Theor Appl Genet (2006) 112: 400–409
DOI 10.1007/s00122-005-0095-z

ORIGINAL PAPER

F. Chen · Z. H. He · X. C. Xia · L. Q. Xia
X. Y. Zhang · M. Lillemo · C. F. Morris

Molecular and biochemical characterization of puroindoline a and b alleles in Chinese landraces and historical cultivars

Received: 20 June 2005 / Accepted: 22 August 2005 / Published online: 13 December 2005
© Springer-Verlag 2005

Abstract Kernel hardness that is conditioned by puroindoline genes has a profound effect on milling, baking and end-use quality of bread wheat. In this study, 219 landraces and 166 historical cultivars from China and 12 introduced wheats were investigated for their kernel hardness and puroindoline alleles, using molecular and biochemical markers. The results indicated that frequencies of soft, mixed and hard genotypes were 42.7, 24.3, and 33.0%, respectively, in Chinese landraces and 45.2, 13.9, and 40.9% in historical cultivars. The frequencies of PINA null, *Pinb-D1b* and *Pinb-D1p* genotypes were 43.8, 12.3, and 39.7%, respectively, in hard wheat of landraces, while 48.5, 36.8, and 14.7%, respectively, in historical hard wheats. A new *Pinb-D1* allele, designated *Pinb-D1t*, was identified in two landraces, Guangtouxianmai and Hongmai from the Guizhou province, with the characterization of a glycine to arginine substitution at position 47 in the coding region of *Pinb* gene. Surprisingly, a new *Pina-D1* allele, designated *Pina-D1m*, was detected in the landrace Hongheshang, from the Jiangsu province, with the characterization of a proline to serine substitution at position 35 in the coding region of *Pina* gene; it was the first novel mutation found in bread wheat, resulting in a hard endosperm with PINA expression. Among the PINA null genotypes, an allele designed as *Pina-D1l*, was detected in five landraces with a cytosine deletion at position 265 in *Pina* locus; while another novel *Pina-D1* allele, designed as *Pina-D1n*, was identified in six landraces, with the characterization of an amino acid change from tryptophan-43 to a 'stop' codon in the coding region of *Pina* gene. The study of puroindoline polymorphism in Chinese wheat germplasm could provide useful information for the further understanding of the molecular basis of kernel hardness in bread wheat.

Keywords Bread wheat (*Triticum aestivum* L.) · Kernel hardness · Puroindolines · *Pina* and *Pinb* alleles

Communicated by M.Morell

F. Chen · Z. H. He (✉) · X. C. Xia (✉) · L. Q. Xia · X. Y. Zhang
Chinese Academy of Agricultural Sciences (CAAS),
Institute of Crop Science/National Wheat Improvement Center,
Zhongguancun South Street 12, Beijing 100081, China
E-mail: zhhe@public3.bta.net.cn
Tel.: +86-10-62170333
Fax: +86-10-68918547
E-mail: xiaxianchun@caas.net.cn

Z. H. He
International Maize and Wheat Improvement Center (CIMMYT),
China Office, c/o CAAS, Zhongguancun South Street 12,
Beijing 100081, China

M. Lillemo
International Maize and Wheat Improvement Center (CIMMYT),
Apdo Postal 6-641, 06600 Mexico DF, Mexico

Present address: M. Lillemo
Department of Plant and Environmental Sciences,
Norwegian University of Life Sciences,
P.O. Box 5003, N-1432 Ås, Norway

C. F. Morris
Washington State University, USDA-ARS Western
Wheat Quality Laboratory, P.O. Box 646394, Pullman,
WA 99164-6394, USA

Introduction

Kernel texture, mainly controlled by one major locus (*Ha)* on the short arm of chromosome 5D, has a large effect on the end-use quality of bread wheat (*Triticum aestivum* L.) and is used for marketing classification (Mattern et al. 1973; Morris 2002). Generally, hard wheat is suitable for bread and other yeast-leavened foods, whereas, soft wheat is used for cookies, cakes and pastries (Morris and Rose 1996). A profound understanding of the genetic basis for kernel hardness resulted from the discovery of friabilin, a M_r 15-kDa protein (Greenwell and Schofield 1986). Friabilin was shown to contain three main components, puroindoline a (PINA), puroindoine b (PINB) and GSP-1; puroindolines represent the molecular-genetic basis of kernel hardness (Morris et al. 1994; Morris 2002; Hogg et al. 2004).

Genes coding for PINA and PINB were located at the *Ha* locus and designated as *Pina-D1* and *Pinb-D1*, respectively (Giroux and Morris 1997). When both genes are in their functional wild type form, the grain texture is soft. However, hard grain texture is the result of mutations in either the *Pina-D1* or *Pinb-D1* locus. Puroindolines are unique among plant proteins because of their tryptophan-rich domains, which have an apparent high affinity for binding lipids (Dubreil et al. 1997). Since hard wheats are associated with the failure to express PINA or with a glycine–to-serine mutation in the tryptophan-rich domain of *Pinb* was found by Giroux and Morris (1997, 1998), several additional *Pinb-D1* alleles have been reported in bread wheat, which result in a change in the kernel hardness from soft to hard (Lillemo and Morris 2000; Morris et al. 2001; Tranquilli et al. 2002; Xia et al. 2005; Chen et al. 2005; Ram et al. 2005). However, characterization of the *Aegilops tauschii* and synthetic hexaploid wheat has identified eight different *Pina* alleles and six unique *Pinb* alleles that are all associated with a soft endosperm (Gedye et al. 2004; Massa et al. 2004).

The most drastic effect on grain hardness is caused by the PINA-null allele that, in several studies, is shown to confer a harder endosperm than, for example, the *Pinb-D1b* allele (Giroux et al. 2000; Martin et al. 2001; Cane et al. 2004; Clarke and Rahman 2005). However, the molecular genetic basis of this mutation has not been known, except for the absence of mRNA transcripts (Giroux and Morris 1998) and, in most cases, a failure to amplify the coding region of *Pina* by flanking PCR markers (Ram et al. 2002; Cane et al. 2004). Recently, Gazza et al. (2005) reported a new variant of the PINA-null allele in cultivars Fortuna and Glenman, with a cytosine deletion at position 267 in the coding region of the *Pina* gene.

