Research & Reviews: Journal of Botanical Sciences

e-ISSN:2320-0189
p-ISSN:2347-2308

Genetic and Bioinformatic Characterization of Puroindoline A and Puroindoline B in Italian Wheat Cultivars

Bruna De Felice*, Francesco Manfellotto and Raffaella D'Alessandro

DISTABIF- Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, University of Campania Luigi Vanvitelli Caserta, Italy

Research Article

Received: 19/07/2017
Accepted: 24/07/2017
Published: 29/07/2017

***For Correspondence**

Bruna De Felice, DISTABIF-Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, University of Campania Luigi Vanvitelli Caserta, Italy.
Tel: ++39-823-274543.

E-mail: bruna.defelice@unina2.it

Keywords: Puroindoline genes, Wheat, Grain texture, Genetic analysis, Bioinformatics.

ABSTRACT

Wheat kernel texture is a critical property influencing wheat raw material processing and end use products features. Wheat grain texture is a genetically determined and inherited trait. Differences among wheat cultivars reside in variable puroindoline a and b (*Pina* and *Pinb)* allelic forms, leading to a continuous range of variable grain hardness. Genetic determination of *Pina* and *Pinb* sequences in a large number of international cultivars has led to the assessment of genetic polymorphisms on wheat grain texture. Here we report the identification of *Pina-D1* and *Pinb-D1* allelic polymorphism in twenty-eight Italian soft wheat cultivars (*Triticum aestivum* L.) and ten Italian durum wheat cultivars (*Triticum turgidum* L. ssp. *durum*). In contrast to previous studies, our results revealed the presence of *Pinb-D1b, Pinb-D1d* and *Pina-D1b* (deletion allele) alleles in Italian cultivars. Since only some puroindoline a and b alleles have been investigated for their direct effect on wheat grain texture, we performed a bioinformatic analysis of puroindoline polymorphisms, to predict the possible impact of the biological function of such proteins on wheat. This investigation could provide useful information to extend the current knowledge about genetic determinants of kernel hardness in soft wheat and indications for breeding and food product applications.

INTRODUCTION

Cereals are human nutrition main components. Different wheat uses request a classification of wheat properties focusing on kernel texture, as "soft" (or bread) wheat (*Triticum aestivum* L.) or as "durum" (very hard) wheat (*Triticum turgidum* L. ssp. *durum*) [1]. However, soft wheat varieties can be further classified into subclasses due to their different texture hardness, ranging from soft to hard. Wheat kernel texture is a critical feature for wheat classification and food industry destination. The common, hexaploid wheat cultivars (*Triticum aestivum* L.) are used for breads, cookies, cakes, and pastries, while the very hard tetraploid cultivars, derived from *Triticum turgidum var. durum* (AABB), are mainly used for Italian-style pastas.

Grain texture is an inherited trait and the principal genes involved in grain texture determination were found to be located in the short arm of chromosome 5D (specifically found in *T. aestivum*). That locus was named Hardness (Ha) locus and is known to control grain hardness in wheat. The softness phenotype is recognized as the dominant trait [2]. This locus contains the puroindoline a (*Pina*) and puroindoline b (*Pinb*) genes (representing the major determinants of wheat kernel texture), which are the subunits of a roughly 13 kDa friabilin, and grain softness protein gene (*Gsp-1*).

PINA and PINB proteins determine grain texture binding the surface of starch granules in the endosperm cells and forming the previously described protein complex. The soft phenotype is determined by the wild-type alleles, named *Pina-D1a* and *Pinb-D1a*, of the Puroindoline a (Pina) and Puroindoline b (Pinb) genes. *Pina-D1* and *Pinb-D1* genes coding sequences are intronless and 447 bp long sharing 70.2% coding sequence. Such genes encode wheat endosperm-specific lipid binding proteins 148 amino acids long with a cysteine-rich backbone and a unique tryptophan-rich domain, which was considered as being responsible for the strong affinity of the Puroindoline-D1 protein to polar lipids [3].

The hard texture phenotype results from various mutations in either one or both of the puroindoline genes or complete absence of puroindoline coding genes, respectively [4]. Indeed, it has been shown that RNA interference (RNAi)-based silencing of Pina and Pinb genes significantly decreased the puroindoline a and puroindoline b proteins in wheat and essentially increased

e-ISSN:2320-0189
p-ISSN:2347-2308

grain hardness [5]. All wheat cultivars to date that have mutations in *Pina* or *Pinb* genes are hard textured, while wheat possessing both the 'soft type' *Pina-D1a* and *Pinb*-D1a sequences are soft textured.

The genetic bases of puroindoline absence in durum wheat is primarily determined by the evolutive events that have led to the speciation of durum and soft wheat. Durum wheat (*Triticum turgidum*) originated about 500,000 years ago from a hybridization event between *Aegilops speltoides* and *Triticum urartu*, resulting in a tetraploid species (genome AABB). A subsequent hybridization event (occurred about 8,000 years ago) between *Triticum turgidum* and *Aegilops tauschii* resulted in the origin of the hexaploid common wheat (*Triticum aestivum*, genome AABBDD). Current evidence suggest that a deletion event occurred at Ha locus during hybridization, leading to the complete absence of those genes in durum wheat. Instead, *Pina* and *Pinb* genes have been provided to common wheat genome by *A. tauschii* (D chromosome series) [6]. Conversely, the third gene belonging to the Ha locus, *Gsp-1*, has been conserved also in durum wheat. The presence of GSP genes and their probable expression in durum wheat suggest a secondary role in determining grain texture [7].

Puroindoline genes arouse scientific interest worldwide for their kernel texture conferring properties. Their investigation in several wheat cultivars has led to the identification of a large number of alleles for both *Pina-D1* and *Pinb-D1* genes [8]. Following their identification, puroindoline genes have been further investigated for their specific correlation to soft wheat kernel texture [9]. In addition to their texture-related function, puroindolines have also been recognized as a-amylase inhibitors (AAIs) and Seed Storage (SS) proteins, having a possible role in protecting plants against pathogens with their bactericidal and fungicidal activities [10,11].

