Molecular Breeding (2006) 17: 191–200
DOI 10.1007/s11032-005-4886-3
© Springer 2006

# Molecular characterization of puroindolines and their encoding genes in *Aegilops ventricosa*

L. Gazza, S. Conti, F. Taddei and N.E. Pogna*
*Istituto Sperimentale per la Cerealicoltura, C.R.A., Section of Applied Genetics, Via Cassia 176, 00191 Rome, Italy; *Author for correspondence (e-mail: pognanorberto@mclink.it; phone: +39-6-3295705; fax: +39-6-3294864)*

Received 1 August 2005; accepted in revised form 3 November 2005

*Key words: Aegilops ventricosa*, Endosperm texture, Puroindolines

## Abstract

Puroindolines, the tryptophan-rich proteins controlling grain hardness in wheat, appeared as two pairs of 13 kDa polypeptides in the Acid-PAGE (A-PAGE) and two-dimensional A-PAGE×SDS-PAGE patterns of starch-granule proteins from wild allotetraploid wheat *Aegilops ventricosa* Tausch. ($2n = 4x = 28$, genomes $D^vD^vN^vN^v$). Puroindoline pair *a1* + *a2* reacted strongly with an antiserum specific for puroindoline-a from common wheat (*Triticum aestivum* L.), whereas puroindoline pair *b1* + *b2* exhibited A-PAGE relative mobilities similar to that of puroindoline-b in *Aegilops tauschii* (Coss.), the D-genome donor to both common wheat and *Ae. ventricosa*. Puroindolines *a2* and *b1* were found to be encoded by alleles *Pina-D1a* and *Pinb-D1h* on chromosome $5D^v$, respectively, whereas puroindolines *a1* and *b2* were assumed to be under the genetic control of chromosome $5N^v$. Puroindoline *a1* encoded by the novel *Pina-N1a* allele exhibited a high level of amino acid variation with respect to puroindoline-a. On the other hand, the tryptophan-rich region of puroindoline *b2* encoded by allele *Pinb-N1a* showed a sequence change from lysine-42 to arginine, with no effect on the amount of protein *b2* accumulated on the starch granules. A partial duplication of the pin-B gene (*Pinb-relic*) was identified about 1100 bp downstream from *Pinb-D1* on chromosome $5D^v$. The present findings are the first evidence of a tetraploid wheat species in which four puroindoline genes are expressed. The potential of *Ae. ventricosa* as a source of genes that may be used to modulate endosperm texture and other valuable traits in cultivated wheat species is discussed.

## Introduction

Puroindolines a (Pin-a) and b (Pin-b) are two α-helical, tryptophan- and cysteine-rich isoforms occurring in the starchy endosperm of the *Triticeae* and *Aveneae* tribes (see Morris 2002 for a review). The lipid-binding properties of these basic proteins, around 13 kDa in size, are assumed to account for their foaming properties (Clark et al. 1994) and permeabilizing effects on bacterial and fungal membranes (Blochet et al. 1993; Krishnamurthy et al. 2001; Jing et al. 2003). Variation in puroindoline composition was also shown to affect grain hardness, crumb structure of bread and rheological properties of wheat dough (Gautier et al. 1994; Dubreil et al. 1997, 1998; Giroux and Morris 1997, 1998; Corona et al. 2001a; Igrejas et al. 2001).

In commom wheat (*Triticum aestivum* L., $2n = 6x = 42$, genomes AABBDD) and *Aegilops tauschii* (Coss.) ($2n = 2x = 14$, genome DD), Pin-a and Pin-b are encoded at the *Pina-D1* and *Pinb-D1* loci, respectively (Greenwell and Schoefield

1989; Giroux and Morris 1997; Gautier et al. 2000; Massa et al. 2004). These loci, along with two degenerated duplications (*PseudoPinb* and *Pinb-relic*) closely linked to *Pinb-D1*, were identified within BAC clones representing the *Hardness* (*Ha*) locus at the extreme telomeric end of the short arm of chromosome 5D (Chantret et al. 2005).

Large amounts of Pin-a and Pin-b accumulate on the starch granules of soft common wheat cultivars possessing wild-type alleles *Pina-D1a* and *Pinb-D1a* (Giroux and Morris 1997, 1998; Lillemo and Morris 2000). On the contrary, reduced amounts of Pin-b, if any, are present in medium-hard cultivars with alleles *b, c, d, e, f* or *g* at *Pinb-D1*. Compared with *Pinb-D1a*, these alleles exhibit either a single amino acid substitution (Lillemo and Morris 2000; Corona et al. 2001b) or a stop codon in the central domain of the Pin-b gene (Lillemo and Morris 2000).

Furthermore, no traces of Pin-a are present on the surface of starch granules of hard common wheat cultivars with a null allele at the *Pina-D1* locus (Giroux and Morris 1998; Lillemo and Morris 2000; Turnbull et al. 2000; Corona et al. 2001a; Gazza et al. 2005). Despite the presence of 'soft' (wild-type) allele *Pinb-D1a*, null Pin-a cultivars show reduced amounts of Pin-b in their starch extracts, suggesting that Pin-a is required for adhesion of Pin-b to the starch granules (Corona et al. 2001a; Gazza et al. 2005).

Recently, eight novel alleles, five at *Pina-D1* and three at *Pinb-D1*, have been described in *Ae. tauschii*, the D-genome donor to common wheat (Massa et al. 2004). Puroindoline genes also occur in diploid wheat species containing A, S, M, C or U genome, but are absent in extra-hard tetraploid wheat species carrying AB (*T. dicoccoides*, *T. dicoccum* and *T. turgidum* ssp *durum*) or AG genomes (*T. timopheevii*) (Gautier et al. 2000; Pogna et al. 2002).

In this paper puroindolines and their encoding genes in wild tetraploid wheat species *Ae. ventricosa* Tausch. are compared with their counterparts in *Ae. tauschii* and common wheat. *Ae. ventricosa* (syn. *T. ventricosum* Ces., $2n = 4x = 28$, genomes $D^vD^vN^vN^v$) contains the D and N genomes derived from *Ae. tauschii* and *Ae. uniaristata*, respectively (Kimber and Zhao 1983). The primary aim of this study was to investigate the expression of puroindoline genes in a tetraploid wheat background, and identify puroindoline alleles of *Ae. ventricosa* that may be transferred into cultivated wheat species to modulate their grain texture characteristics.

