Prevalence of puroindoline alleles in wheat varieties from eastern Asia including the discovery of a new SNP in puroindoline b

Hiroyuki Tanaka[1]*, Craig F. Morris[2], Mika Haruna[1] and Hisashi Tsujimoto[1]

[1]*Laboratory of Plant Genetics and Breeding Science, Faculty of Agriculture, Tottori University, Tottori 680-8553, Japan* and [2]*USDA-ARS Western Wheat Quality Laboratory, Pullman, WA 99164-6394, USA*

Abstract

Kernel texture (grain hardness) in common wheat (*Triticum aestivum* L.) is of primary technological importance and is largely determined by puroindoline gene sequence and expression. We investigated the puroindoline haplotypes of 246 Asian common wheat varieties. All but three were conclusively characterized for puroindoline a and b haplotypes. Of the total, 174 possessed the soft *Pina-D1a/Pinb-D1a* 'wild-type' gene sequences with SKCS hardness indexes (HI) ranging from 13.5 to 61.8. Among the remaining 72 varieties with HIs of 56.1–97.8, nearly half (30) were *Pina-D1a/Pinb-D1b*, 4 were *Pina-D1a/Pinb-D1c*, 19 were *Pina-D1a/Pinb-D1p*, 10 were *Pina-D1b/Pinb-D1a* ('a-null'), 3 were *Pina-D1l/Pinb-D1a*, 2 possessed a new C-to-T SNP mutation at position 382, which is tentatively designated *Pinb-D1ab*, 1 was a 'double null' with neither puroindoline a nor b expressed and no PCR-detectable gene sequence, and 3 had undetermined/ambiguous puroindoline a sequence but possessed *Pinb-D1a*. The double null was the hardest of all varieties tested with an HI of 97.8. The frequency of soft and hard varieties and puroindoline hardness haplotype varied depending on the origin of the varieties. The lowest frequency of hard varieties occurred in Korea and south-western Japan. Tibet and Pakistan also had low frequencies of hard varieties. The highest frequency of hard varieties appeared in north-east China followed by north-west China and Nepal. Within Asia, the *Pinb-D1p* allele appears in a region extending from north-eastern China through Inner Mongolia, north-western China, Xinjiang and Tibet, with the greatest frequency in north-western China. This allele was also present in Pakistan and Afghanistan, but not found in Japan, and may have been dispersed along the 'Silk Road'. All three *Pina-D1l* varieties came from China. The newly discovered SNP originated in Afghanistan and the 'double null' in Xinjiang.

Keywords: grain hardness; mutation; puroindoline; *Triticum aestivum* L; wheat

Introduction

Grain hardness (kernel texture) is one of the most important end-use quality characteristics of cultivated common wheat (*Triticum aestivum* L.), as it has a profound effect on milling, processing and utilization.

Based on the texture of the mature kernel, common wheat varieties are typically classified as 'hard' or 'soft'. Soft wheat varieties are used for cakes, cookies, pastries and some types of noodles, whereas hard wheat varieties are used for breads and other yeast-leavened foods (Morris and Rose, 1996).

Much of the variation in wheat grain hardness is controlled by a single locus, called *hardness* (*Ha*) (Symes, 1965; Baker, 1977), which is located on the short arm

---

*Corresponding author. E-mail: htanaka@muses.tottori-u.ac.jp

of chromosome 5D (Mattern *et al.*, 1973; Law *et al.*, 1978). Greenwell and Schofield (1986) reported the occurrence of 'friabilin', a small group of Mr 15k proteins that are abundant on the surface of water-washed starch granules from soft wheat, scarce on the surface of hard wheat starch granules, and absent in durum wheat. The genes encoding friabilin were tightly linked to the *Ha* locus (Jolly *et al.*, 1993; Sourdille *et al.*, 1996; Campbell *et al.*, 1999). Friabilin is composed mainly of two proteins, puroindolines a and b, and the respective genes are *Pina-D1* and *Pinb-D1* (Morris, 2002). These puroindoline proteins contain unique tryptophan-rich domains with apparent affinity for phospholipids on the surface of starch granules (Gautier *et al.*, 1994; Greenblatt *et al.*, 1995; Bhave and Morris, 2008a, b).