As the largest producer and consumer of “soft and hard“ wheat in the entire world, Italy is a center of wheat, holding a highly diverse stock of wheat germplasm, proving useful in both applied and basic research efforts to give insight into the biology of wheat plant. Germplasm and the scientific method of breeding provide the foundation for bountiful wheat harvests.

In our research, we aimed to investigate *Pina-D1* and *Pinb-D1* allele polymorphism in Italian *Triticum aestivum* and *Triticum turgidum* ssp. *durum* cultivars, to evaluate possible genetic differences related to phenotypic variability in kernel texture. For this purpose, we amplified and sequenced *Pina-D1* and *Pinb-D1* genes from each cultivar and performed bioinformatic analyses to assess possible phenotypic consequences on grain texture.

METHODS

Plant Material

Ten Italian *Triticum turgidum* ssp. *durum* and 28 Italian *Triticum aestivum* grain cultivar were provided by (CRA-SCS), Battipaglia (Salerno, Italy) and CRA–SCS (Verona, Italy). For each cultivar, grains were grounded and leaves were used to extract the DNA.

Fresh leaves form wheat cultivars were collected and immediately grinded in liquid nitrogen. Samples were stored at -80°C until DNA extraction was performed.

DNA Extraction from Wheat Leaves

DNA was extracted from 1.6 grams of grinded wheat leaves following the CTAB extraction method [12]. DNA yield and purity was assessed a using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). Purity of the extracted DNA was based on the 260/280 and on the 260/230 O.D. ratios. DNA integrity was assessed running extracted samples on 1% (w/v) agarose gel stained with ethidium bromide.

PCR Amplification of Puroindoline and Control Genes

PCR amplifications were carried out on each DNA sample. As positive control, *Gsp1* gene was amplified (because of its presence also on A- and B- wheat genomes, other than on D-genome). *Gsp1* primer sequences were Gsp-1F-5’-CTTCCTCCTAGCTTTCCTTG-3’ and Gsp-1R- TAGTGATGGGGATGTTGCAG-3’ [13]. Primers used for amplification of *Pina-D1* and *Pinb-D1* were Pina-F-5’-ATGAAGGCCCTCTTCCTCA-3’, Pina-R-5’-TCACCAGTAATAGCCAATAGTG-3’, Pinb-F- 5’-ATGAAGACCTTATTCCTCCTA-3’, Pinb-R- 5’-TCACCAGTAATAGCCACTAGGGAA-3’, respectively [3]. In samples missing *Pina-D1* amplification, a second pair of primers was used to assess the presence of *Pina*-null allele (*Pina-D1b*). Their sequences were: Pina-D1b-F-5’-AATACCACATGGTTCTAGATACTG-3’ and Pina-D1b-R- GCAATACAAAGGACCTCTAGATT-3’ [14].

PCR reactions were performed in 25 µL total volume containing 25 pmol of each primer, 250 µM of each dNTP, 1x PCR buffer, 1.5 mM of $MgCl_2$, 0,5 units of Taq DNA polymerase (Promega, USA) and 100 ng of DNA template. *GSP1*, *Pina-D1* and *Pinb-D1* amplification program comprised an initial denaturation at 95°C for 5 min, followed by 35 cycles of 95°C for 40 s, 55°C for 50s, 72°C for 1min and a final extension at 72°C for 7min. *Pina-D1b* amplification program comprised an initial denaturation at 95°C for 5min, followed by 35 cycles of 95°C for 40s, 57°C for 50s, 72°C for 1min and 72°C for 7min.

The expected *Gsp1* amplicon was 467 base pairs (bp) in length, while *Pina-D1* and *Pinb-D1* were 447 bp in length and *Pina-D1b* was 778 bp. PCR amplification was subsequently assessed running the amplified product on 1,2% (w/v) agarose gel stained with ethidium bromide and comparing amplicon size with a DNA size standard for gel electrophoresis (DNA Ladder 100 bp plus, Applichem, Darmstadt, Germany; DNA molecular weight marker XVII 500 bp ladder, Roche Diagnostics, Mannheim, Germany).

e-ISSN:2320-0189
p-ISSN:2347-2308

Extracted DNA fragments were re-amplified by PCR, cloned using TA Cloning kit (Invitrogen, USA) and sequenced. Each *Pina* and *Pinb* PCR product was cut from the gel, purified with the Qiaquick Gel Extraction Kit (Qiagen, Hilden, Germany), cloned into the pMOSBlue vector, pMOSBlue blunt ended cloning kit, (Amersham, Piscataway, NJ, USA), and used to transform the DH5α competent cells (Stratagene, La Jolla, CA). Inserts of the desired size were evaluated using PCR with T7 and M13 primers.

DNA Sequence Analysis

Recombinant clones were sequenced from both strands using the Big Dye Terminator v3.1 Cycle Sequencing Kit via the automatic sequencing system ABI PRISM 3100 automatic DNA sequencer (Applied Biosystems, Foster City, CA).

Restriction Sites Allele Identification

To assess the identity of the specific allele polymorphism a restriction assay was performed on *Pinb-D1* amplicons, using *SapI* restriction enzyme (predicted to cut only the *Pinb-D1b* allele) (Thermo Scientific, USA). We digested 10 µL of amplicon, 5 units of *SapI* enzyme, 2 µL of buffer and incubated for 16 hours at 37°C. Inactivation was conducted at 65°C for 2 hours. Gel electrophoresis was performed on a 3% agarose gel stained with ethidium bromide. To detect the two digested fragments, and compared to pBR322 DNA-*MspI* digest molecular marker (New England Biolabs, Ipswich, MA, USA).