China is the largest producer and consumer of wheat in the world. Wheat is planted in ten agro-ecological zones that are further divided into 26 sub-zones, with winter, facultative, and spring wheats sown both in autumn and spring (He et al. 2001). Chinese wheat germplasm differs from that of other countries in several aspects. China is the secondary origin for wheat with a broad diversity in the germplasm. Chinese wheat cultivars are early in maturing, in order to suit the double cropping system, and they are mainly used to make steamed bread and noodles. Although several breeding programs were established in the 1930 s and 1940 s, improved cultivars were basically not used by farmers until the 1950 s. Currently, there are 13,370 landraces at the national gene bank. This is a unique resource since they were collected from various parts of China in the early 1950 s. More than 2,000 cultivars have been released in China since 1949, and breeding efforts in the 1950 s and early 1960 s were largely dependent on the crossing of outstanding Chinese landraces with elite introductions from USA, Italy, and other countries. Both Chinese landraces and historical cultivars are unselected populations for grain hardness since an intensive selection for quality improvement did not start until the late 1990 s. Therefore, Chinese landraces and historical cultivars are excellent resources to study the allelic variation of PINA and PINB in bread wheat. *Pinb-D1b* is the most dominant hardness allele in current Chinese cultivars (Xia et al. 2005), but two new alleles have also been recently discovered in this germplasm: *Pinb-D1p* was detected in ten hard genotypes (Xia et al. 2005) and *Pinb-D1q* was present in Jingdong 11 (Chen et al. 2005).

The main objectives of the present study were to characterize the distribution of puroindoline alleles in Chinese landraces and historical cultivars, to identify new mutations at *Pina* and *Pinb* loci, and to uncover the evolution of puroindoline alleles in Chinese wheats since the 1930 s.

Materials and methods

Wheat germplasm

In total, 219 landraces and 166 historical cultivars of bread wheat, developed from the 1930 s to the 1980 s, and 12 introduced wheats were used for the investigation of SKCS hardness and puroindoline alleles. These accessions are from a candidate core collection of 5,129 lines, which were selected from 23,705 accessions collected or bred in China, based on clustering within wheat regions or sub-regions by passport data and additional character data of the accessions (Zhang et al. 2002; Dong et al. 2003). Therefore, they represent very well the landmark landraces and historical cultivars in China from the 1930 s to 1980 s, which were collected from the 16 provinces of Gansu, Hebei, Beijing, Shaanxi, Shandong, Henan, Anhui, Hunan, Guizhou, Jiangsu, Sichuan, Zhejiang, Guangdong, Guangxi, Fujian, and Yunnan in China. They are available in the Chinese Wheat Genetic Germplasm Center, the Institute of Crop Science, the Chinese Academy of Agriculture Science (CAAS). In addition to the Chinese accessions, 12 exotic cultivars, i.e., Abbondanza, Ardito, Funo, and St 1472/506 from Italy, Early Premium from USA, Predgornaia 2 and Skorospelka L1 from Russia, Lovrin 10 and Lovrin 13 from Romania, Orofen from Chile, Suwon 86 from Korea, and Quality from Australia, were also included, since they were the most frequently used introductions in Chinese breeding programs before the 1980 s (Zhuang 2003).

The 219 landraces, 166 historical cultivars, and 12 introduced wheats were planted at the Luoyang Agricultural Research Institute in the 2002–2003 season, according to local management practices. Both winter, facultative, and spring types can grow well in Luoyang and additional supports were used to ensure that no lodging was present in the trial. After harvest, all the wheat samples were cleaned. Falling number tests indicated that they were free of sprouting damage.

Hardness measurement

Kernel hardness was measured with 300-kernel samples of each genotype, using the Perten single kernel characterization system (SKCS) 4100, following the manufacturer's operation procedure (Perten Instruments North America Inc., Springfield, IL, USA). The mean, standard deviation, and distribution of SKCS hardness data were used to classify the tested genotypes into soft, mixed, and hard types. The SKCS produces a four-class frequency distribution of hardness data for each cultivar with class limits of <33, 34–46, 47–59, and >60.

DNA isolation and PCR characterization of puroindoline alleles

The genomic DNA of three kernels, from each surveyed hard genotype based on the SKCS classification, was extracted separately from pulverized kernels, following a method modified from Lagudah et al. (1991), in order to verify the purity of the sample and detect bona fide alleles. Two pairs of allele-specific primers were used for the detection of *Pinb-D1b* (glycine-46 to serine-46), i.e., 5′-ATG AAG GCC CTC TTC CTCA-3′ (upstream primer) and 5′-CTC ATG CTC ACA GCC GCT-3′ (downstream primer), for the detection of the substitution of G to A at position 223 in the nucleotide sequence of *Pinb* gene, and 5′-ATG AAG GCC CTC TTC CTCA-3′ (upstream primer) and 5′-CTC ATG CTC ACA GCC GCC-3′ (downstream primer) for the detection of no nucleotide change at position 223 of the *Pinb* sequence (Giroux and Morris 1997, 1998). The genotypes with PINA null were detected by the SDS-PAGE of Triton X-114-soluble proteins as described below. *Pinb-D1c*, with the Proline-60 mutation in hard wheats, was detected by the site-specific cleavage of the polymerase chain reaction (PCR)-amplified product of *Pinb*. Two units of the restriction enzyme *PvuII*, together with the supplied reaction buffer and bovine serum albumin (BSA), was directly added to the PCR-amplified *Pinb* in a final reaction volume of 30 μl and incubated at 37°C for 2 h, following the method by Lillemo and Morris (2000).

The full-length *Pinb* was amplified using the sense-strand terminal primer 5′-ATG AAG ACC TTA TTC CTC CTA-3′ and the anti-sense terminal primer 5′-TCA CCA GTA ATA GCC ACT AGG GAA-3′ (Gautier et al. 1994). The sense-strand terminal primer 5′-ATG AAG GCC CTC TTC CTCA-3′ and the antisense-strand terminal primer 5′-TCA CCA GTA ATA GCC AAT AGTG-3′ were used to amplify the *Pina* gene (Gautier et al. 1994).