To investigate a direct causal relationship between puroindolines and grain hardness, rice, which has no homologues to the puroindolines (Gautier *et al.*, 2000), was transformed with wild-type puroindoline genes (Krishnamurthy and Giroux, 2001). Textural analysis of the transgenic rice seeds indicated that expression of wild-type puroindoline genes reduced rice grain hardness. These data supported the hypothesis that puroindolines play an important role in controlling kernel hardness.

As mentioned previously, durum and related tetraploid wheat varieties lack the genes for puroindolines and consequently have very hard kernels. Upon allopolyploidization, wherein *T. turgidum* ssp. *dicoccoides* (Körn. ex Asch & Graebn.) hybridized with *Aegilops tauschii* (Coss.) forming *T. aestivum*, the puroindolines contributed by *Ae. tauschii* restored kernel softness. All soft wheat varieties examined so far carry these ancestral 'wild-type' puroindoline gene alleles that have been designated *Pina-D1a* and *Pinb-D1a* (Giroux and Morris, 1997, 1998; Morris, 2002; Bhave and Morris, 2008a, b). Hard wheat varieties have specific mutations in either of the *Pina-D1* or *Pinb-D1* genes (Morris, 2002; Morris and Bhave, 2007).

Subsequent to its formation about 7000 years ago in the south-east coastlands of the Caspian Sea (Tsunewaki, 1966, Nishikawa *et al.*, 1980), common wheat spread to nearly every corner of the arable earth. During its transmission to Europe, Africa and southern and eastern Asia, mutations were selected, which conferred better adaptation to a wide range of climates. For instance, in China, there are areas of high mountains and severe arid deserts. Thus, the earliest eastward spread of wheat from its centre of origin was not a simple expansion of an ecological continuum. In the process of transmission to eastern Asia, the food products of wheat consumption changed from baked breads (generally flat breads) to steamed bread and noodles.

The genetic composition of wheat varieties reveals important information as to the movement and utilization of germplasm. We previously reported that limited but specific variation of seed storage proteins occur in Japanese common wheat varieties (Tanaka *et al.*, 2003). We also investigated the diversity of the low-molecular-weight glutenin subunit genes in Asian common wheat varieties (Tanaka *et al.*, 2005). The puroindolines are especially useful in this regard and show clear differences in allele/haplotype frequency based on region (Giroux and Morris, 1998; Lillemo and Morris, 2000; Corona *et al.*, 2001; Morris *et al.*, 2001; Pogna *et al.*, 2002; Ram *et al.*, 2002; Bagulho *et al.*, 2003; Cane *et al.*, 2004; Chen *et al.*, 2005; Gazza *et al.*, 2005; Huang and Röder, 2005; Ikeda *et al.*, 2005; Xia *et al.*, 2005; Chang *et al.*, 2006; Lillemo *et al.*, 2006; Ravel *et al.*, 2006; Pickering and Bhave, 2007). These surveys on wheat varieties have also been instrumental in discovering new mutations (alleles) in the puroindolines, thus expanding our understanding of these important genes as well as identifying valuable genetic resources for wheat improvement.

In the present study, we surveyed a large number of wheat varieties from eastern Asia for kernel texture and puroindoline haplotype. The frequency of previously known puroindoline haplotypes in Asian common wheat varieties is presented. Furthermore, we report the discovery of a new SNP in puroindoline b.

Materials and methods

Plant materials

We examined 246 varieties of common wheat from Asia (Tables 1 and 2). All of the varieties except those that are designated as 'Norin' in Japan are landraces. Japan was divided into two regions, north-eastern Japan (Hokkaido, Tohoku, Kanto and Hokuriku/Nagano) and south-western Japan (Tokai/Kinki, Chugoku and Shikoku/Kyushu; Nakamura *et al.*, 1999), based on climate. Many varieties within each region were bred using germplasm within the same region. Fukunaga and Inagaki (1985) reported genealogical pedigrees of Japanese wheat lines and showed that the cultivars within each region were similar to each other, but those between the regions were differentiated. China was grouped into seven regions: eastern China (Shandong, Jiangsu, Anhui, Zhejiang and Fujian Provinces), north-western China (Shaanxi, Gansu and Ningxia Provinces), north-eastern China (Heilongjiang, Jilin and Liaoning Provinces), Sichuan, Inner Mongolia, Xinjiang and Tibet (Tsujimoto *et al.*, 1998). A few varieties were identified without a specific province and are indicated as only 'China'. We also examined the AABB tetraploid durum (*T. turgidum* ssp. *durum* Desf. [Husn.]) variety, 'Stewart'.