Bioinformatic Analyses

NE cutter V2.0 was used to evaluate restriction enzyme sites maps in each known *Pina-D1* and *Pinb-D1* alleles with the aim to evaluate differences in the presence/absence or in the position of restriction sites. Such evaluation allowed to perform specific PCR amplicon digestions in order to confirm the presence of allelic variants. The nature of cloned sequences was confirmed by performing computer-based similarity searches with known *Pina* and *Pinb* sequences in the NCBI database using BLASTN algorithm [15] to assess the specificity of the amplified product against the entire database.

Direct comparison between wild type puroindoline sequences *Pina-D1a* and *Pinb-D1a* and sequenced products was performed using Clustal W program [16] and verifying, for each sequence position, the nucleotide correspondence to the reference allele. The nucleotide sequences were translated using the ExPASy translation tool and translated sequences were submitted to PROVEAN software tool to predict whether amino acid substitutions or indels could have an impact on the biological function of the analysed protein. Cut-off value to predict a protein alteration as deleterious was set at -2.50 [17].

RESULTS

Gsp-1, Pina-D1 and Pinb-D1 Genes Amplification

All analysed wheat DNA samples gave *GSP*-1 gene amplification (the above mentioned specific 467 bp amplicon), confirming the suitability of extracted DNA for puroindoline genes investigation. Puroindoline gene amplification was confirmed in each soft wheat DNA sample. **Figure 1** shows the amplification results obtained for Claudio, Guadalupe and Adelaide cultivars.

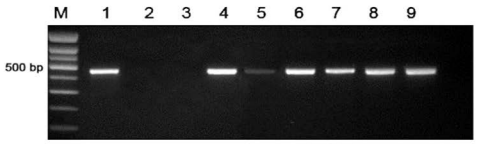

Figure 1. Typical *Gsp-1, Pina-D1* and *Pinb-D1* genes amplification in *Triticum aestivum* and *Triticum turgidum* ssp. durum wheat samples. Lane M: 100 bp ladder; Lanes 1, 2 and 3: *Claudio durum* wheat *Gsp-1, Pina* and *Pinb* amplification; Lanes 4, 5 and 6: Guadalupe bread wheat, *Gsp-1*, *Pina* and *Pinb* amplification; Lanes 7, 8 and 9: Adelaide bread wheat, *Gsp-1*, *Pina* and *Pinb* amplifications.

As we did not get any *Pina-D1* amplification in some of the analysed cultivars, we performed a PCR assay to specifically evaluate the presence of null *Pina-D1* allele (*Pina-D1b*) **(Figure 2)**.

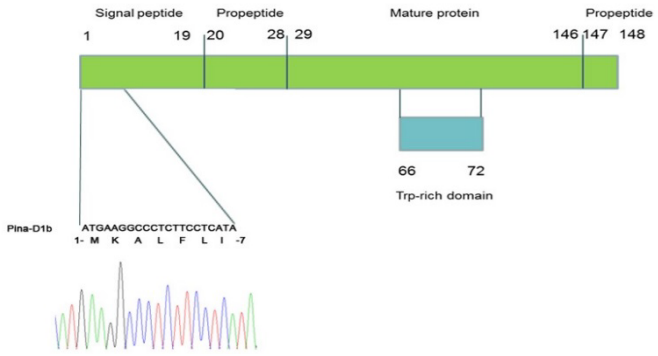

Figure 2. Domain structure of *Pina-D1* null allele variant from *Triticum aestivum* (Profeta and Tirex cultivar). Deduced amino acid sequence is shown for the 23-nucleotide long *Pina* coding sequence belonging to *Pina-D1b* null allele.

e-ISSN:2320-0189
p-ISSN:2347-2308

The cultivars assessed for the large deletion were all durum wheat cultivars. Moreover, as Guadalupe cultivar was previously reported to lack *Pina-D1* amplification, we performed a *Pina-D1b* specific PCR on that cultivar. Beside Guadalupe, we got *Pina-D1b* amplicon for Amidon, Profeta, Ciano and Sibilia cultivars [12]. **Figure 3** shows *Pina-D1b* amplification positive results.

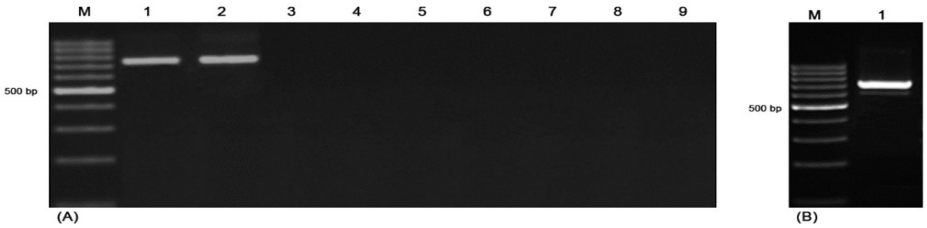

Figure 3. (A) *Pina-D1b* (null allele) amplification in *Profeta (Triticum aestivum*) and six *Triticum turgidum ssp*. durum wheat samples. Lane 1-10: Profeta and Sibilia (bread wheats), Cirillo, Giemme, Saragolla, Claudio, Core, Creso and Zenit *Pina-D1b* amplification; Lane M: Marker XVII 500 bp (Roche). **(B)** Pina-D1b (null allele) amplification in Guadalupe cultivar. Lane 1: Guadalupe *Pina-D1b* amplification; Lane M: Marker 100 bp.

Pina and Pinb Genes Sequence Analysis

In order to confirm the specificity of *Pina-D1, Pinb-D1* and *Pina-D1b* amplification and the specific allelic form, we excised from agarose gels, cloned and sequenced the amplicons of interest. BLASTN analyses revealed that the obtained amplifications were specific for the expected genes when compared to public sequence databases.