The PCR amplification was conducted in a PTC-200 Peltier Thermocycler (Gene Company, China). The PCR reaction was performed in 25 μl volumes containing 10 pmol of each primer, 250 μM of each of dNTP, 1×PCR buffer, 1.5 mM of $MgCl_2$, 0.5 unit of *Taq* DNA polymerase (Promega, USA), and 100 ng of template DNA. The samples, denaturated at 94°C for 2 min, were submitted to 35 cycles of 45 s denaturation at 94°C, 1 min annealing at 58°C, and 1 min elongation at 72°C, with a final extension of 5 min at 72°C in the end. The PCR products were analyzed on 1.5% (w/v) agarose gels, stained with ethidium bromide, and visualized using UV light.

Isolation of Triton-soluble proteins and SDS-PAGE detection of PINA null genotypes

Wheat kernels from the surveyed hard wheats were pulverized with a hammer between sheets of weighing paper. Triton-soluble proteins were extracted from the crushed kernels by phase partitioning, using the Triton X-114 detergent, following the procedure described by Morris and Massa (2003) with slight modifications. Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed in the standard method, using 13.5% T and 2.6% C resolving gel and 1 mm thick minigels (Morris et al. 1994). The stacking gel was 4% T and 2.67% C. Piperazine diacrylamide (PDA) was used as a cross linker, instead of bis-acrylamide, in order to produce a much cleaner background after staining. Two 8-well gels were run at 40 mA in a stacking gel and changed at 12 W when the dye arrived in the resolving gel. The gels were silver-stained with trichloroacetic acid (TCA) and methanol fixation, as described by Morris and Massa (2003), and the staining process was stopped with citric acid.

Standard check cultivars Chinese Spring and Falcon, representing the wild type (soft wheat) and PINA null (*Pina-D1b* hard wheat), respectively, were used in each gel as controls.

DNA sequencing

For the hard genotypes, except for *Pinb-D1b*, PINA null and *Pinb-D1c*, a 447 bp PCR fragment amplified from the template DNA of each surveyed kernel, with full-length *Pinb* specific primers, was sequenced from two strands by the Augct Biotechnology Company (http://www.augct.com). *Pinb* fragments from ten kernels and *Pina* fragments from five kernels, of the hard landrace cultivars Hongmai and Guangtouxianmai, developed in Xiuwen County, Guizhou, were sequenced from two strands by the Augct Biotechnology Company (http://www.augct.com) and the Bioasia Biotechnology Company (http://www.bioaisa.cn), respectively; while *Pina* fragments from ten kernels and *Pinb* fragments from five kernels, in the 12 hard landrace cultivars Hongheshang, Sanyuehuang, Baikezaomai, Xiaoyuhua, Chengdugaungtou, Guangtouxiaomai, Xianmai, Zhuantoubaike, and Baimangchun (two genotypes with a common name from Jianhu County and Guanyun County, respectively), Yazuizi and Yazuixiaomai, were sequenced from both strands by the above-mentioned two companies.

Sequence alignments were performed using DNAMAN software.

Statistical analysis

SAS 8.0 software and LSD multiple comparison were used to compute averages of the SKCS hardness index and to examine their difference among various puroindoline alleles.

Results

Distributions of SKCS grain hardness and puroindoline alleles in Chinese landraces and historical cultivars

In all, 173 landraces, historical wheats, and introduced cultivars were classified as soft, 77 as mixed, and 147 as hard type, according to the SKCS hardness index. As presented in Table 1, the frequencies of soft, mixed and hard genotypes were 42.7, 24.3, and 33.0%, respectively, in Chinese landraces, and 45.2, 13.9, and 40.9%, respectively, in historical cultivars. In comparison with landraces and historical cultivars, the frequency of hard genotypes in current cultivars increases notably (62.5%), with a corresponding decrease in the frequencies of soft and mixed hard types (25.1 and 12.4%).

For the soft types, both *Pina* and *Pinb* genes should be in the wild-type allelic form, according to the current model of the relationship between the puroindoline genes and kernel texture. The sequencing of both *Pina* and *Pinb*, in eight randomly selected soft wheat genotypes, confirmed the hardness model (data not shown). The mixed wheats were not genotyped by SKCS because some of them probably contained two or more genotypes (Chen et al. 2005). Further analysis will be performed for mixed wheats in the future.

The single kernel characterization system hardness index of the surveyed 147 hard wheats ranged from 48 to 100. Of these cultivars, 55, 34, and 39 genotypes belonged to *Pinb-D1b*, PINA null and *Pinb-D1p*, respectively, based on the PCR amplification with *Pinb-D1b* specific primer sets, SDS-PAGE of Triton X-114 soluble proteins and sequencing of the *Pinb* fragment. As presented in Table 2, the frequencies of PINA null, *Pinb-D1b*, and *Pinb-D1p* were 43.8, 12.3, and 39.7%, respectively, in hard landraces, and 48.5, 36.8, and 14.7%, respectively, in hard historical cultivars. All of the introduced hard cultivars (Early Premium, Predgornaia 2 and Skorospelka L1, Lovrin 10, and Lovrin 13, Orofen, and Suwon 86) confer the *Pinb-D1b* genotype. No *Pinb-D1c* allele was found with site-specific cleavage of the PCR-amplified fragment of the *Pinb* gene in this survey. Compared with landraces and historical cultivars, the frequencies of PINA null and *Pinb-D1p* was significantly reduced, while the frequency of *Pinb-D1b* was largely increased in the current cultivars.