Table 1. Soft wheat varieties (*Pina-D1a/Pinb-D1a*, SKCS hardness index < 50, $n = 169$) from eastern Asia classified according to geographic origin

Afghanistan	KU3066, KU3068
Bhutan	KU7070, KU7180, KU7182
China (unknown)	KT020-529, KT020-533, KT020-539, KT020-540, KT020-546, KT020-602
China (east)	KT020-528, KT020-538, KT020-544, KT020-545, KT020-547, KT020-551, KT020-553, KT020-554, KT020-557, KT020-558, KT020-562, KT020-564, KT020-565, KT020-568, KT020-569, KT020-571, KT020-572, KT020-574, KT020-575
China (inner Mongolia)	KT020-603, KT020-604, KT020-608, KT020-609
China (north-east)	KT020-517, KT020-520
China (north-west)	KT020-576, KT020-578, KT020-583, KT020-587, KT020-594, KT020-596
China (Sichuan)	KT020-541, KT020-552, KT020-559, KT020-563, KT020-566, KT020-567, KT020-570, KT020-598, KT020-599
China (Tibet)	KU13501, KU13502, KU13503, KU13504, KU13505, KU13604, KU13649, KU13650, KU13651, KU13652, KU13657, KU13808, KU13809, KU13810, KU13811
China (Xinjiang)	KT020-586
Japan (Chugoku)	Akatsukikomugi (Norin 83), Eshima, Fukuwasekomugi (Norin 128), Hirakikomugi, Hiroshimashipurei, Norin 52, Norin 74, Shirasagikomugi (Norin 95), Taikankomugi, Ushiokomugi (Norin 105), Yushoki
Japan (Hokkaido)	Akakawaaka, Honkei 124, Norin 29, Sapporoharukomugi
Japan (Hokuriku/Nagano)	Hokuriku 30, Igachikugooregon, Kokeshikomugi (Norin 89), Myokoukomugi (Norin 86), Shiranekomugi (Norin 131), Zenkoujikomugi (Norin 109)
Japan (Kanto)	Aka, Akabouzu, Akakomugi, Fujimikomugi (Norin 98), Fukuhokomugi (Norin 124), Kanagawawasekomugi, Koborehachikoku, Norin 68, Saitama 27, Sakyukomugi (Norin 91), Shirodaruma, Shirodarumasai 1, Sunekiri 15, Wasenyubai, Yuyakekomugi (Norin 76)
Japan (Shikoku/Kyushu)	Aburakomugi, Asakazekomugi (Norin 123), Asozairai, Chikushikomugi (Norin 121), Danchikomugi (Norin 93), Ebisukomugi (Norin 87), Eshimashinriki, Fukuraku, Gogatsukomugi (Norin 118), Hatamasari (Norin 80), Hayakomugi, Hayamehaikara, Hiyokukomugi (Norin 107), Igachikugo, Iyokomugi (Norin 79), Junreikomugi (Norin 96), Minaminokomugi (Norin 125), Nichirinkomugi (Norin 103), Norin 20, Norin 45, Norin 5, Norin 61, Saikai 102, Saikai 106, Sakigakekomugi (Norin 112), Setokomugi (Norin 120), Shiroganekomugi (Norin 117), Shirowasekomugi (Norin 122), Soujukuakage, Yutakakomugi
Japan (Tohoku)	Furutsumasari (Norin 94), Hachimankomugi, Hitsumikomugi (Norin 88), Kitakamikomugi (Norin 97), Nanbukomugi (Norin 82), Norin 10, Norin 39, Okukomugi (Norin 90), Sanjaku, Shimofusakomugi (Norin 101), Susonokomugi (Norin 77), Wakamatsukomugi (Norin 127)
Japan (Tokai/Kinki)	Gifuwasekomugi, Kobushikomugi (Norin 110), Norin 26, Norin 32, Norin 43, Norin 59, Omasekomugi (Norin 106), Sanshukotake, Shinchinko 1, Shinrikikomugi, Shirochabo, Shironankin
Korea	KT020-513, KT020-514, KT020-515, KT020-516
Pakistan	KU3333, KU3335, KU3342, KU3355, KU3356, KU3358, KU3359, KU3363