## Materials and methods

### *Plant material*

Hulled grains (1000-seeds wt = 21.4 g) of *Ae. ventricosa* Accession L36 and *Ae. tauschii* ssp *typica* Accession L35 (1000-seeds wt = 9.2 g) were from a wheat collection grown at the Istituto Sperimentale per la Cerealicoltura, Rome, Italy, were used in the present study. Common wheat cv. Bolero was analysed as well.

### *Acid-PAGE (A-PAGE) and two-dimensional A-PAGE×SDS-PAGE*

Puroindolines were extracted from air-dried starch granules obtained as described previously (Corona et al. 2001a). Starch granules (50 mg) were suspended in a solution containing 50 mM NaCl and 50% (v/v) propan-2-ol. After sonication for a few seconds, the suspension was vortexed at room temperature for 1 h and then centrifuged at $8{,}000 \times g$ for 10 min. Proteins in the supernatant were precipitated with two volumes of acetone at −20 °C overnight, and then air-dried. Before loading for A-PAGE fractionation, proteins were suspended in 50 μl of 8.5 mM sodium lactate buffer (3.4 g/l of 97% NaOH adjusted to pH 3.1 with lactic acid) and mixed with a half volume of 50% (v/v) glycerol, containing 0.1% (w/v) pyronine Y. Electrophoresis was performed at pH 3.1 as described previously (Gazza et al. 2005).

Two-dimensional A-PAGE×SDS-PAGE was carried out as described by Redaelli et al. (1995). After the first dimension (A-PAGE), the gel was incubated at room temperature for 45 min in an equilibration solution containing 0.0625 M Tris–HCl, pH 6.8, 2% (w/v) SDS, 40% (w/v) glycerol and 5% β-mercaptoethanol. The gel was then loaded onto an SDS-PAGE gel prepared with 20% acrylamide (T = 20%, C = 0.05%), 0.375 M Tris–HCl, pH 8.4 and 0.1% (w/v) SDS, and run until the tracking dye reached the bottom of the gel. A 0.25% (w/v) solution of Coomassie Brillant Blue R250 in 6% trichloracetic acid was used to fix and stain both the A-PAGE and SDS-PAGE gels.

*Western blotting*

Western blotting was performed in a Bio-Rad semi-dry transfer cell using nitrocellulose membranes (Hybond-C Extra, AmershamBiosciences, Sweden). A-PAGE gels were equilibrated for 20 min in a transfer buffer, pH 8.3, containing 25 mM Tris, 192 mM glycine, 1% $\beta$-mercaptoethanol and 20% (v/v) methanol in distilled water, transferred between two double layers of 3 MM chromatography paper (Whatman, UK), pre-equilibrated in the same transfer buffer without $\beta$-mercaptoethanol, and electro-blotted at 23 V for 50 min. After the transfer, membranes were maintained for 1 h in PBS buffer containing 5% (w/v) blocking agent (AmershamBiosciences, Sweden), and incubated for 16 h in the same buffer containing 0.2% (w/v) blocking agent and a 1:500 dilution of a polyclonal Pin-a- or Pin-b-specific antiserum developed according to Krishnamurthy and Giroux (2001). After incubation with a 1:2500 dilution of a goat anti-rabbit horseradish peroxidase conjugate (Promega, USA), blots were stained with 4-chloro-1-naphthol and hydrogen peroxide. The antisera were prepared by standard methods (Primm Srl, Milan, Italy).

*DNA isolation, PCR amplification and cloning*

Genomic DNAs were isolated from young leaves using the procedure of Dellaporta et al. (1983). The samples were denaturated at 94 °C for 5 min before the addition of *Taq*-polymerase. The sense strand primer 5′-CCTCGGACACCTTGTTAA-3′ (named 'Pina-prom') and the antisense strand primer 5′-TCACCAGTAATAGCCAATAGTG-3′ ('Pina-low') were used to amplify the Pin-a gene using 35 cycles of 1 min of denaturation at 94 °C, 1.5 min annealing at 54 °C and 2 min elongation at 72 °C. Amplification of the Pin-a gene was also performed with the sense primer 5′-ATGAAG GCCCTCTTCCTCA-3′ ('Pina-up') coupled with either 'Pina-low' or the antisense primer 5′-TCA-TAAATTATTCCATGACCA-3′ ('Pina-ter') following the protocol described above except that the annealing temperature was 50 °C. The Pin-b gene was amplified with the sense strand primer 5′-CACATGATTCTAAATAC-3′ ('Pinb-prom') and the antisense strand primer 5′-TCACCAGTAAT AGCCACTAGGGAA-3′ ('Pinb-low') using 35 cycles of 1 min of denaturation at 94 °C, 1.5 min annealing at 54 °C and 2 min elongation at 72 °C. The 3′ boundary of the Pin-b gene was amplified with the sense strand primer 5′-GTAAGGATTAT GTGATGGGGT-3′ ('Pinb-int') and the antisense strand primer 5′-GAGATCTACATGAAGCAG CT-3′ ('Pinb-rel') using 35 cycles of 1 min of denaturation at 94 °C, 1.5 min annealing at 58 °C and 1.5 min elongation at 72 °C. All PCR amplifications were followed by a final cycle with an extension of 7 min at 72 °C. The PCR products were separated on 1.8% agarose gels, stained with ethidium bromide and visualized under UV.

The amplification products were eluted from the agarose gel with the Nucleospin-Extract kit (Macherey-Nagel, Germany) and cloned using the TOPO® TA Cloning Kit for Sequencing containing One Shot TOP10 Electrocomp *E. coli*, according to the manufacturer's instructions (Invitrogen, USA).

*DNA sequencing*

Plasmids with the inserted DNAs were obtained by the lysis method using Wizard Plus SV Minipreps DNA Purification System (Promega, USA), and sequenced on a PerkinElmer ABI Prism 377 DNA sequencer, using the dideoxynucleotide chain termination method. Direct sequencing was also performed on PCR fragments eluted from the agarose gel as described above. Sequence data of *Ae. ventricosa* from this article have been deposited in the GenBank data libraries.