Table 1 Frequencies of three hardness types in Chinese landraces, historical and current cultivars, and introduced wheats

Type	Sample No.	Soft (%)	Mixed (%)	Hard (%)
Landraces	219	42.7	24.3	33.0
Historical cultivars	166	45.2	13.9	40.9
Current cultivars[a]	251	25.1	12.4	62.5
Introduced cultivars	12	33.4	8.3	58.3
Total	648	35.2	16.7	48.1

[a]Data of current cultivars from Xia et al. (2005)

Pinb-D1p, characterized as a new *Pinb* frame-shift mutation, resulting in PINB null (Xia et al. 2005), was identified in 39 landraces and historical cultivars with the SKCS hardness index ranging from 50 to 79. Further analysis indicated that the *Pina* gene, in those genotypes with *Pinb-D1p*, was the wild type (*Pina-D1a*), based on the SDS-PAGE of Triton X-114 soluble proteins and sequencing of the *Pina* fragment. Interestingly, of the 29 landraces with *Pinb-D1p*, 18 were from the Gansu province (Table 3). The other 11 accessions with *Pinb-D1p* were distributed in seven different provinces, i.e., Beijing, Hebei, Henan, Shaanxi, Jiangsu, Guizhou, and Guangxi. Most of the *Pinb-D1p* genotypes in historical cultivars could be traced back to landraces; for example, the *Pinb-D1p* allele in five accessions (Shijiazhuang 407, Nongda 183, Nongda 311, Nongda 36, and Huabei 187) of Beijing and Hebei were derived from Yanda 1817, based on pedigree information (He et al. 2001).

Detailed data on the influence of three puroindoline mutations (PINA null, *Pinb-D1b*, and *Pinb-D1p*) on kernel hardness was shown in Fig. 1. Among these puroindoline mutations, the SKCS hardness index of PINA null (70.3) was significantly higher than that of both *Pinb-D1b* (61.6) and *Pinb-D1p* (60.5).

Novel puroindoline a and b alleles involved in a single nucleotide substitution

Two hard landraces, Guangtouxianmai and Hongmai, originating from Xiuwen County, Guizhou province, did not belong to any of the genotypes *Pinb-D1b*, *Pina-D1b*, and *Pinb-D1c*, based on the PCR amplification with *Pinb-D1b* specific primer sets, the SDS-PAGE of Triton X-114 soluble proteins and site specific cleavage of the PCR-amplified *Pinb* fragment. Subsequent DNA sequencing of the *Pinb* gene, from ten kernels of each of the two landraces, revealed a point mutation involving a base G to C substitution at the 226th nucleotide in the coding sequence of the *Pinb* gene, which results in a glycine (GGC) to arginine (CGC) substitution at position 47 in the deduced amino acid sequence of PINB (Table 4). Sequencing *Pina*-amplified fragments indicated that both landraces possessed the wild-type *Pina* allele. The *Pinb* allele in these two landraces was different from any one of previously reported *Pinb* mutations. The single nucleotide mutation, with G to C substitution at the 226th nucleotide of the *Pinb* locus, could be designated as

Table 2 Frequencies of different puroindoline alleles in Chinese landraces, historical, and current wheat cultivars

Genotype	Phenotype	Number	Landrace (%)	Historical cultivar (%)	Current cultivar[a] (%)
PINA null[b]	Hard	81	43.8	48.5	13.4
Pina-D1m/Pinb-D1a	Hard	1	1.4	0	0
Pina-D1a/Pinb-D1b	Hard	159	12.3	36.8	76.5
Pina-D1a/Pinb-D1d	Hard	2	0	0	1.7
Pina-D1a/Pinb-D1p	Hard	52	39.7	14.7	8.4
Pina-D1a/Pinb-D1t	Hard	2	2.8	0	0

[a]Data of current cultivars from Xia et al. (2005)
[b]PINA null is composed of *Pina-D1b/Pinb-D1a*, *Pina-D1l/Pinb-D1a*, and *Pina-D1n/Pinb-D1a*

Table 3 Presence of novel puroindoline alleles in Chinese landraces and historical cultivars

Genotype	Origin	Sample No.	Cultivar
Pina-D1l/Pinb-D1a	Jiangsu	3	Sanyuehuang, Baikezaomai, Xiaoyuhua
	Sichuan	1	Chengduguangtou
	Guangxi	1	Guangtouxiaomai
Pina-D1m/Pinb-D1a	Jiangsu	1	Hongheshang
Pina-D1n/Pinb-D1a	Jiangsu	6	Xianmai, Zhuantoubaike, Baimangchun (2)[a], Yazuizi, and Yazuixiaomai
Pina-D1a/Pinb-D1p	Gansu	18	Hongchunmai (2), Jinbaojin, Hongguangtou, Tiexiaomai, Jinghuangmai, Duanmai, and Xiaohongmai (3), Xiaomai, Mimai, Hongguangtou, Hongxiaomai, Hongmangmai, Changmangmang, Heshangtou, and Juangmangheshangtou
	Shaanxi	6	Mazhamai, Bima 1, Siqiangmai, Jingyang 302 (2),and Qinmai 4
	Jiangsu	5	Baishuituanlimai, Daliuleng, Liying 5, Xuzhou 15, and Suzhou 8060
	Beijing	3	Nongda 183, Youmanghong 7,and Fengkang 7
	Hebei	2	Shijiazhuang 407 and Tang 85-5032
	Guangxi	2	Guangtoumai and Hongkemai
	Henan	1	Youzimai
	Anhui	1	Xiaonong 8218
	Guizhou	1	Hongmai
Pina-D1a/Pinb-D1t	Guizhou	2	Guangtouxianmai and Hongmai

[a]Number of accessions with a common name originated from different places

Pinb-D1t, according to the 2005 Supplement of the Wheat Gene Catalogue (McIntosh et al. 2005). The SKCS hardness index (means ± SD) and frequency distribution of Guangtouxianmai and Hongmai were 68 ± 16, 64 ± 15, and 3-9-13-75, 3-6-23-68, respectively. Both of them were classified as class 1, a hard type.