The varieties were maintained at Tottori University as a part of the National Bioresources Project − Wheat, Japan.

Kernel texture and puroindoline analyses

Kernel texture of all 246 varieties was determined using the Perten Single Kernel Characterization System (SKCS) 4100 (Perten Instruments Inc., Springfield, IL) on a 50-kernel sample each of clean, unbroken wheat kernels, according to the manufacturer's instructions (see also AACC International, 2000).

Puroindoline gene sequence analysis was conducted by one of two procedures, depending on whether the analysis was performed in Japan or the United States. Total genomic DNA was extracted from young leaves (Japan) or dry mature seeds (USA) by the CTAB method (Murray and Thompson, 1980) and used as template for PCR. The PCR primers used to amplify the puroindoline genes were those reported by Gautier *et al.* (1994) as follows: 5′-ATGAAGGCCCTCTTCCTCA-3′ and 5′-TCACCAGTAATAGCCAATAGTG-3′ for the amplification of the full-length (447 bp) puroindoline a gene, and 5′-ATGAAGACCTTATTCCTCCTA-3′ and 5′-TCACCA-GTAATAGCCACTAGGGAA-3′ for the amplification of the full-length (447 bp) puroindoline b gene.

PCR amplification (Japan) was performed using TaKaRa Ex *Taq* DNA polymerase (2.5 U, TaKaRa) in 100 μl of reaction buffer (TaKaRa, 2 mM MgCl$_2$) containing 200 ng of genomic DNA, 200 μM of each dNTP and 40 pmol of each primer. The PCR conditions were 93°C for 4 min followed by 35 cycles of 94°C for 60 s, 53°C

Table 2. Hard wheat varieties (SKCS hardness index >50, $n = 77$) from eastern Asia classified according to *Puroindoline* haplotype and geographic origin

Pina-D1a/Pinb-D1a	
Bhutan	KU7001, KU7005
Nepal	KU4704, KU4705, KU4739
Pina-D1a/Pinb-D1b	
China (east)	KT020-537, KT020-550, KT020-555, KT020-556
China (inner Mongolia)	KT020-605, KT020-610
China (north-east)	KT020-518, KT020-519, KT020-521, KT020-523, KT020-525
China (north-west)	KT020-582, KT020-592
China (Sichuan)	KT020-548
China (unknown)	KT020-527, KT020-542, KT020-591, KT020-601
Japan (north-east)	Aobakomugi (Norin 81), Hanagasakomugi (Norin 116), Haruminori (Norin 111), Hikarikomugi (Norin 85), Horoshirikomugi (Norin 114), Miyaginokomugi (Norin 102), Mukakomugi (Norin 108), Norin 42, Takunekomugi (Norin 115), Toyohokomugi (Norin 119)
Japan (south-west)	Kobouzu
Nepal	KU4703
Pina-D1a/Pinb-D1c	
Japan (north-east)	Haruhikari (Norin 104), Mikunikomugi (Norin 100), Norin 75
China (unknown)	KT020-580
Pina-D1a/Pinb-D1p	
Afghanistan	KU3070
China (east)	KT020-526, KT020-536, KT020-543, KT020-573
China (inner Mongolia)	KT020-606, KT020-607
China (north-east)	KT020-522
China (north-west)	KT020-577, KT020-588, KT020-589, KT020-597, KU481, KU482
China (Tibet)	KU13513A
China (Xinjiang)	KT020-585
China (unknown)	KT020-532, KT020-600
Pakistan	KU3357
Pina-D1a/Pinb-D1ab	
Afghanistan	KU3062, KU3069
Pina-D1b/Pinb-D1a	
Afghanistan	KU3067
Bhutan	KU7004, KU7029
China (east)	KT020-549
China (north-east)	KT020-524
China (north-west)	KT020-590
China (Sichuan)	KT020-561
Japan (north-east)	Haruyutaka (Norin 130)
Nepal	KU4706, KU4738
Pina-D1l/Pinb-D1a	
China (east)	KT020-534, KT020-535
China (unknown)	KT020-531
Double null	
China (Xinjiang)	KT020-584
Unknown/Pinb-D1a	
China (east)	KT020-560
Japan (south-west)	Sadabouzu
Nepal	KU4740