All amplified sequences were aligned with the reference alleles (*Pina-D1a* and *Pinb-D1a*) using ClustalW program to evaluate the specific genotype and the translated sequence was obtained using ExPASy translation tool. Among soft wheat cultivars, the most common genotypes resulted *Pina-D1a/Pinb-D1b* (found in Adelaide, Pandas, Colfiorito and Dorico cultivars) and *Pina-D1a/Pinb-D1d* (Aubusson and Bologna), but *Pina-D1a-b/PinbD1a* (Guadalupe) genotype was also found (we obtained positive results for both *Pina-D1a* and *Pina-D1b* amplifications). Besides, five soft wheat cultivars, Amidon, Profeta, Ciano, Guadalupe and Sibilia showed the presence of the *Pina-D1* null allele (*Pina-D1b/Pinb-D1a* genotype).

Table 1. shows the identified *Pina-D1/Pinb-D1* genotypes for each *T. aestivum* and *T. turgidum* ssp. durum wheat. *Pina-D1* sequences were found to be corresponding to the wild type allele (*Pina-D1a*, Genbank DQ363911) for all remaining bread wheat cultivars, so that the predicted codified amino acidic sequence is identical to the wild type puroindoline a protein. *Pinb-D1* amplicons sequencing revealed a higher level of heterogeneity in investigated cultivars, as two different alleles were identified beside the wild-type in *Triticum aestivum*. Indeed, most analysed cultivars showed the wild type version of *Pinb-D1* allele when aligned with the reference sequence (*Pinb-D1a*, Genbank DQ363913). Their predicted codified protein sequence was corresponding to the wild type version of the protein *Pinb*.

Table 1. *Pina-D1/Pinb-D1* genotypes for investigated wheat cultivars.

Wheat cultivar	Pina-D1 genotype Triticum turgidum ssp. durum	Pinb-D1 genotype
Claudio	none	none
Cirillo	none	none
Cosmodur	none	none
Core	none	none
Creso	none	none
Giemme	none	none
Grazia	none	none
Neodur	none	none
Saragolla	none	none
Zenit	none	none
Triticum aestivum		
Adelaide	*Pina-D1a*	*Pinb-D1b*
Amidon	*Pina-D1b (null allele)*	*Pina-D1a*
Aubusson	*Pina-D1a*	*Pinb-D1d*
Aurelio	*Pina-D1a*	*Pinb-D1a*
Bolero	*Pina-D1a*	*Pinb-D1a*
Bologna	*Pina-D1a*	*Pinb-D1d*
Centauro	*Pina-D1a*	*Pinb-D1a*
Ciano	*Pina-D1b (null allele)*	*Pinb-D1a*
Colfiorito	*Pina-D1a*	*Pinb-D1b*
Dorico	*Pina-D1a*	*Pinb-D1b*
Francia	*Pina-D1a*	*Pinb-D1a*

e-ISSN:2320-0189
p-ISSN:2347-2308

Genio	*Pina-D1a*	*Pinb-D1a*
Gladio	*Pina-D1a*	*Pinb-D1a*
Guadalupe	*Pina-D1a/Pina-D1b*	*Pinb-D1a*
Lampo	*Pina-D1a*	*Pinb-D1a*
Leone	*Pina-D1a*	*Pinb-D1a*
Libero	*Pina-D1a*	*Pinb-D1a*
Livio	*Pina-D1a*	*Pinb-D1a*
Mosè	*Pina-D1a*	*Pinb-D1a*
Neviano	*Pina-D1a*	*Pinb-D1a*
Oscar	*Pina-D1a*	*Pinb-D1a*
Pascal	*Pina-D1a*	*Pinb-D1a*
Sagittario	*Pina-D1a*	*Pinb-D1a*
Pandas	*Pina-D1a*	*Pinb-D1b*
Profeta	*Pina-D1b* (null allele)	*Pinb-D1a*
Salgemma	*Pina-D1a*	*Pinb-D1a*
Serena	*Pina-D1a*	*Pinb-D1a*
Sibilia	*Pina-D1b* (null allele)	*Pinb-D1a*

However, other cultivars showed nucleotide mismatches compared to the wild type allele and, specifically, Colfiorito, Dorico, Adelaide and Pandas showed a sequence corresponding to *Pinb-D1b* allele (G223A nucleotide change, Genbank DQ363914) and Aubusson and Bologna cultivar showed the presence of *Pinb-D1d* allele (T217A, sequence not deposited in Genbank,) **(Figure 4)**.