## Results

*Electrophoretic fractionation of starch-granule proteins in* Ae. tauschii *and* Ae. ventricosa

As expected, A-PAGE fractionation of starch-granule proteins in common wheat cv. Bolero (Figure 1, lane 3) revealed two prominent bands corresponding to Pin-a and Pin-b encoded by the 'soft' (wild-type) *Pina-D1a* and *Pinb-D1a* alleles, respectively (Corona et al. 2001a). Two major bands (*a3* and *b3)* were also observed in Accession L35 of *Ae. tauschii* (Figure 1, lane 1), whereas Accession L36 of tetraploid wheat *Ae. ventricosa* exhibited two pairs of polypeptides designated

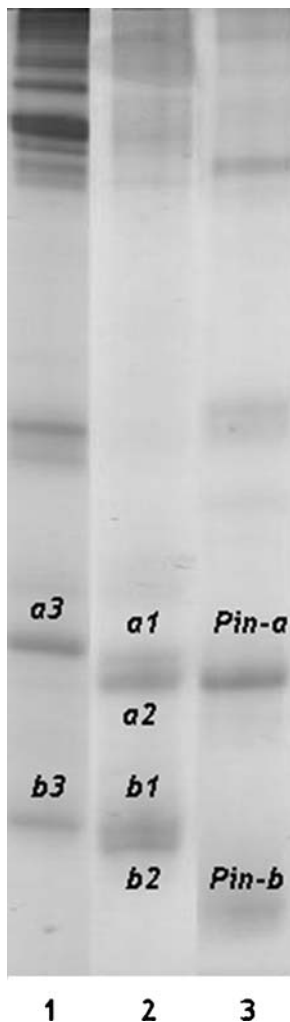


*Figure 1.* A-PAGE fractionation of starch-granule proteins from (1) *Aegilops tauschii* Accession L35, (2) *Ae. ventricosa* Accession L36 and (3) *Triticum aestivum* cv. Bolero. Puroindolines a (Pin-a) and b (Pin-b) in cv. Bolero are encoded by 'soft' (wild-type) alleles *Pina-D1a* and *Pinb-D1a*, respectively. Letters *a* and *b* indicate puroindolines in *Aegilops tauschii* and *Ae. ventricosa*.

*a1* + *a2* and *b1* + *b2* (Figure 1, lane 2). Band *a1* was fainter and slightly slower than band *a2*, which displayed the same relative mobility as Pin-a in common wheat cv. Bolero. On the other hand, bands *b1* and *b2* occurred in similar amounts in the cathodic region of the gel, their relative mobilities being comparable to that of band *b3* in *Ae. tauschii*, but significantly slower than that of Pin-b in cv. Bolero.

When fractionated by two-dimensional electrophoresis (A-PAGE in the first dimension and SDS-PAGE in the second), protein pairs *a1* + *a2* and *b1* + *b2* from *Ae. ventricosa* appeared as four spots in the $M_r$ 13–15 kDa region, proteins *b1* + *b2* being slightly faster than proteins *a1* + *a2* in the second dimension Figure 2). Two spots corresponding to proteins *a3* and *b3* were observed in the same 13–15 kDa region of the A-PAGE×SDS-PAGE pattern of *Ae. tauschii* (data not shown).

### *Western blotting of starch-granule proteins in* Ae. tauschii *and* Ae. ventricosa

The polyclonal antiserum developed against the 16-mer DRASKVIQEAKNLPPR sequence in the C-terminal region of mature Pin-a (Krishnamurthy and Giroux 2001) reacted strongly with proteins *a1*, *a2* and *a3* fractionated by A-PAGE (Figure 3, lanes 1 and 2), suggesting that these latter polypeptides belong to the Pin-a family. On the contrary, the polyclonal Pin-b-specific antiserum developed against the 16-mer GEVFKQLQRAQSLPSK epitope in the C-terminal region of mature Pin-b (Krishnamurthy and Giroux 2001) reacted only with Pin-b in cv. Bolero, whereas it gave no reaction with the starch extracts from *Ae. tauschii* or *Ae. ventricosa* (data not shown).

### *PCR amplification of puroindolines in* Ae. tauschii *and* Ae. ventricosa

The sense strand 'Pina-prom' and the antisense strand 'Pina-ter' are located 859 bp upstream of and 84 bp downstream from the coding DNA sequence (CDS) of the *Pina-D1a* allele, respectively (Chantret et al. 2005), whereas the sense strand 'Pina-up' and the antisense strand 'Pina-low' are positioned at the 5′ and 3′ ends of the CDS of this allele, respectively (Gautier et al. 1994).

Direct sequencing of the 1188 bp PCR fragment amplified from *Ae. tauschii* DNA with the primer pair 'Pina-prom' + 'Pina-low' revealed the presence of the *Pina-D1d* allele (Figure 4a). This allele differs from wild-type *Pina-D1a* in a nucleotide substitution that determines a change from arginine-58 to glutamine (Massa et al. 2004) (Figure 4b). Amplification of *Ae. ventricosa* DNA with the primer pairs 'Pina-prom' + 'Pina-low' or 'Pina-up' + 'Pina-ter' gave two PCR fragments, which were cloned in *E. coli*. Sequencing of 10

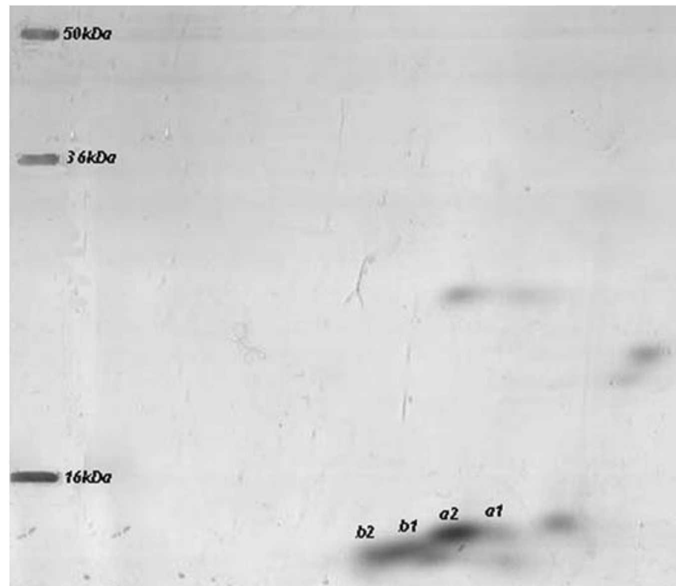


*Figure 2*. Two-dimensional (A-PAGE×SDS-PAGE) fractionation of starch-granule proteins from *Ae. ventricosa* Accession L36. Letters indicate puroindolines a and b. Molecular weight markers are shown at the left-hand side.