A hard wheat cultivar Hongheshang, originating from Guanyun County, Jiangsu province, did not belong to any known mutation, either. Sequencing the *Pinb* gene in Hongheshang indicated that it had the wild type allele (*Pinb-D1a*). However, the sequencing of *Pina* revealed a new *Pina-D1* mutation, involving a base C–T substitution at the 187th nucleotide in the coding sequence of the *Pina* gene, resulting in a proline (CCG) to serine (TCG) substitution at position 35 in the deduced amino acid sequence of PINA (Table 5). The single nucleotide mutation, with C–T substitution at the 187th nucleotide of the *Pina* gene, could be designated as *Pina-D1m*, according to the 2005 Supplement of the Wheat Gene Catalogue (McIntosh et al. 2005). The SKCS hardness index (means ± SD) and frequency distribution of Hongheshang were 73 ± 13 and 1-3-9-87. It was classified as class 1, a hard type.

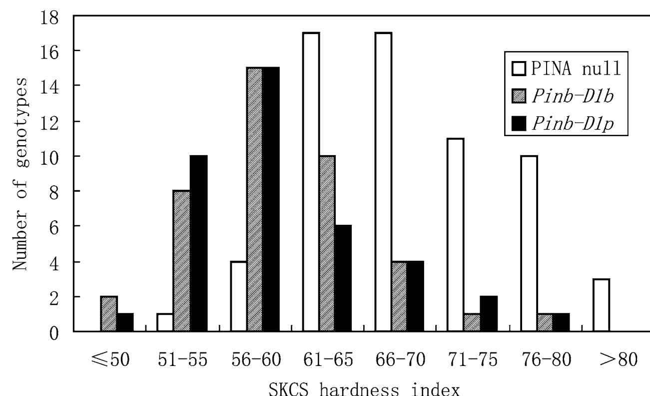

Fig. 1 SKCS hardness distribution for PINA null, *Pinb-D1b* and *Pinb-D1p* in hard Chinese landraces and historical cultivars Note: PINA null is composed of *Pina-D1b/Pinb-D1a*, *Pina-D1l/Pinb-D1a*, and *Pina-D1n/Pinb-D1a*

Molecular basis of the PINA-null genotype

In all, 65 PINA null genotypes, among the 147 hard landraces and historical cultivars, were identified by the SDS-PAGE of Triton X-114-extracted proteins. The electrophoresis of six randomly selected hard cultivars is shown in Fig. 2, in which the PINA protein was absent in the cultivars Yazuixiaomai, Baimangchun, Falcon (a control for PINA null), Yishanxiaomai and Sanyuehang

Table 4 Nucleotide and deduced amino acid sequence changes in the currently reported *Pinb* alleles in bread wheat

Allele	Position																											References
	13	14	15	16	…	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	
Pinb-D1a	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CTG	AGC	CAG	Giroux and Morris (1997)
(Soft)	Q	E	R	P	…	W	P	T	K	W	W	K	G	G	C	E	H	E	V	R	E	K	C	C	K	Q	L	S	Q	
Pinb-D1b	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	TGG	AAG	AGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CTG	AGC	CAG	Giroux and Morris (1998)
(Hard)	Q	E	R	P	…	W	P	T	K	W	W	K	S	G	C	E	H	E	V	R	E	K	C	C	K	Q	L	S	Q	
Pinb-D1c	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CCG	AGC	CAG	Lillemo and Morris (2000)
(Hard)	Q	E	R	P	…	W	P	T	K	W	W	K	G	G	C	E	H	E	V	R	E	K	C	C	K	Q	P	S	Q	
Pinb-D1d	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	AGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CTG	AGC	CAG	Lillemo and Morris (2000)
(Hard)	Q	E	R	P	…	W	P	T	K	W	R	K	G	G	C	E	H	E	V	R	E	K	C	C	K	Q	L	S	Q	
Pinb-D1e	CAG	GAG	CGG	CCG	…	TGA	CCC	ACA	AAA	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CTG	AGC	CAG	Morris et al. (2000)
(Hard)	Q	E	R	P	…	*	P	T	K	W	W	K	G	G	C	E	H	E	V	R	E	K	C	C	K	Q	L	S	Q	
Pinb-D1f	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	TGA	AAG	GGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CTG	AGC	CAG	Morris et al. (2000)
(Hard)	Q	E	R	P	…	W	P	T	K	W	*	K	G	G	C	E	H	E	V	R	E	K	C	C	K	Q	L	S	Q	
Pinb-D1g	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	TGG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGA	TGC	AAG	GAG	CTG	AGC	CAG	Morris et al. (2000)
(Hard)	Q	E	R	P	…	W	P	T	K	W	W	K	G	G	C	E	H	E	V	R	E	K	*	C	K	Q	L	S	Q	
Pinb-D1p	CAG	GAG	CGG	CCG	…	TGG	CCC	AC-A	AAT	GGT	GGA	AGG	GCG	GCT	GTG	AGC	ATG	AGG	TTC	GGG	AGA	AGT	GCT	GCA	AGG	AGC	TGA	GCC	AG	Xia et al. (2005)
(Hard)	Q	E	R	P	…	W	P	T	N	G	G	R	A	A	V	S	M	R	F	G	R	S	A	A	S	S	*	A		
Pinb-D1q	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	TTG	AAG	GGC	GGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CTG	AGC	CAG	Chen et al. (2005)
(Hard)	Q	E	R	P	…	W	P	T	K	W	L	K	G	G	C	E	H	E	V	R	E	K	C	C	K	Q	L	S	Q	

Table 4 (Contd.)