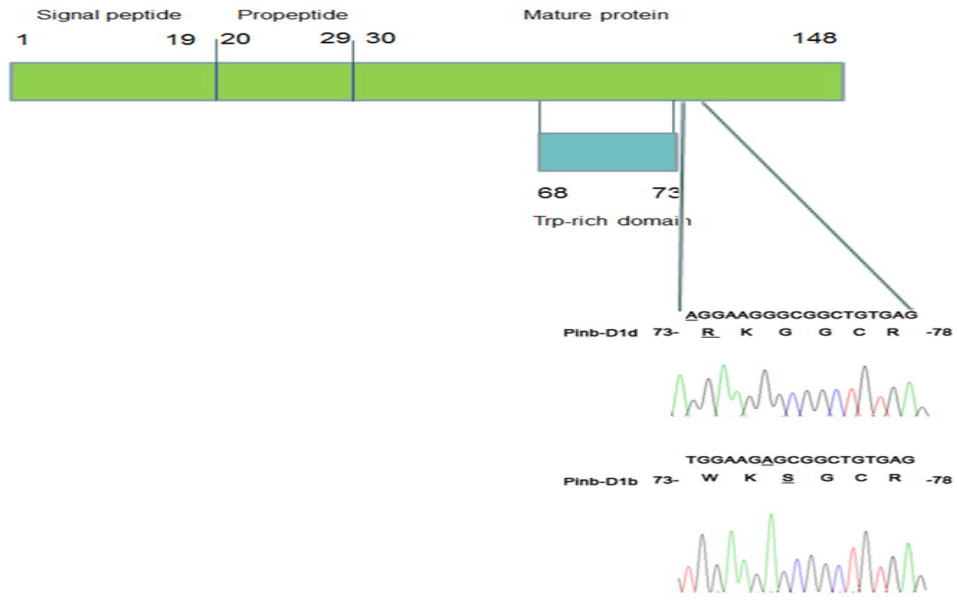

Figure 4. Domain structure of the *Pinb-D1b* and *Pinb-D1d* allelic variants from *Triticum aestivum.* Colfiorito, Dorico, Adelaide and Pandas cultivar showed the presence of *Pinb-D1b* allele (G223A nucleotide change) while Aubusson and Bologna cultivar *Pinb-D1d* allele variant (T217A) leading to G75S and W73R aminoacidic changes respectively in the predicted codified protein.

Those nucleotide changes are causative of an aminoacidic change in the predicted codified protein and, in details, *Pinb-D1b* leads to W73R and *Pinb-D1d* to G75S amino acidic changes **(Figure 5)**.

Aminoacid position	1	2	3	4	5	6	7	.....	72	73	74	75	76	77	78	79	80	.....	146	147	148	149
PINB	M	K	T	L	F	L	L	...	K	W/R	K	G/S	G	C	E	H	E	...	Y	Y	W	*
Pinb-D1a	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Adelaide	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	AGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Colfiorito	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	AGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Dorico	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	AGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Pandas	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	AGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Duilio	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Guadalupe	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Levante	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Aubusson	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	TGG	AAG	AGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Bologna	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	TGG	AAG	AGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
Tirex	ATG	AAG	ACC	TTA	TTC	CTC	CTA	...	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	...	TAT	TAC	TGG	TGA
	***	***	***	***	***	***	***		***	***	***	***	***	***	***	***	***		***	***	***	***
Nucleotide position	123	456	789	111 012	111 345	111 678	122 901	...	222 111 456	222 112 789	222 222 012	222 222 345	222 222 678	222 233 901	222 333 234	222 333 567	222 334 890	...	444 333 678	444 344 901	444 444 234	444 444 567

Figure 5. *Pinb-D1b* reference allele nucleotide sequence (Genbank DQ363914) and *T. aestivum ssp.* durum amplified sequences. Deduced *Pinb* amino acid sequence is shown. Nucleotide sequences variations underlined and aminoacidic sequences showed.

e-ISSN:2320-0189
p-ISSN:2347-2308

Pinb-D1b allele amplification was confirmed performing *SapI* restriction digestions on *Pinb-D1* amplicons. Restriction site analyses predicted the presence of a specific site recognized by that enzyme only in *Pinb-D1b* allele, arising from the polymorphism at nucleotide 223 and resulting in a single cut in position 214 of the amplicon. As expected, only in Colfiorito, Dorico, Adelaide and Pandas *Pinb-D1* amplicons were digested and agarose gel electrophoresis showed two fragments of 214 bp and 233 bp **(Figure 6)**.

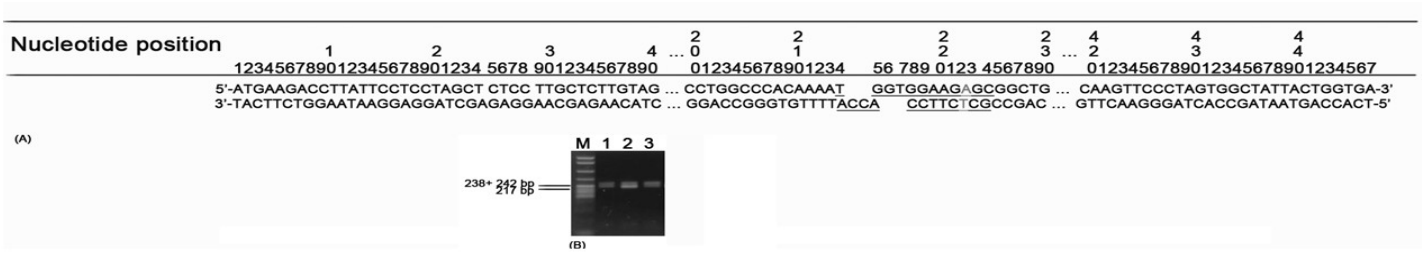

Figure 6. *Pinb-D1b* allele digested by *SapI* restriction enzyme. **(A)** Nucleotide position, forward and reverse sequences are reported. *SapI* cutting site is shown as an interruption in the sequence and nucleotide sequence recognized by *SapI* is underlined. *Pinb-D1b* polymorphic site is reported in grey. **(B)** *SapI* restriction analysis of *Pinb-D1b* allele. Lane M: pBR322 DNA-MspI digest molecular marker; Lane 1: Levante *Pinb* amplicon digestion; Lane 2: Adelaide *Pinb* amplicon digestion; Lane 3: Pandas *Pinb* amplicon digestion.

Protein Function Prediction Analysis

Protein function prediction effect of published wheat-related genetic polymorphisms leading to an amino acidic change was assessed using PROVEAN software, after translating nucleotide sequences into their corresponding protein sequences. In this analysis, all known *Pina-D1* and *Pinb-D1* gene variation determining an amino acidic change in wheat were included.

Bioinformatic prediction analysis has assessed whether PINA and PINB amino acid substitutions or indels had an impact on the biological function of the protein. The prediction was based on the change, caused by a variation, in the similarity of query sequence to closely related sequences collected through BLAST. Prediction results allowed the identification of 5 deleterious *Pina-D1* (*Pina-D1b, Pina-D1l, Pina-D1m, Pina-D1n, Pina-D1p*) and 9 deleterious *Pinb-D1* amino acidic alterations (*Pinb-D1aa, Pinb-D1ab, Pinb-D1c, Pinb-D1e, Pinb-D1g, Pinb-D1p, Pinb-D1r, Pinb-D1s, Pinb-D1u*) **(Table 2)**.