clones from each transformation event revealed two different Pin-a genes in *Ae. ventricosa* Accession L36 (Figure 4a). One Pin-a gene (GenBank DQ124421) was found to correspond to wild-type allele *Pina-D1a*, whereas the other gene (GenBank DQ124419), provisionally called *Pina-N1a*, exhibited 12 single nucleotide polymorphisms (SNPs) in its CDS, six of them being transitions. Amongst the 10 non-synonymous SNPs, five were clustered in the 33 bp region at the extreme 3′ end of the CDS. One of these SNPs changed G to A at position 443, and resulted in a TAG stop codon

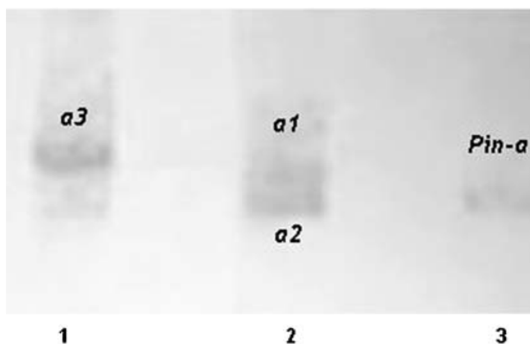


*Figure 3*. Reaction of anti-Pin-a antiserum against starch-granule proteins fractionated by A-PAGE from (1) *Ae. tauschii* Accession L35, (2) *Ae. ventricosa* Accession L36 and (3) *T. aestivum* cv. Bolero. Letters indicate puroindolines.

adjacent to the 'normal' TGA stop codon (Figure 4). As expected, amplification of *Ae. ventricosa* DNA with the terminal primers 'Pina-up' + 'Pina-low', which are present in all *Pina-D1* alleles described so far (Massa et al. 2004), gave a 447 bp fragment corresponding to allele *Pina-D1a*.

The sense strand 'Pinb-prom' is located 432-bp upstream of the CDS of allele *Pinb-D1a*, whereas the antisense strand 'Pinb-low' is at the 3′ end of this CDS (Chantret et al. 2005). The 878 bp PCR fragment amplified from *Ae. tauschii* DNA with the primer pair 'Pinb-prom' + 'Pinb-low' was found to correspond to allele *Pinb-D1i* (Massa et al. 2004). On the other hand, sequencing of 10 *E. coli* clones transformed with an 878 bp fragment amplified from *Ae. ventricosa* DNA with the same primer pair revealed two different Pin-b genes (Figure 4a). One gene (GenBank DQ124422) was identical to the *Pinb-D1h* allele observed in *Ae. tauschii* by Massa et al. (2004), whereas the other gene (GenBank DQ124420), provisionally called *Pinb-N1a*, was considerably different from alleles *h*, *i* or *a* at the *Pinb-D1* locus (Figure 4a). Compared with wild-type *Pinb-D1a*, the entire coding region of *Pinb-N1a* displayed 13 non-synonymous SNPs, one of them resulting in the change from lysine-42 to arginine in the tryptophan-rich region of the encoded protein (Figure 4b).

