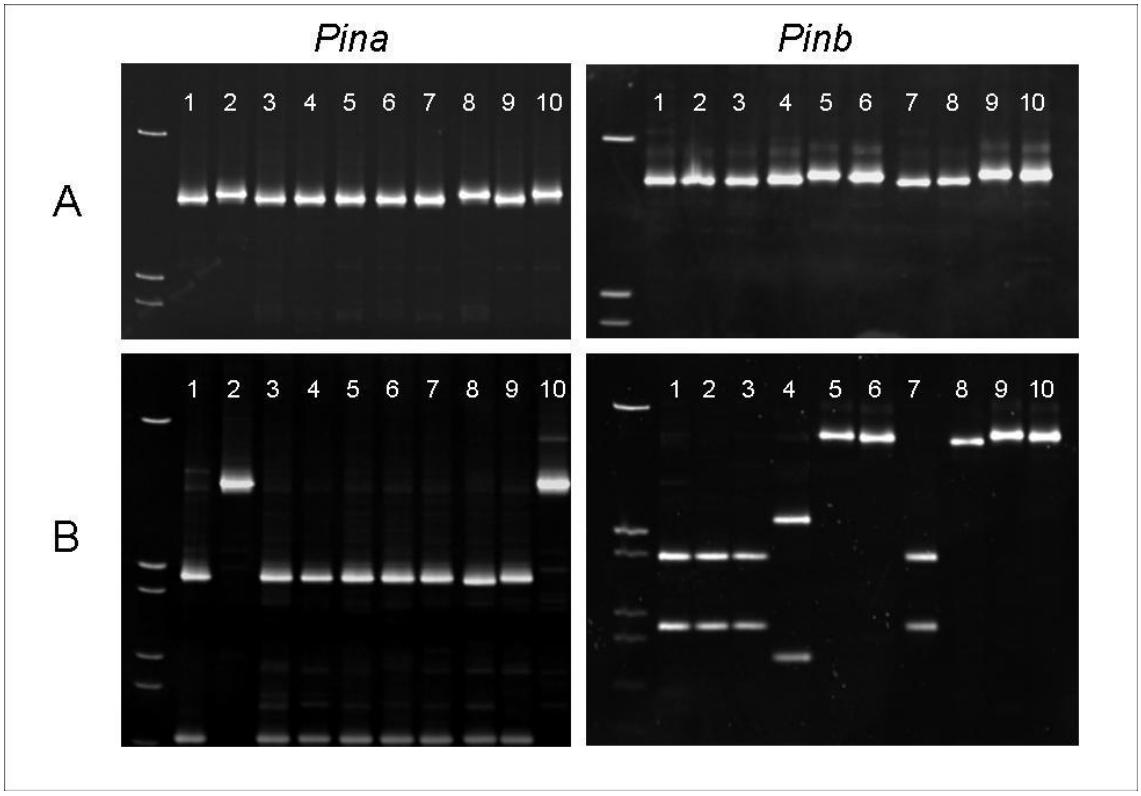

Fig. 2 Alignment of genomic DNA sequences of the *Pina* and *Pinb* genes.

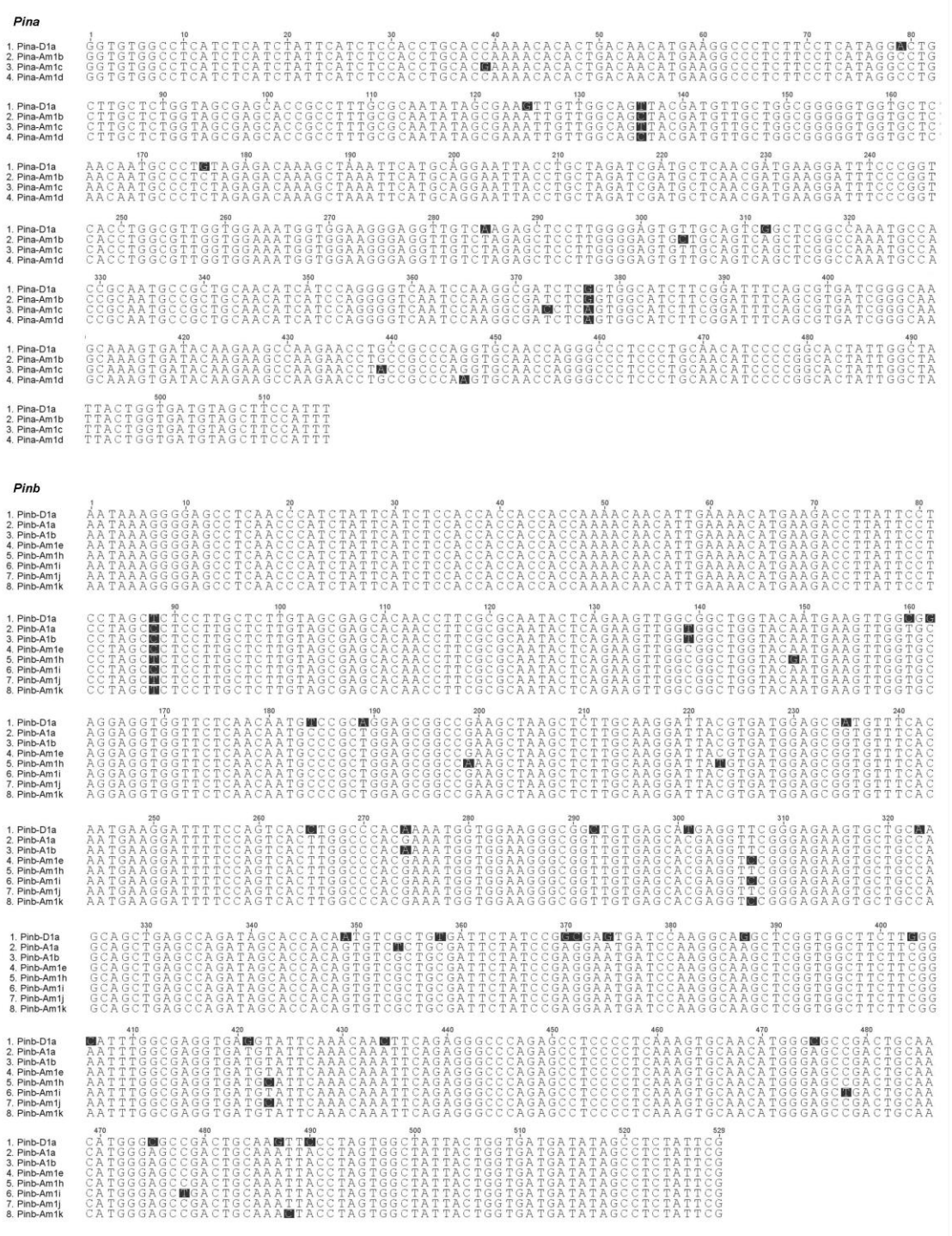

Fig. 3 Alignment of the predicted Pin-a and Pin-b proteins from common wheat (*Pina-D1a*) and einkorn alleles described in this study. Arrows indicate Cys residues. Tryptophan rich residues are indicated by boxes. Black stars indicate polymorphism between einkorn sequences.