Table 2. Missense *Pina-D1*/*Pinb-D1* wheat alleles and predicted effect on codified protein function (n.a.=not available). Multiple sites mutations in the same polypeptide have been analysed as individual mutations. *in association with the wild type allele of the other puroindoline gene if not otherwise specified. fs=frameshift. Aminoacidic positions are numbered starting from the initial methionine.

Allele	Nucleotide variation	Aminoacidic variation	Provean Score	Protein variation effect	Previously reported phenotype*
Wheat-related Pina-D1 alleles					
Pina-D1b	Deletion from nt 24 to the end	Deletion from aa 8 to the end	-212.987	Deleterious	Hard texture, no PINA protein on starch granules and reduced PINB
Pina-D1k	Complete *Pina-D1* deletion	-	n.a.	n.a.	Hard texture, absence of PINA and PINB proteins
Pina-D1l	265Cdel	K89fs	-94.340	Deleterious	Hard texture, PINA null
Pina-D1m	C187T	P63S	-3.298	Deleterious	Hard texture
Pina-D1n	G212A	W71X	-3.909	Deleterious	Hard texture, PINA null
Pina-D1p	T38A	V13E	-2.306	Neutral	n.a.
	410delC	C138fs	-11.15	Deleterious	
Pina-D1q	C417A, A418C	N139K, I140L	-2.009	Neutral	n.a.
Wheat-related Pinb-D1 alleles					
Pinb-D1aa	C96A, 213delA	K71fs	-126.926	Deleterious	n.a.
Pinb-D1ab	C382T	Q128X	-2.844	Deleterious	n.a.
Pinb-D1b	G223A	G75S	-0.326	Neutral	Hard grain texture; low levels of starch surface friabilin
Pinb-D1c	T266C	L89P	-6.496	Deleterious	Hard endosperm
Pinb-D1d	T217A	W73R	-1.289	Neutral	Medium- hard texture
Pinb-D1e	G204A	W68X	-3.287	Deleterious	Hard texture, PINB null
Pinb-D1f	G219A	W73X	-2.226	Neutral	Hard texture, PINB null
Pinb-D1g	C255A	C85X	-10.136	Deleterious	Hard texture, PINB null
Pinb-D1l	A220G	K74E	-0.693	Neutral	n.a.
Pinb-D1p	213delA	K71fs	-126.926	Deleterious	Hard texture, PINB null
Pinb-D1q	G218T	W73L	-2.258	Neutral	Hard texture
Pinb-D1r	128insG	E43fs	-161.281	Deleterious	Hard texture, PINB null
Pinb-D1s	128insG, G204A	E43fs	-161.412	Deleterious	n.a.

e-ISSN:2320-0189
p-ISSN:2347-2308

Pinb-D1t	G226C	G76R	-1.058	Neutral	Hard texture
Pinb-D1u	127delG	E43fs	-166.639	Deleterious	Mixed or hard texture
Pinb-D1v	G22A	A8T	-2.243	Neutral	Hard texture
	C25A	L9I	-0.879	Neutral	
Pinb-D1w	G431T	S144I	-0.297	Neutral	Hard texture

As expected, the most deleterious protein alteration resulted *Pina-D1b* leading to a very large deletion of PINA sequence (only the first seven amino acids are codified by the deleted gene). Conversely, other already known amino acidic changes found in our cultivars (*Pinb-D1b* and *Pinb-D1d*), also if leading to an alteration in the protein sequence (G75S and W73R), were predicted to have a neutral phenotypic effect. Restriction enzyme site prediction allowed to identify a unique restriction site in *Pinb-D1b* variant, i.e., *SapI* restriction site cutting at nucleotide position 214 of *Pinb-D1b* amplicon and resulting in a single cut in the amplicon and giving as results 214 bp and 233 bp long fragments.

DISCUSSION

Puroindolines role in determining wheat grain softness has been investigated in recent years. The presence of variations in one or both puroindolines (or the absence of puroindoline proteins) leads to a hard phenotype, ranging from the soft phenotype to the extremely hard phenotype encountered in durum wheat [18]. This model recognizes the softening effect of puroindoline proteins.

Up today, many *Pina-D1* and *Pinb-D1* alleles have been identified in different geographic bread wheat cultivars from around the world. In wheat 7 different *Pina* alleles and 17 different *Pinb* alleles have been described. [19]. The most frequent *Pina* allele leading to the hard phenotype is a large deletion encompassing most of *Pina* coding sequence (the deletion starts after the 23th nucleotide from the initial ATG, encompassing 15,380 bp), that is the *Pina.D1b* allele, while the most frequent_encountered *Pinb* variant is *PinB-D1b*. In most geographic regions, soft wheat cultivars with the *Pinb-D1b* allele are commonly found, while *Pinb-D1p* allele is prevalent in Chinese cultivars [4,20]. *Pina-D1* sequence point variations are less frequently encountered, but up to date three SNP alleles leading to an aminoacidic change in the mature protein sequence (*Pina-D1m, n* and *q*), a SNP and single nucleotide deletion allele (determining a change in the leader peptide and causing a frameshift in the coding sequence) (*Pina-D1p*), one single nucleotide deletion allele (*Pina-D1l*) and one complete gene deletion (*Pina-D1k*) have been described [21].