**(a)**

```
Pina-D1a   ATGAAGGCCC TCTTCCTCAT AGGACTGCTT GCTCTGGTAG CGAGCACCGC CTTTGCGCAA TATAGCGAAG TTGTTGGCAG 80
Pina-D1d   ATGAAGGCCC TCTTCCTCAT AGGACTGCTT GCTCTGGTAG CGAGCACCGC CTTTGCACAA TATAGCGAAG TTGTTGGCAG 80
Pina-N1a   ATGAAGGCCC TCTTCCTCAT AGGCCTGCTT GCTCTGGTAG CGAGCACCGC CTTTGCGCAA TATAGCGAAG TTGTTGGCAG 80
                                    *                                   *

Pina-D1a   TTACGATGTT GCTGGCGGGG GTGGTGCTCA ACAATGCCCT GTAGAGACAA AGCTAAATTC ATGCAGGAAT TACCTGCTAG 160
Pina-D1d   TTACGATGTT GCTGGCGGGG GTGGTGCTCA ACAATGCCCT GTAGAGACAA AGCTAAATTC ATGCAGGAAT TACCTGCTAG 160
Pina-N1a   TTACGATGCT GCTGGCGGGG GTGGTGCTCA ACAATGCCCT GTAGAGACAA AGCTAGATTC ATGCAGGAAT TACCTGCTAG 160
                   *                                                   *

Pina-D1a   ATCGATGCTC AACGATGAAG GATTTCCCGG TCACCTGGCG TTGGTGGAAA TGGTGGAAGG GAGGTTGTCA AGAGCTCCTT 240
Pina-D1d   ATCGATGCTC AACGATGAAG GATTTCCCGG TCACCTGGCG TTGGTGGAAA TGGTGGAAGG GAGGTTGTCA AGAGCTCCTT 240
Pina-N1a   ATCGATGCTC AACGATGAAG GATTTCCCGG TCACCTGGCG TTGGTGGAAA TGGTGGAAGG GAGGTTGTCT AGAGCTCCTT 240
                                                                                      *

Pina-D1a   GGGGAGTGTT GCAGTCGGCT CGGCCAAATG CCACCGCAAT GCCGCTGCAA CATCATCCAG GGGTCAATCC AAGGCGATCT 320
Pina-D1d   GGGGAGTGTT GCAGTCAGCT CGGCCAAATG CCACCGCAAT GCCGCTGCAA CATCATCCAG GGGTCAATCC AAGGCGATCT 320
Pina-N1a   GGGGAGTGTT GCAGTCGGCT CGGCCAATTG CCACCGCAAT GCCGCTGCAA CATCATCCAG GGGTCAATCC AAGGCGATCT 320
                            *           *

Pina-D1a   CGGTGGCATC TTCGGATTTC AGCGTGATCG GGCAAGCAAA GTGATACAAG AAGCCAAGAA CCTGCCGCCC AGGTGCAACC 400
Pina-D1d   CGGTGGCATC TTCGGATTTC AGCGTGATCG GGCAAGCAAA GTGATACAAG AAGCCAAGAA CCTGCCGCCC AGGTGCAACC 400
Pina-N1a   CGGTAGCATC TTTGGATTTC AGCGTGATCG GGCAAGCAAA GTGATACAAG AAGCCAAGAA CCTGCCGCCC AGGTGCAACC 400
               *        *

Pina-D1a   AGGGCCCTCC CTGCAACATC CCCGGCACTA TTGGCTATTA CTGGTGA 447
Pina-D1d   AGGGCCCTCC CTGCAACATC CCCGGCACTA TTGGCTATTA CTGGTGA 447
Pina-N1a   AGGGCCCTCC CTGCGACATC CGCAGCACTA GTGGCTATTA CTAGTGA 447
                          *       *         *            *


Pinb-D1a   ATGAAGACCT TATTCCTCCT AGCTCTCCTT GCTCTTGTAG CGAGCACAAC CTTCGCGCAA TACTCAGAAG TTGGCGGCTG 80
Pinb-D1h   ATGAAGACCT TATTCCTCCT AGCTCTCCTT GCTCTTGTAG CGAGCACGAC CTTCGCGCAA TACTCAGAAG TTGGCGGCTG 80
Pinb-D1i   ATGAAGACCT TATTCCTCCT AGCTCTCCTT GCTCTTGTAG CGAGCACGAC CTTCGCGCAA TACTCAGAAG TTGGCGGCTG 80
Pinb-N1a   ATGAAGACCT TATTCCTCCT AGCTCTCCTT GCTCTTGTAG CGGGCACAAC CTTCGCGCAA TACTCAGAAG TTGGCGGCTG 80
                                                         *    *

Pinb-D1a   GTACAATGAA GTTGGCGGAG GAGGTGGTTC TCAACAATGT CCGCAGGAGC GGCCGAAGCT AAGCTCTTGC AAGGATTACG 160
Pinb-D1h   GTACAATGAA GTTGGTGCAG GAGGTAGTTC TCAACAATGC CCGCTGGAGC GGCCGAAGCT AAGCTCTTGT AAGGATTATG 160
Pinb-D1i   GTACAATGAA GTTGGTGCGG GAGGTAGTTC TCAACAATGC CCGCTGGAGC GGCCGAAGCT AAGCTCTTGT AAGGATTATG 160
Pinb-N1a   GTACAATGAA GTTGGTGCAG GAGGTGGTTC TCAACAATGC CCGATGGAGC GGCCGAACCT AAGCTCTTGC AAGGATTACG 160
                           * **       *              *    **             *            *         *

Pinb-D1a   TGATGGAGCG ATGTTTCACA ATGAAGGATT TTCCAGTCAC CTGGCCCACA AAATGGTGGA AGGGCGGCTG TGAGCATGAG 240
Pinb-D1h   TGATGGGGTG GTGTTTCACA ATGAAGGATT TTCCATTCAC TTGGCCCACG AAATGGTGGA AGGGCGGTTG TGAGCATGAG 240
Pinb-D1i   TGATGGGGTG GTGTTTCACA ATGAAGGATT TTCCATTCAC TTGGCCCACG AAATGGTGGA AGGGCGGTTG TGAGCATGAG 240
Pinb-N1a   TGATGGAGCG GTGTTTCACA ATGAAGGATT TTCCAGTCAC TTGGCCCACG AGATGGTGGA AGGGCGGTTG TGAGCACGAG 240
                 * *  *                          *     *        *  *                *         *

Pinb-D1a   GTTCGGGAGA AGTGCTGCAA GCAGCTGAGC CAGATAGCAC CACAATGTCG CTGTGATTCT ATCCGGCGAG TGATCCAAGG 320
Pinb-D1h   GTTCGGGAGA ACTGCTGCAA GCAGCTGAGC CAGATAGCAC CACAGTGTCG CTGCGATTCT ATCCGAGGAA TGATCCAAGG 320
Pinb-D1i   GTTCGGGAGA ACTGCTGCAA GCAGCTGAGC CAGATAGCAC CACAGTGTCG CTGCGATTCT ATCCGAGGAA TGATCCAAGG 320
Pinb-N1a   GTTCGGGAGA AGTGCTGCCA GCAGCTGAGC CAGATAGCAC CACAGTGTCG CTGCGATTCT ATCCGAGGAA TGATCCAAGG 320
                       *      *                            *         *            **  *

Pinb-D1a   CAGGCTCGGT GGCTTCTTGG GCATTTGGCG AGGTGAGGTA TTCAAACAAC TTCAGAGGGC CCAGAGCCTC CCCTCAAAGT 400
Pinb-D1h   CAAGCTCGGT GGCTTCTTTG GAATTTGGCG AGGTGATGTA TTCAAAAAAA TTCAGAGGGC CCAGAGCCTC CCCTCAAAGT 400
Pinb-D1i   CAAGCTCGGT GGCTTCTTTG GAATTTGGCG AGGTGATGTA TTCAAAAAAA TTCAGAGGGC CCAGAGCCTC CCCTCAAAGT 400
Pinb-N1a   CAAGCTCGGT GGCTTCTTCG GAATTTGGCG AGGTGATGTA TTCAAACAAA TTCAGAGGGC CCAGAGACTC CCCTCAAAGT 400
             *                *   *               *          *  *                  *

Pinb-D1a   GCAACATGGG CGCCGACTGC AAGTTCCCTA GTGGCTATTA CTGGTGA 447
Pinb-D1h   GCAACATGGG AGCCGACTGC AAATTCCCTA GTGGCTATTA CTGGTGA 447
Pinb-D1i   GCAACATGGG AGCCGACTGC AAATTCCCTA GTGGCTATTA CTGGTGA 447
Pinb-N1a   GCAACATGGG AGCCGACTGC AAATTCCCTA GTGGCTATTA CTGGTGA 447
                      *            *
```