Allele	Position 13	14	15	16	…	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	References
Pinb-D1r	CAG	GAG	GCG	GCC	…	CTG	GCC	CAC	AAA	ATG	GTG	GAA	GGG	CGG	CTG	TGA	GCA	TGA	GGT	TCG	GGA	GAA	GTG	CTG	CAA	GGA	GCT	GAG	CCA	Ram et al. (2005), McIntosh et al. (2005)
(Hard)	Q	Q	A	A	…	L	A	H	K	M	V	E	G	R	L	*	A	*	G	S	G	E	V	L	E	G	A	E	P	
Pinb-D1s	CAG	GGA	GCG	GCC	…	CTG	ACC	CAC	AAA	ATG	GTG	GAA	GGG	CGG	CTG	TGA	GCA	TGA	GGT	TCG	GGA	GAA	GTG	CTG	CAA	GGA	GCT	GAG	CCA	Ram et al. (2005), McIntosh et al. (2005)
(Hard)	Q	Q	A	A	…	L	T	H	K	M	V	E	G	R	L	*	A	*	G	S	G	E	V	L	E	G	A	E	P	
Pinb-D1t	CAG	GAG	CGG	CCG	…	TGG	CCC	ACA	AAA	TGG	TGG	AAG	GGC	CGC	TGT	GAG	CAT	GAG	GTT	CGG	GAG	AAG	TGC	TGC	AAG	GAG	CTG	AGC	CAG	In this paper
(Hard)	Q	E	R	P	…	W	P	T	K	W	W	K	G	R	C	E	H	E	V	R	E	K	C	C	K	Q	L	S	Q	

Table 5 Nucleotide and deduced amino acid sequence changes of new PINA alleles detected in Chinese wheat landraces

Allele	Position 34	35	36	37	38	39	40	41	42	43	…	51	52	53	54	55	56	57	58	59	60	61	62
Pina-D1a	TTC	CCG	GTC	ACC	TGG	CGT	TGG	TGG	AAA	TGG	…	CTC	CTT	GGG	GAG	TGT	TGC	AGT	CGG	CTC	GGC	CAA	ATG
(Soft)	F	P	V	T	W	R	W	W	K	W	…	L	L	G	E	C	C	S	R	L	G	Q	M
Pina-D1l	TTC	CCG	GTC	ACC	TGG	CGT	TGG	TGG	AAA	TGG	…	CTC	CTT	GGG	GAG	TGT	TGC	AGT	CGG	CTC	GGC	–AAA	TG
(Hard)	F	P	V	T	W	R	W	W	K	W	…	L	L	G	E	C	C	S	R	L	G	K	
Pina-D1m	TTC	TCG	GTC	ACC	TGG	CGT	TGG	TGG	AAA	TGG	…	CTC	CTT	GGG	GAG	TGT	TGC	AGT	CGG	CTC	GGC	CAA	ATG
(Hard)	F	S	V	T	W	R	W	W	K	W	…	L	L	G	E	C	C	S	R	L	G	Q	M
Pina-D1n	TTC	CCG	GTC	ACC	TGG	CGT	TGG	TGG	AAA	TAG	…	CTC	CTT	GGG	GAG	TGT	TGC	AGT	CGG	CTC	GGC	CAA	ATG
(Hard)	F	P	V	T	W	R	W	W	K	*	…	L	L	G	E	C	C	S	R	L	G	Q	M

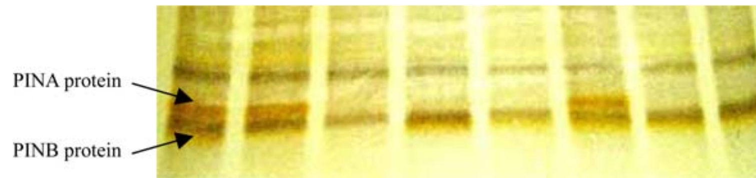

Fig. 2 SDS-PAGE of Triton X-114 extracted protein from a single kernel in six Chinese landrace wheat cultivars. From left to right: Hongheshang, Guangtouxianmai, Yazuixiaomai, Baimangchun, Falcon, Chinese Spring, Yishanxiaomai, and Sanyuehang; Chinese Spring and Falcon were the standard check cultivars for presence and absence of PINA, respectively. *Arrows* indicate the expected bands of PINA and PINB

but present in Hongheshang, Guangtouxianmai and Chinese Spring (a control for the wild type). Of these 65 PINA null genotypes, the PCR fragments of the *Pina* gene could be amplified in 11 landraces, but not in the other 54 accessions with *Pina* specific primers for the coding region. The sequencing *Pinb* of the 11 landraces indicated that they had the wild-type allele (*Pinb-D1a*). Subsequently, sequencing their *Pina* gene revealed a cytosine deletion at position 265 in five PINA-null landraces, including Sanyuehuang, Baikezaomai, Xiaoyuhua, Chengduguangtou and Guangtouxiaomai (Table 3). This mutation leads to a shift in the open reading frame (ORF) in the coding region of the *Pina* gene and, subsequently, a stop codon at position 93 in the deduced amino acid sequence of PINA. This mutation is identical to the allele previously detected and described as *Pina-D1c* by Gazza et al. (2005), which we have renamed as *Pina-D1l*, according to the 2005 Supplement of the Wheat Gene Catalogue (McIntosh et al. 2005) (Table 5). Moreover, a single nucleotide substitution was detected in the coding region of the *Pina* gene in six other PINA-null landraces, i.e., Xianmai, Zhuantoubaike, and Baimangchun (two cultivars with a common name from Jianhu County and Guanyun County, respectively), Yazuizi and Yazuixiaomai, characterized by a G–A change at position 212, resulting in tryptophan-43 to a ‘stop’ codon (Table 5). This new *Pina-D1* allele was designated as *Pina-D1n*, according to the 2005 Supplement of the Wheat Gene Catalogue (McIntosh et al. 2005).

Discussion

Evolution of puroindoline alleles in Chinese landraces, historical and current cultivars

Improvement of disease resistance and yield potential has been the major breeding objectives in Chinese breeding programs and historical cultivars were mostly developed by crossing Chinese landraces with introduced cultivars. Current wheat cultivars were predominantly developed by intercrossing historical cultivars or improved elite advanced lines (He et al. 2001). Among the parents of popular wheat cultivars, 26 Chinese landraces and 12 introduced cultivars were used most frequently in breeding programs (Zhuang 2003), in which 54.1% of landrace cultivars and 58.3% of introduced cultivars belong to the hard types. The 12 introduced parents can be divided into two groups. Four Italian cultivars, including Abbondanza, Ardito, Funo, and St 1472/506, are of the soft type and are widely used in breeding programs in the southern part of the Yellow and Huai Valleys (Zone II) and the Yangtze regions (Zones III and IV). The other eight introduced parents are hard types, with the predominant *Pinb-D1b* allele, and were widely used in northern China (Zones I and II). From Chinese landraces and historical cultivars to currently popular cultivars, the reduction of soft and mixed wheats and the increase of hard wheats are largely due to the increasing use of hard type parents and a favorable selection of hard genotypes in breeding programs, although soft types are also acceptable if they have outstanding agronomic performances.