Pinb-D1 sequence variations are more numerous and frequently encountered than *Pina-D1* alleles. In details, 8 SNPs alleles leading to an amino acidic change in leader peptide or in the mature protein sequence (*Pinb-D1b, Pinb-D1c, Pinb-D1d, Pinb-D1l, Pinb-D1q, Pinb-D1t, Pinb-D1v, Pinb-D1w*), 4 SNPs alleles leading to a nonsense mutation in the mature protein sequence (*Pinb-D1e, Pinb-D1f, Pinb-D1g, Pinb-D1ab*), 5 single nucleotide deletions creating a frameshift in the protein sequence (*Pinb-D1p, Pinb-D1r, Pinb-D1s, Pinb-D1u, Pinb-D1aa*) [22].

In addition to the 17 described sequence variations, complete *Pinb-D1* deletion has been described [23]. Even if frequently the hard phenotype encountered in *T. aestivum* cultivars is determined by alterations in *Pina* or *Pinb* sequences, some reports of contemporary deletion of both genes has been described [19]. A comparative analysis of the resulting phenotype expressed when *Pina* and *Pinb* are mutated has been provided. The “null allele” *Pina-D1b* is responsible for a harder phenotype than *Pinb-D1b and Pinb-D1p* in 2005 and *Pina-D1m* is reported to induce a harder phenotype than *Pina-D1b* (null allele) mutation [12, 13].

Moreover, cultivars with *Pina-null/Pinb-null* alleles show the highest hardness index when compared to different combinations of *Pina-D1/Pinb-D1* alleles [24]. Therefore, correlation between kernel texture and *Pina* and *Pinb* genotype can provide useful information for the improvement of wheat quality, and to extend the current understanding of the molecular and genetic mechanisms determining wheat grain texture. Puroindoline gene expression and protein quantification analyses suggested that the presence of PINA directly affects wheat grain hardness, while PINB influences the same phenotypic feature affecting polypeptide three-dimensional stability and its affinity for polar lipids [25]. According to the critical role of puroindoline a, *Pina* transcription and protein synthesis have been found to be higher than *Pinb* in soft wheat [14].

Contrasting data emerged by transcription level analyses of *Pina-D1* or *Pinb-D1* genes. They have shown no differences in soft and *Pina-D1a/Pinb-D1b* hard wheat in some studies, while *Pina-D1a* mRNA expression has been found as reduced in *Pinb-D1b* hard wheat compared to soft in the research of Capparelli. PINB protein level was drastically reduced in PINA null (*Pina-D1b/Pinb-D1a*) hard wheat. However, the absence of puroindoline a (determined by both complete absence of puroindoline a gene or *Pina-D1b* allele) has been previously demonstrated to exert a critical effect in a drastic reduction of puroindoline b expression. Although PINA and PINB are currently supposed to be the principal molecules involved in grain texture determination in wheat, other factors could have a role in that complex trait definition. Indeed, Puroindoline b-2 genes, similarly to *Gsp-1*, have been identified also in durum wheat, suggesting the presence of a multigene family in the A- and/or B-genome, The current knowledge about puroindoline involvement in wheat texture determination has posed the basis for the comprehension of genetic determinants of this important phenotypic trait, but emerging data about puroindoline gene duplications, puroindoline-related genes identification and their regulated expression control suggest that more complex molecular mechanisms are involved [26].

e-ISSN:2320-0189
p-ISSN:2347-2308

Italy, among the Mediterranean countries, has the longest tradition in wheat breeding and its germplasm can be considered as one of the richest and most valuable. Since in Italy wheat germplasm has been collected starting from the 1947, our investigation, aimed to identify novel puroindoline gene alleles characterizing Italian wheat cultivars. For the first time, we found that three durum wheat cultivars had null *Pina* allele (complete absence of amplification for Duilio and Levante). We investigated the presence of null *Pina* allele in cultivars lacking *Pina* amplification. The specific study of *Pina-D1b* allele amplification and the subsequent sequence analysis of the obtained amplicon has allowed the detailed identification of *Pina* null allele, providing the confirmation that the lack of *Pina-D1* gene amplification is due to the presence of that specific allele and no other genomic deletions encompassing *Pina-D1* gene. Interestingly, we obtained, for the first time, different results from Corona, which reported lack of amplification of *Pina-D1* in Guadalupe cultivar and consequently assigned the *Pina-D1b* allele. We obtained the amplification of *Pina-D1a* allele on that cultivar confirmed by sequence analysis. And, when we analysed the specific *Pina-D1b* amplification, we obtained the expected amplicon for the null allele. We could suppose that the presence of both alleles in Guadalupe cultivar is determined by heterozygosity for the alternative forms of *Pina-D1* alleles, which can influence the grain texture [27].

Regarding soft wheat cultivars we identified nucleotide mismatches compared to the wild type allele and, specifically, Colfiorito, Dorico, Adelaide and Pandas showed a sequence corresponding to *Pinb-D1b* allele leading to W73R an amino acidic change and Aubusson and Bologna cultivar showed the presence of *Pinb-D1d* allele causative G75S amino acidic changes. Although gene sequence analysis represents the most reliable result for genotyping, the development of simple and fast assays to identify the specific *Pina* or *Pinb* allele represent a necessity to extend genetic evaluation to a larger sample and to acquire more information about genetic determinants of grain texture. We evaluated the possibility of *Pinb-D1b* allele identification through a specific digestion assay (*SapI* restriction analysis), therefore we performed restriction analysis on Levante, Adelaide and Pandas *Pinb-D1b* amplicons as they include restriction sites resulting in two fragments visible through gel electrophoresis.

As reported, functional properties of puroindolines reside on their structure. The most critical region for lipid binding is the tryptophan rich loop region (PINA-D1a motif WRWWKWWK; PINB-D1a motif WPTKWWK). To understand the functional effect of *Pina* and *Pinb* variants on wheat grain texture, we performed a bioinformatics analysis [28]. Although most of the predictions revealed the deleterious effect of puroindoline sequence alteration in determining the hard texture phenotype, the bioinformatics prediction, based on the comparison of the query sequence with related sequences published in international databases, was not able to identify variations affecting functional aminoacidic regions, such as the tryptophan rich-domain (TRD) of puroindolines., *Pinb-D1b, d,* are predicted as neutral, although they are located in the region very next to TRD, presumably influencing the functionality of the synthesized polypeptide [29-33].