**(b)**

```
Pina-D1a   MKALFLIGLL ALVASTAFAQ YSEVVGSYDV AGGGGAQQCP VETKLNSCRN YLLDRCSTMK DFPVTWRWWK WWKGGCQELL 80
Pina-D1d   MKALFLIGLL ALVASTAFAQ YSEVVGSYDV AGGGGAQQCP VETKLNSCRN YLLDRCSTMK DFPVTWRWWK WWKGGCQELL 80
Pina-N1a   MKALFLIGLL ALVASTAFAQ YSEVVGSYDA AGGGGAQQCP VETKLDSCRN YLLDRCSTMK DFPVTWRWWK WWKGGCLELL 80
                                          *                 *                                 *

Pina-D1a   GECCSRLGQM PPQCRCNIIQ GSIQGDLGGI FGFQRDRASK VIQEAKNLPP RCNQGPPCNI PGTIGYYW 148
Pina-D1d   GECCSQLGQM PPQCRCNIIQ GSIQGDLGGI FGFQRDRASK VIQEAKNLPP RCNQGPPCNI PGTIGYYW 148
Pina-N1a   GECCSRLGQL PPQCRCNIIQ GSIQGDLGSI FGFQRDRASK VIQEAKNLPP RCNQGPPCDI RSTSGYY- 147
                *   *                    *                                *  ** *   *


Pinb-D1a   MKTLFLLALL ALVASTTFAQ YSEVGGWYNE VGGGGGSQQC PQERPKLSSC KDYVMERCFT MKDFPVTWPT KWWKGGCEHE 80
Pinb-D1h   MKTLFLLALL ALVASTTFAQ YSEVGGWYNE VGAGGSSQQC PLERPKLSSC KDYVMGWCFT MKDFPFTWPT KWWKGGCEHE 80
Pinb-D1i   MKTLFLLALL ALVASTTFAQ YSEVGGWYNE VGAGGSSQQC PLERPNLSSC KDYVMGWCFT MKDFPVTWPT RWWKGGCEHE 80
Pinb-N1a   MKTLFLLALL ALVAGTTFAQ YSEVGGWYNE VGAGGGSQQC PMERPNLSSC KDYVMERCFT MKDFPVTWPT RWWKGGCEHE 80
                          *                    *  *      *   *          **         *     *

Pinb-D1a   VREKCCKQLS QIAPQCRCDS IRRVIQGRLG GFLGIWRGEV FKQLQRAQSL PSKCNMGADC KFPSGYYW 148
Pinb-D1h   VRENCCKQLS QIAPQCRCDS IRGMIQGKLG GFFGIWRGDV FKKIQRAQSL PSKCNMGADC KFPSGYYW 148
Pinb-D1i   VREKCCQQLS QIAPQCRCDS IRGMIQGKLG GFFGIWRGDV FKQIQRAQRL PSKCNMGADC KFPSGYYW 148
Pinb-N1a   VREKCCQQLS QIAPQCRCDS IRGMIQGKLG GFFGIWRGDV FKQIQRAQRL PSKCNMGADC KFPSGYYW 148
              *  *                 **   *     *     *    **    *
```

*Figure 4.* (a) Nucleotide sequences of puroindoline alleles and (b) deduced amino acid sequences of puroindolines a and b in common wheat cv. Bolero, *Ae. tauschii* Accession L35 and *Ae. ventricosa* Accession L36. Polymorphic sites are indicated by stars.

The 3′ non-coding region of the *Pinb-D1h* allele in *Ae. ventricosa* (GenBank DQ124422) was explored by PCR amplification with the sense strand primer 'Pinb-int', which occurs at positions 149–169 in the CDS of allele *Pinb-D1h*, coupled with the antisense strand primer 'Pinb-rel', which is located 1101 bp downstream from allele *Pinb-D1i* in *Ae. tauschii* (Chantret et al. 2005). The 1400 bp PCR product obtained with these primers was 99.8% identical to the 3′ DNA boundary of allele *Pinb-D1h* in *Ae. tauschii* Accession CPI 110799 (GenBank AY159804, Turnbull et al. 2003), and contained the *Pinb-relic* sequence, a short duplication of the 5′ region of the Pin-b gene (Chantret et al. 2005).

In conclusion, *Ae. tauschii* Accession L35 was shown to possess alleles *Pina-D1d* and *Pinb-D1i*, whereas *Ae. ventricosa* Accession L36 proved to be quite unique in having alleles *Pina-D1a* and *Pinb-D1h*, along with novel puroindoline alleles *Pina-N1a* and *Pinb-N1a*.

## Discussion

Genes coding for puroindolines were found in all the diploid wheat species analysed so far, including *Aegilops* species closely related to the B-genome progenitor (Tranquilli et al. 1999; Gautier et al. 2000; Morris 2002; Massa et al. 2004). The loss of puroindoline genes from the A, B or G genomes in polyploid *Triticum* species (wild and domesticated) was claimed to be a classical example of gene elimination induced by polyploidy, and probably driven by illegitimate recombination events involving different transponsable elements (Chantret et al. 2005). These genomic rearrangements did not occur in the $D^v$ and $N^v$ genomes of wild wheat *Ae. ventricosa*. In fact, A-PAGE fractionation of starch extracts and PCR amplification of genomic DNAs provided the first evidence of a tetraploid wheat species in which four puroindoline genes are expressed. In particular, the starch granules of *Ae. ventricosa* Accession L36 were found to be associated with two proteins (*a1* and *a2)*, which reacted strongly with a Pin-a-specific antibody. Upon A-PAGE, puroindoline *a2* (Figure 1, lane 2) appeared as an intense band with a relative mobility identical to that of wild-type Pin-a encoded by allele *Pina-D1a* in common wheat cv. Bolero. Therefore, puroindoline *a2* was assumed to be encoded by *Pina-D1a* on chromosome $5D^v$ of *Ae. ventricosa* (Figure 4a). On the other hand, protein *a1*, which appeared as a faint band slightly slower than puroindoline *a2*, was assigned to the novel *Pina-N1a* allele on chromosome $5N^v$. This allele exhibited a high level of nucleotide diversity in its CDS with respect to *Pina-D1* alleles. Compared with the sole arginine-to-glutamine substitution at position 58 in the Pin-a variants observed in *T. aestivum* and *Ae. tauschii* (Morris 2002; Massa et al. 2004), puroindoline *a1* in *Ae. ventricosa* showed nine amino acid substitutions. Furthermore, one SNP in *Pina-N1a* caused the loss of the last tryptophan residue in the C-terminal region of the protein. According to Gautier et al. (1994), Pin-a is synthesized as a preprotein, and contains the cleavable YYW tripeptide at its C-terminal end. As the terminal tryptophan residue exerts an important role in the translocation of Pin-a through the membranes in developing wheat seeds (Schiffer et al. 1992), its truncation in puroindoline *a1* could have altered the information necessary for addressing this protein to the starchy endosperm, as suggested by the reduced amount of protein *a1* on starch granules. Moreover, amino acid replacements at other critical positions of puroindoline *a1* could have impaired its affinity to starch, as observed in Pin-b variants encoded by alleles *b* or *d* at *Pinb-D1* in common wheat (Corona et al. 2001a, b). Remarkably, the 5′ boundary of allele *Pina-N1a* (GenBank DQ124419) was found to be 97% identical to that of *T. monococcum* GenBank AJ302092.1 and AY622786.1.