In comparison with puroindoline alleles in Chinese landraces and historical cultivars, *Pinb-D1b* was sharply increased in the currently popular cultivars, while the frequencies of PINA null and *Pinb-D1p* alleles were gradually reduced during the past decades. Cultivars with the *Pinb-D1b* allele were also selected in various countries in Northern Europe, North America, Australia, and Japan (Lillemo and Morris 2000; Morris et al. 2001; Cane et al. 2004; Ikeda et al. 2005). This genotype is slightly superior to others in its milling and breadmaking characteristics (Martin et al. 2001; Nagamine et al. 2003; Cane et al. 2004) that, to some extent, might be a reason for the increase of the *Pinb-D1b* allele in Chinese improved cultivars.

Until now, the *Pinb-D1p* allele has only been found in Chinese wheat cultivars (Xia et al. 2005, Ikeda et al. 2005) and is shown in this study to be very prevalent in landraces from Gansu Province. It is possible that China, and even the Gansu province, was the origin of the *Pinb-D1p* allele, since not much exchange of germplasm among provinces was recorded in early China. Further study is needed to confirm the origin of the *Pinb-D1p* allele. *Pinb-D1p* involves a shift in the ORF and a truncated *Pinb* transcript. The SKCS hardness of the PINA null is much higher than that of the *Pinb-D1p* (i.e., PINB null, Fig. 1), which provides additional evidence that the PINA protein plays a more important role in kernel hardness than the PINB, which is consistent with previous studies (Capparelli et al. 2003; Clarke and Rahman 2005). The SKCS hardness of the PINA

null is also much higher than that of the *Pinb-D1b*, but the presence of the PINA null is very high in Chinese landraces. More work is needed to understand this, since most wheat was milled by manual methods in the major wheat growing areas before the 1970 s.

Molecular basis of *Pinb-D1* mutations

Based on the model outlined in Lillemo and Morris (2000), the tryptophan-rich domain in PINB, consisting of residues 39–45, is positioned in the loop between helix one and two and probably stabilized by a disulphide bond. Giroux and Morris (1997) predicted the tryptophan-rich domain to be in a beta-sheet conformation. The glycine to arginine change at position 47 (*Pinb-D1t*) occurs in a loop between the tryptophan-rich domain and the second helix, and the reduced flexibility introduced by the arginine residue may alter the lipid binding abilities of the tryptophan-rich domain and, hence, result in the hard endosperm of wheat kernels with *Pinb-D1t*, just as the effect of the glycine to serine change on the protein structure that is described by Giroux and Morris (1997). Among the known mutations in *Pinb*, position 44 has a higher rate of mutation than other positions and three mutation types, *Pinb-D1d* (Trp-44 to Arg-44), *Pinb-D1f* (Trp-44 to stop codon) and *Pinb-D1q* (Trp-44 to Leu-44), which have been detected at this position (Lillemo and Morris 2000; Morris et al. 2001; Chen et al. 2005). In contrast, all other mutations in *Pinb* occur at different positions and the mutation occurring at position 47 of *Pinb* locus was just reported in this survey (Table 4).

Molecular basis of *Pina-D1* mutations

Previously, the PINA null (*Pina-D1b*) was the only hardness mutation known at the *Pina-D1* locus and was simply described as a lack of PINA protein expression (Giroux and Morris 1997, 1998; Lillemo and Morris 2000; Morris et al. 2001). New insight came with the discovery of the *Pina-D1l* allele, which is a variant of the PINA null with a cytosine deletion at position 265, as shown in the present study and also reported by Gazza et al. (2005). We checked the sequence of *Pina* in common wheat and found that the 267th position is a base A instead of C and that a cytosine deletion at the 267th position at the *Pina* locus, reported by Gazza et al. (2005), should be at the 265th position, which is consistent with the sequencing results of *Pina* in the present study. The name *Pina-D1c*, designated by Gazza et al. (2005), was in conflict with a previous report for an allele in *A. tauschii* by Gedye at al. (2004). Therefore, the PINA-null type, due to a cytosine deletion at position 265 in the coding region of the *Pina* gene, was renamed as *Pina-D1l* (McIntosh et al. 2005).

A second variant of PINA null, *Pina-D1n*, was detected with a single nucleotide substitution (G–A). The occurrence of the *Pina-D1n* allele was identified in six genotypes from six different counties of the Jiangsu province, indicating that *Pina-D1n* was prevalent in the landrace cultivars confined to that region. Five genotypes with *Pina-D1l* were detected in landraces from the Jiangsu, Sichuan, and Guangxi provinces, suggesting that *Pina-D1l* has a wider distribution. Among the other PINA null genotypes, no PCR fragment could be amplified with *Pina* specific primers, indicating that some big changes might have occurred at the *Pina-D1* locus in this genotype, as discussed by several authors (Cane et al. 2004; Gazza et al. 2005).

Pina polymorphism, *Pina-D1c–Pina-D1j*, was previously only found in *A. tauschii* (Massa et al. 2004; Gedye et al. 2004); however, the *Pina-D1m* allele reported in this study is the first amino acid substitution in the PINA of bread wheat that is known to result in hard endosperm. Bihan et al. (1996) predicted by infrared and Raman spectroscopy that *Pina* had a secondary structure composed of approximately 30% α-helices (four helices, linked together by flexible loops), 30% β-sheets, and 40% unordered structure at pH 7. The deduced amino acid sequence of *Pina* suggested that the tryptophan-rich domain consisted of residues 38–44 based on the sequence reported by Gautier et al. (1994). The proline to serine change at 35 (*Pina-D1m)* would then occur in a loop between the first helix and the tryptophan-rich domain, and the reduced flexibility introduced by the serine residue might alter the lipid-binding abilities of the tryptophan-rich domain (Bihan et al. 1996). Proline is known as an alpha helix breaker due to its irregular side chain constraints and sterics (Piela et al. 1987; Yun et al. 1991).