for 90 s and 72°C for 120 s. A final cycle with an extension of 10 min at 72°C completed the reaction.

PCR amplification (USA) was performed using *Taq* DNA polymerase (0.5 U, Promega, Madison, WI) in 25 µl of 1× *Taq* DNA polymerase reaction buffer (1.5 mM MgCl$_2$) containing 100 ng of genomic DNA, 250 µM of each dNTP and 10 pmol of each primer. The PCR conditions were 94°C for 4 min followed by 35 cycles of 94°C for 60 s, 58°C for 90 s and 72°C for 120 s. A final cycle

with an extension of 10 min at 72°C completed the reaction.

The products were separated in 1.5% (w/v) agarose gel, stained with ethidium bromide and visualized using UV light. Each amplified fragment was extracted from the agarose gel using QIAquick Gel Extraction Kit (QIAGEN) and inserted into the cloning site of the plasmid vector pGEM-T Easy (Promega, Japan), or purified from the dNTPs and oligonucleotide primers with

exonuclease I and shrimp alkaline phosphatase (ExoSAP-IT, UBS, Cleveland, OH).

DNA sequences were determined on a DNA sequencer (ABI PRISM 3100 Genetic Analyzer, PerkinElmer Applied Biosystems Division, Foster City, CA) using the BigDye Terminator v3.0 (Japan) or v3.1 (USA) Cycle Sequencing Kit (Applied Biosystems). The sequencing primers were as follows: M13 (Japan) and those described above for primary gene amplification (Gautier *et al.*, 1994). The sequence results were visualized using Sequencing Scanner v1.0 (Applied Biosystems) and analyzed using GENETYX v7.0 (GENETYX).

Isolation of Triton X-114-soluble proteins, which includes the puroindolines, and protein electrophoresis were carried out on single kernels according to the procedure of Giroux and Morris (1998) (Japan and USA).

## Results

Kernel texture

SKCS hardness indexes (HIs) for all the Asian common wheat varieties ranged from KU3333 (13.5) to KT020-584 (97.8) (Fig. 1), with a corresponding range in standard deviation (50 kernels in each variety) from 7.5 to 17.2. Standard deviation values in this range are indicative of homogeneous samples. Generally, the arbitrary separation of kernel texture classes is set at 50, since the SKCS 4100 kernel texture algorithm was designed to emulate the forerunning ground-grain near-infrared reflectance (NIR) spectrometer system, which sets a

mean value of soft wheat varieties at 25, a mean value of hard wheat varieties at 75 and a separation of classes at 50 (AACC International, 2000, Approved Method 39-70A) (Morris, 2002). Using 50, then, as a delimiter for class boundaries, the 'soft group' comprised 169 varieties (SKCS HI < 50) and the 'hard group' comprised 77 varieties (SKCS HI > 50). The soft wheat varieties (*Pina-D1a/Pinb-D1a*, HI < 50) are presented in Table 1 according to their geographic origin.

Characterization of the puroindoline genes

We next sought to resolve the molecular genetic basis for the hard kernel phenotype exhibited by those varieties with HI > 50. It is now well established that the presence/absence and sequence polymorphism of the puroindolines explains both the major classifications into 'soft' and 'hard', as well as some of the variation within the hard class (Morris, 2002; Bhave and Morris, 2008a, b; Morris and Bhave, 2007). Since, it also seems, at present, that puroindoline b is more polymorphic than puroindoline a, at least in *T. aestivum*, we began our analysis by obtaining gene sequence for puroindoline b. To provide additional coverage of the germplasm, we included the first three lines with HI ranked just less than 50 (KT020-546, HI 48.8; KT020-529, HI 49.3; KU7070, HI 49.5).