Two Pin-b polypeptides (*b1* and *b2)* were also found to interact with the starch granules of *Ae. ventricosa*. Puroindoline *b1* (Figure 1, lane 2) was assigned to allele *Pinb-D1h* on chromosome $5D^v$ because its A-PAGE mobility was identical to that of band *b3* (Figure 1, lane 1). In fact, this latter protein in *Ae. tauschii* Accession L35 is encoded by allele *Pinb-D1i*, which differs from allele *Pinb-D1h* in a sole synonymous SNP (Massa et al. 2004). As a consequence, puroindoline *b2* was assumed to be under the genetic control of the novel *Pinb-N1a* allele on chromosome $5N^v$. The 5′ boundary of this allele was found to be 99% identical to that of *Pinb-D1i* in *Ae. tauschii* (Chantret et al. 2005). By contrast, *Pinb-N1a* differed from wild-type *Pinb-D1a* in 13 non-synonymous SNPs, one of them resulting in the change from lysine-42 to arginine in the

tryptophan-rich region of the protein. A single amino acid substitution in the tryptophan-rich domain of Pin-b in hard common wheat cultivars containing alleles *Pinb-D1b* or *Pinb-D1d* (Giroux and Morris 1997; Lillemo and Morris 2000) was found to result in a low amount of Pin-b on the surface of starch granules (Corona et al. 2001a, b). On the contrary, the interaction of puroindoline *b2* with starch granules was not affected by the lysine-to-arginine conservative mutation. The same amino acid replacement at position 42 was observed in avenoindoline b, a puroindoline-like protein from *Avena sativa* (Gautier et al. 2000).

Upon A-PAGE, puroindoline *b3* in *Ae. tauschii*, and puroindolines *b1* and *b2* in *Ae. ventricosa* moved slower than wild-type Pin-b in *T. aestivum* cv. Bolero (Figure 1). The calculated isoelectric points of *b1* and *b3* (10.4), *b2* (10.5) and wild-type Pin-b (10.7) were in agreement with this electrophoretic pattern. Similarly, the isoelectric point (10.2) of puroindoline *a3* in *Ae. tauschii* was in agreement with the reduced mobility of this protein compared with wild-type Pin-a (pI = 10.5) in cv. Bolero.

Starch-granule proteins from *Ae. tauschii* and *Ae. ventricosa* did not react with the antiserum developed against the GEVFKQLQRAQSLPSK sequence of wild-type puroindoline-b, likely because this sequence underwent to three amino acid replacements in puroindolines *b1*, *b2* and *b3* (Figure 4b).

A partial duplication of the Pin-b gene (*Pinb-relic)* was found to be located downstream from the Pin-b gene on chromosome 5D in *Ae. tauschii* and *T. aestivum* (Chantret et al. 2005). The presence of *Pinb-relic* on chromosome $5D^v$ in *Ae. ventricosa* indicated that the 3′ boundary of *Pinb-D1* has been largely conserved during phylogenesis of this species from *Ae. tauschii* and *Ae. uniaristata*.

As mentioned above, alleles *Pina-D1a* and *Pinb-D1h* of *Ae. ventricosa* were previously observed in *Ae. tauschii*. However, amongst the 50 *Ae. tauschii* accessions studied by Massa et al. (2004), allele *Pinb-D1h* was found in combination with alleles *Pina-D1c* or *Pina-D1d*. Therefore, the novel *Pina-D1a* + *Pinb-D1h* combination in *Ae. ventricosa* likely resulted from inter-locus recombination between puroindoline genes on chromosome 5D.

The growing interest of biochemists, geneticists and breeders for puroindolines is largely due to their lipid binding and foaming properties, which have been found to contribute significantly to texture and flavour of cereal food and beverage (Clark et al. 1994; Dubreil et al. 1998; Igrejas et al. 2001). Furthermore, interactions of puroindolines with starch granules and biological membranes have been related to their endosperm softening effects and antimicrobial activity, respectively (Blochet et al. 1993; Krishnamurthy et al. 2001; Morris 2002).

Diploid ancestor wheats were found to be a rich resource of genetic variation for Pin-b, whereas Pin-a proved to be highly conserved in these species (Tranquilli et al. 1999; Gautier et al. 2000; Massa et al. 2004). The present results are strong evidence that tetraploid wheat *Ae. ventricosa* could constitute an additional valuable resource of puroindoline genes. Remarkably, the genetic variation in this species was found to include Pin-a, which is the puroindoline isoform with the highest affinity for polar lipids, biological membranes and starch granules (Husband et al. 1995; Dubreil et al. 1997; Jing et al. 2003; Gazza et al. 2005). In this context it is noteworthy that *Ae. ventricosa* has been widely used in wheat breeding to transfer genes for resistance to rust (*Sr38, Yr17* and *Lr17*), cereal cyst nematode (*CreX*) or eyespot (*Pch1*) caused by *Pseudocercosporella herpotrichoides* into common or durum wheat by homologous or homoeologous recombination (Maia 1967; Jahier et al. 1978; Worland et al. 1988; Friebe et al. 1996; Seah et al. 2000; Huguet-Robert et al. 2001).