Acknowledgements This project was funded by the National Basic Research Program (2002CB11300), National 863 program (2003AA207090) and National Natural Science Foundation of China (30260061).

References

Bihan TL, Blochet JE, Desormeaux A, Marion D, Pezolet M (1996) Determination of the secondary structure and conformation of puroindolines by infrared and Raman spectroscopy. Biochemistry 35:12712–12722

Cane K, Spackman M, Eagles HA (2004) Puroindoline genes and their effects on grains quality traits in southern Australian wheat cultivars. Aust J Agric Res 55:89–95

Capparelli R, Borriello G, Giroux MJ, Amoroso MG (2003) Puroindoline a-gene expression is involved in association of puroindolines to starch. Theor Appl Genet 107:1463–1468

Chen F, He ZH, Xia XC, Lillemo M, Morris CF (2005) A new puroindoline b mutation present in Chinese winter wheat cultivar Jingdong 11. J Cereal Sci 42:267–269

Clarke B, Rahman S (2005) A microarray analysis of wheat grain hardness. Theor Appl Genet 110:1259–1267

Dong YS, Cao YS, Zhang XY, Liu SC, Wang LF, You GX, Pang BS, Li LH, Jia JZ (2003) Development of candidate core collections in Chinese common wheat germplasm. J Plant Genet Resour 4:1–8 (in Chinese)

Dubreil L, Compoint JP, Marion D (1997) Interaction of puroindolines with wheat flour polar lipids determining their foaming properties. J Agric Food Chem 45:108–116

Gautier MF, Aleman ME, Guirao A, Marion D, Joudrier P (1994) *Triticum aestivum* puroindolines, two basic cysteine rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Bio 25:43–57

Gazza L, Nocente F, Ng PKW, Pogna NE (2005) Genetic and biochemical analysis of common wheat cultivars lacking puroindoline a. Theor Appl Genet 110:470–478

Gedye KR, Morris CF, Bettge AD (2004) Determination and evaluation of the sequence and testural effects of the puroindoline a and puroindoline b genes in a population of synthetic hexaploid wheat. Theor Appl Genet 109:1597–1063

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

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

Giroux MJ, Talbert L, Habernicht DK, Lanning A, Hemphill A, Martin JM (2000) Association of puroindoline sequence type and grain hardness in hard red spring wheat. Crop Sci 40:370–374

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

He ZH, Rajaram S, Xin ZY, Huang GZ (2001) A history of wheat breeding in China. CIMMYT, Mexico, DF, pp 1–94

Hogg AC, Sripo T, Beecher B, Martin JM, Gorpux MJ (2004) Wheat puroindolines interact to form friabilin and control wheat grain hardness. Theor Appl Genet 108:1089–1097

Ikeda TM, Ohnishi N, Nagamine T, Oda S, Hisatomi T, Yano H (2005) Identification of new puroindoline genotypes and their protein products among wheat cultivars. J Cereal Sci 41:1–6

Lagudah ES, Appels R, McNeil D (1991) The *Nor-D3* locus of *Triticum tauschii*: natural variation and genetic linkage to markers in chromosome 5. Genome 34:387–395

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

Martin JM, Frohberg RC, Morris CF, Talbert LE, Giroux MJ (2001) Milling and bread baking traits associated with puroindoline sequence type in hard red spring wheat. Crop Sci 41:228–234

Massa AN, Morris CF, Gill BS (2004) Sequence diversity of puroindoline-a, puroindoline-b, and the grain softness protein genes in *Aegilops tauschii* coss. Crop Sci 44:1808–1816

Mattern PJ, Morris R, Schmidt JW, Johnson VA (1973) Location of genes for kernel properties in wheat cultivar 'Cheyenne' using chromosome substitution lines. In: Sears ER, Sears LMS (eds) Proceedings of the 4th international wheat genetics symposium. University of Missouri, Columbia, Mo, pp 703–707

McIntosh RA, Devos KM, Dubcovsky J, Rogers WJ, Morris CF, Appels R, Anderson OD (2005) Catalogue of gene symbols for wheat: 2005 supplement, published online at: http://wheat.pw.usda.gov/ggpages/wgc/2005 upd.html

Morris CF (2002) Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Mol Bio 48:633–647

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

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

Morris CF, Massa AN (2003) Puroindoline genotype of the U.S. national institute of standards & technology reference material 8441, wheat hardness. Cereal Chem 80:674–678

Morris CF, Rose SP (1996) Wheat. In: Henry RJ, Kettlewell PS (eds) Cereal grain quality. Chapman and hall, NY, pp 3–54

Nagamine T, Ikeda TM, Yanagisawa T, Yanaka M, Ishikawa N (2003) The effects of the hardness allele *Pinb-D1b* on the flour quality of wheat for Japanese white salty noodles. J Cereal Sci 37:337–342

Piela L, Nemethy G, Shceraga HA (1987) Proline-induced constraints in α–helices. Biopolymers 26:1587–1600

Ram S, Boyko E, Giroux MJ, Gill BG (2002) Null mutation in puroindoline a is present in Indian wheats: puroindoline genes are located in the distal part of 5DS. J Plant Biochem Biotech 11:79–83

Ram S, Jain N, Shoran J, Singh R (2005) New frame shift mutation in puroindoline b in Indian wheat cultivars Hyb65 and NI5439. J Plant Biochem Biotech 14:45–48

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

Xia LQ, Chen F, He ZH, Chen XM, Morris CF (2005) Occurrence of puroindoline alleles in Chinese winter wheats. Cereal Chem 82:38–43

Yun RH, Anderson A, Hermans J (1991) Proline in α–helices: stability and conformation studied by dynamics simulation. Proteins 10:219–228

Zhang XY, Pang BS, You GX, Wang LF, Jia JZ, Dong YC (2002) Allelic variation and genetic diversity at *Glu-1* loci in Chinese wheat germplasms. Agric Sci China 1:1074–1082

Zhuang QS (2003) Chinese wheat improvement and pedigree analysis. Chinese Agriculture Press, Beijing, pp 1–681