Cloning and direct PCR sequencing of puroindoline b identified three of the known mutations associated with hard kernel trait in wheat. These were the G-to-A (designated hereafter 'G/A') SNP at 226 that confers the Gly-46
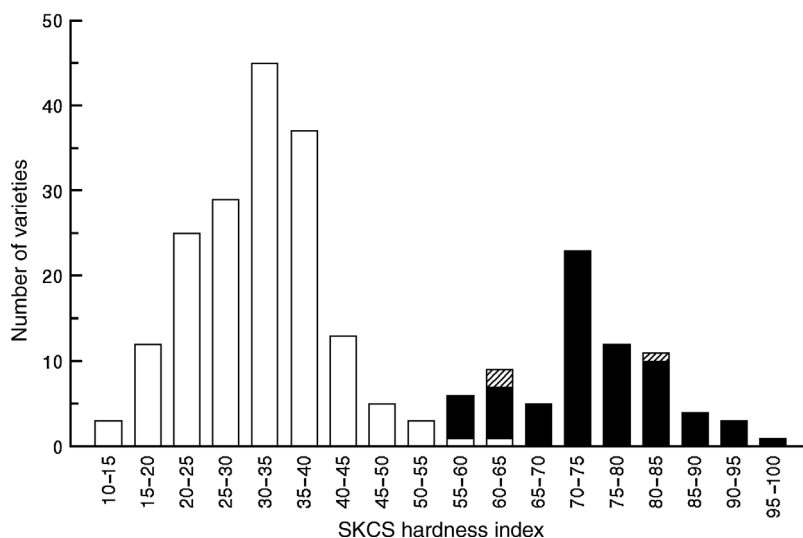
Fig. 1. Frequency distribution histogram of SKCS hardness index of 246 common wheat varieties from Asia. Varieties with the soft 'wild-type' puroindoline haplotype of *Pina-D1a/Pinb-D1a* are denoted by 'open' bars; varieties with the a characterized 'hard' puroindoline haplotype and the 'double null' variety KT020-584 are denoted by solid bars and three varieties with *Pinb-D1a* but unknown puroindoline a are denoted by 'hatched' bars.

to Ser change, the T/C SNP at 266 that confers the Leu-60 to Pro change and the deletion of an A at 210–213. Because the A deletion occurs somewhere in the wild-type sequence of CAAAAT, we have assigned its location to the last of the A's, in other words, at position 213. This SNP produces an open reading frame shift after Thr-41. The current allele assignments for these mutations are *Pinb-D1b*, *Pinb-D1c* and *Pinb-D1p*, respectively (Morris and Bhave, 2007). Of the 80 varieties examined (i.e. 77 with HI > 50, plus three with HI of 48.8–49.5), these three puroindoline b sequences accounted for 53 (Table 2).

Sequencing of puroindoline b led to the discovery of a new SNP that has not been previously described. It is a C/T SNP at position 382 and converts the CAG (Glu) codon to a TAG stop signal (Fig. 2). This SNP was provided the *Pinb-D1ab* allelic designation (Table 2) (Morris and Bhave, 2007). This novel sequence was present in accessions KU3062 and KU3069, both from Afghanistan, and was obtained via cloning (Japan) and from six individual kernels (USA). Its sequence has been entered into GenBank (AB302894).

In all other varieties except one, sequencing produced the 'wild-type' puroindoline b sequence, which is designated *Pinb-D1a*. Repeated efforts to amplify puroindoline b from KT020-584 (HI = 97.8, China, Xinjiang) were unsuccessful.