## Acknowledgements

The authors are grateful to Dr Andrea Brandolini for kindly providing seeds of *Ae. ventricosa* and to Mauro Citterich for his technical assistance in DNA sequencing.

## References

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

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

Clark D.C., Wilde P.J. and Marion D. 1994. The effect of lipid binding protein on the foaming properties of beer containing lipid. J. Inst. Brew. 100: 23–25.

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

Corona V., Gazza L., Zanier R. and Pogna N.E. 2001b. A tryptophan-to-arginine change in the tryptophan-rich domain of puroindoline b in five French bread wheat cultivars. J. Genet. Breed. 55: 187–189.

Dellaporta S.L., Wood J. and Hicks J.B. 1983. A plant DNA minipreparation: version II. Plant Mol. Biol. Rep. 1: 19–21.

Dubreil L., Compoint J.P. and Marion D. 1997. Interaction of puroindoline with wheat flour polar lipids determines their foaming properties. J. Agric. Food Chem. 45: 108–116.

Dubreil L., Méliande S., Chiron H., Compoint J.-P., Quillien L., Branlard G. and Marion D. 1998. Effect of puroindolines on the breadmaking properties of wheat flour. Cereal Chem. 75: 222–229.

Friebe B., Jiang J., Raupp W.J., McIntosh R.A. and Gill B.S. 1996. Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. Euphytica 91: 59–87.

Gautier M.F., Aleman M.E., Guirao A., Marion D. and 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 M.-F., Cosson P., Guirao A., Alary R. and 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.

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

Giroux M. and Morris C.F. 1997. A glycine to serine change in puroindoline b is associated with grain hardness and low levels of starch-surface friabilin. Theor. Appl. Genet. 95: 857–864.

Giroux M.J. and Morris C.F. 1998. Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. PNAS 95: 6262–6266.

Greenwell P. and Schofield J.D. 1989. The chemical basis of grain hardness and softness. In: Helsinki H. Salovaara (ed.), Wheat End-Use Properties. University of Helsinki and Lahti Research Training Center, pp. 59–72.

Huguet-Robert V., Dedryver F., Roder M.S., Korzun V., Abelard P., Taguy A.M., Jaudeau B. and Jahier J. 2001. Isolation of a chromosomally engineered durum wheat line carrying the *Aegilops ventricosa Pch1* gene for resistance to eyespot. Genome 44: 345–349.

Husband F., Wilde P.J., Marion D. and Clark D.C. 1995. Comparison of the foaming and interfacial properties of two related lipid-binding proteins from wheat in the presence of a competitive surfactant. In: Dickinson E. and Lorient D. (eds), Food Macromolecules and Colloids. R. Soc. Chem, London, pp. 285–296.

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

Jahier J., Doussinault G., Dosba F. and Burgeois F. 1978. Monosomic analysis of resistance to eyespot in the variety "Roazon". Proc. 5th Intl Wheat Genetics Symp., New Dehli, India, Indian Society of Genetics and Plant Breeding Publ., Indian Agricultural Research Institute, New Delhi, India, pp. 437–440.

Jing W., Demcoe A.R. and Vogel H.J. 2003. Conformation of a bactericidal domain of puroindoline a: structure and mechanism of action of a 13-residue antimicrobial peptide. J. Bacteriol. 185: 4938–4947.

Kimber G. and Zhao Y.H. 1983. The D genome of the Triticeae. Can. J. Genet. Cytol. 25: 581–589.

Krishnamurthy K., Balconi C., Sherwood J.E. and Giroux M.J. 2001. Wheat puroindolines enhance fungal disease resistance in transgenic rice. Mol. Plant–Microbe Interact. 14(10): 1255–1260.

Krishnamurthy K. and Giroux M.J. 2001. Expression of wheat puroindoline genes in transgenic rice confers grain softness. Nat. Biotechnol. 19: 162–166.

Lillemo M. and Morris C.F. 2000. A leucine to proline mutation in puroindoline b is frequently present in hard wheats from Northern Europe. Theor. Appl. Genet. 100: 1100–1107.

Maia N. 1967. Obtention de blés tendres résistants au piétin-verse (*Cercosporella herpetrichoides*) par croissement interspécifiques. C.R. Acad. Sci. Fr. 53: 149–154.

Massa A.N., Morris C.F. and Gill B.S. 2004. Sequence diversity of puroindoline-a, puroindoline-b, and the grain softness protein genes in *Aegilops tauschii* Coss. Crop Sci. 44: 1808–1816.

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

Pogna N.E., Gazza L., Corona V., Zanier R., Niglio A., Mei E., Palumbo M. and Boggini G. 2002. Puroindolines and kernel hardness in wheat species. In: Ng P.K.W. and Wrigley C.W. (eds), Wheat Quality Elucidation. AACC, St. Paul, Minnesota, USA, pp. 155–169.

Redaelli R., Morel M.-H., Autran J.-C. and Pogna N.E. 1995. Genetic analysis of low $M_r$ glutenin subunits fractionated by two-dimensional electrophoresis (A-PAGE×SDS-PAGE). J. Cereal Sci. 21: 5–13.

Seah S., Spielmeyer W., Jahier J., Sivasithamparam K. and Lagudah E.S. 2000. Resistance gene analogs within an introgressed chromosomal segment derived from *Triticum ventricosum* that confers resistance to nematode and rust pathogens in wheat. Mol. Plant–Microbe Interact. 13: 334–341.

Schiffer M., Chang C.H. and Stevens F.J. 1992. The functions of tryptophan residues in membrane proteins. Protein Engin. 5: 213–214.

Tranquilli G., Lijavetzky D., Muzzi G. and Dubcovsky J. 1999. Genetic and physical characterization of grain texture-related loci in diploid wheat. Mol. Gen. Genet. 262: 846–850.

Turnbull K.-M., Gaborit T., Marion D. and Rahman S. 2000. Variation in puroindoline polypeptides in Australian wheat cultivars in relation to grain hardness. Aust. J. Plant Physiol. 27: 153–158.

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

Worland A.J., Law C.N., Hollins T.W., Koebner R.M.D. and Guira A. 1988. Location of a gene for resistance to eyespot (*Pseudocercosporella herpotrichoides*) on chromosome 7D of bread wheat. Plant Breed. 101: 43–51.