CONCLUSION

In conclusion, our study has isolated and characterised in wheat Italian cultivars new allelic polymorphisms. However, the comparison between our bioinformatic results and data emerging from our studies underline the need of functional experiments to validate *in silico* predictions. Future studies are warranted to clarify the mechanisms underlying the role of *Pina/Pinb* variations on wheat grain texture.

ACKNOWLEDGMENTS

We thank CRA–SCS (Salerno, Italy) and CRA–SCS (Verona, Italy) for providing seeds of wheat species. Author Contributions: BDF conceived and designed the experiments; BDF, MB, FM, and EB performed the experiments. BDF and EB analyzed the data. BDF and EB contributed reagents/materials/analysis tools. BDF wrote the paper.

REFERENCES

1. Raba DN, et al. Comparative studies regarding the panification performances of some wheat species used in food processing. JAPT. 2001;16:257-261.
2. Pauly A, et al. Wheat (*Triticum aestivum* L. and *T. turgidum* L. ssp. durum) Kernel Hardness: I. Current view on the role of puroindolines and polar lipids. Comp Rev Food Sci Food Safety. 2013;12:413-426.
3. Gautier MF, et al. *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol. 1994;25:43-57.
4. Li W, et al. Recurrent deletions of puroindoline genes at the grain hardness locus in four independent lineages of polyploid wheat. Plant Physiol. 2008;146:200-212.
5. Gasparis S, et al. The RNA-mediated silencing of one of the Pin genes in allohexaploid wheat simultaneously decreases the expression of the other, and increases grain hardness. J Exp Bot. 2011;62:4025-4036,
6. Li GY, et al. Molecular characterization of allelic variations at *Pina* and *Pinb* loci in Shandong wheat landraces, historical and current cultivars. J Cereal Sci. 2008;47:510-517.
7. Chantret N, et al. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell. 2005;17:1033-1045.

e-ISSN:2320-0189
p-ISSN:2347-2308

8. Morris CF and Bhave M. Reconciliation of D-genome puroindoline allele designations with current DNA sequence data. J Cereal Sci. 2008;48:277-287.
9. Morris CF. Puroindolines: The molecular genetic basis of wheat grain hardness. Plant Mol Biol. 2002;48:633-647.
10. Jing W, et al. Conformation of a bactericidal domain of puroindoline a: Structure and mechanism of action of a 13-residue antimicrobial peptide. J Bacteriol. 2003;185:4938-4947.
11. Krishnamurthy K, et al. Wheat puroindolines enhance fungal disease resistance in transgenic rice. Mol Plant Microbe Interact. 2001;14:1255-1260.
12. Corona V, et al. 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. 2001;34:243-250.
13. Bhave M and Morris CF. Molecular genetics of puroindolines and related genes: Allelic diversity in wheat and other grasses. Plant Mol Biol. 2008;66:205-219.
14. Bhave M and Morris CF. Molecular genetics of puroindolines and related genes: Regulation of expression, membrane binding properties and applications. Plant Mol Biol. 2008;66:221-231.
15. Chen F, et al. Molecular and biochemical characterization of puroindoline a and b alleles in Chinese landraces and historical cultivars. Theor Appl Genet. 2006;112:400-409.
16. Giroux MJ and Morris CF. Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. Proc Natl Acad Sci USA. 1998;95:6262-6266.
17. Giroux MJ and Morris CF. A glycine to serine change in puroindoline-b is associated with wheat grain hardness and low levels of starch-surface friabilin. Theor Appl Genet. 1997;95:857-864.
18. McIntosh RA, et al. Catalogue of gene symbols for wheat: 2006 supplement. 2006.
19. Gazza L, et al. Genetic and biochemical analysis of common wheat cultivars lacking puroindoline a. Theor Appl Genet. 2005;110:470-478.
20. Chen F, et al. Discovery, distribution and diversity of Puroindoline-D1 genes in bread wheat from five countries (*Triticum aestivum* L.). BMC Plant Biol. 2013;13:125.
21. Amoroso MG, et al. Real time RT-PCR and flow cytometry to investigate wheat kernel hardness: Role of puroindoline genes and proteins. Biotechnol Lett. 2004;26:1731-1737.
22. Capparelli R, et al. Puroindoline A-gene expression is involved in association of puroindolines to starch. Theor Appl Genet. 2003;107:1463-1468.
23. Turnbull KM, et al. Variation in puroindoline polypeptides in Australian wheat cultivars in relation to grain hardness. Aus J Plant Physiol. 2000;55:89-95.
24. Lillemo M, et al. Analysis of puroindoline a and b sequences from *Triticum aestivum* cv. 'Penawawa' and related taxa. Euphytica. 2002;126:321-331.
25. Kuchel H, et al. Genetic dissection of grain yield in bread wheat I QTL analysis. Theor Appl Genet. 2007; 115:1029-1041.
26. Pickering PA and Bhave M. Comprehensive analysis of Australian hard wheat cultivars shows limited puroindoline allele diversity. Plant Sci. 2007;172:371-379.
27. Choi Y. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus Variants of Another protein. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB '12). ACM: New York, NY, USA. 2012;414-417.
28. Doyle JJ and Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13-15.
29. Chen F, et al. Molecular characterization of the Puroindolinea-D1b allele and development of an STS marker in wheat (*Triticum aestivum* L.). J Cereal Sci. 2010;52:80-82.
30. Vincze T, et al. A program to cleave DNA with restriction enzymes. Nucleic Acids Res. 2003;31:3688-3691.
31. Altschul SF, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-3402.
32. Larkin MA, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947-2948.
33. Gasteiger E, et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3784-3788.