Next, we characterized puroindoline a from the remaining hard varieties with HI > 50 (17 that did not carry an identified mutation in puroindoline b, i.e. alleles *Pinb-D1b, c, p* and *ab*), plus the three varieties with HIs of 48.8–49.5. Analysis of puroindoline a is more problematic because applying the puroindoline a primers of Gautier *et al.* (1994) do not produce a product in those varieties carrying the *Pina-D1b* mutation, a deletion of 15,380 bp (Giroux and Morris, 1998; Bhave

and Morris, 2008a; Morris and Bhave, 2007). Nevertheless, wild-type puroindoline a sequence was obtained for KT020-529, KT020-546, KU4704, KU4705, KU4739, KU7001, KU7005 and KU7070 (Tables 1 and 2). With the exception of KU4739, these varieties represented seven of the eight softest varieties included in this set of 80 varieties with HI ≥ 48.8. The hardest of these eight lines, KU4739, had an HI of 61.8. In summary, these eight lines had the wild-type *Pina-D1a/Pinb-D1a* haplotype and HIs from 48.8 to 61.8 (Tables 1 and 2).

Three varieties were found to have a single-base-pair deletion of a C at 264–265 (assigned to 265) in puroindoline a, such that a frame shift in the open reading frame was expected from GGC *C*AA ATG to GGC AAA TG and a sequence change from Glu-61 to Lys, and thereafter. The varieties possessing this mutation were KT020-531, KT020-534 and KT020-535, with HIs of 71.2–86.0 (Table 2). This SNP has been designated *Pina-D1l* (Morris and Bhave, 2007).

Of the remaining varieties, 11 consistently produced no product with the puroindoline a primers, whereas three produced ambiguous or inconsistent results after repeated testing. Of the 11, one was KT020-584, which also produced no puroindoline b product and was the hardest of all 246 varieties. It was assumed to be potentially 'null' for both puroindoline genes. The other ten possessed wild-type puroindoline b sequence, and were thus assumed to be *Pina-D1b/Pinb-D1a*.
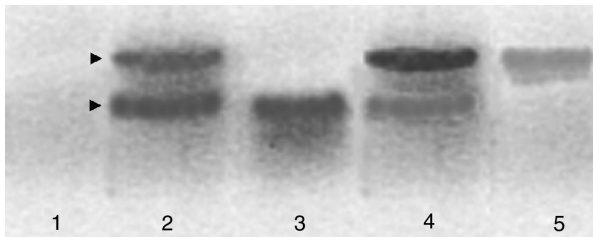
Puroindoline gene expression

We next analyzed puroindoline gene expression of all the varieties by analyzing for the presence of puroindoline a and b proteins in SDS–PAGE (Fig. 3). Generally, SDS–PAGE produced clear results and were consistent with the DNA results according to puroindoline haplotypes. Figure 3 provides examples of the typical results encountered: puroindolines were not present in the durum cv. Stewart, Norin 61 is typical of soft wheat varieties and expresses puroindolines a and b at similar relative levels, Haruyutaka is an 'a-null' (*Pina-D1b*) and lacks puroindoline a expression, Haruminori has the *Pinb-D1b* mutation, which expresses puroindoline b at a lower level relative to puroindoline a, and Haruhikari has the *Pinb-D1c* mutation, which apparently interferes with either the expression or the isolation of puroindoline b.

Of the previously mentioned ten varieties that possessed the wild-type puroindoline b sequence but produced no puroindoline a PCR product, none exhibited puroindoline a on SDS–PAGE, therefore supporting their classification as having the *Pina-D1b/Pinb-D1a* haplotype (Table 2).

```
Nucleic acid
               3               3               3
               7               8               9
               012 345 678 901 234 567 890
Pinb-D1a       ctt cag agg gcc cag agc ctc
Pinb-D1ab      ... ... ... ... t.. ... ...


Amino acid
                                       1   1
               9   9   9   9   9       0   1
               5   6   7   8   9       0   1
Pinb-D1a       L   Q   R   A   Q       S   L
Pinb-D1ab      –   –   –   –   stop    –   –
```

Fig. 2. Partial nucleic acid and amino-acid sequence of puroindoline b showing the wild-type (*Pinb-D1a*) and new SNP mutation (*Pinb-D1ab*). The C/T mutation at position 382 (numbering begins from the ATG start codon) produces a premature stop codon at processed protein position 99 (Glu-to-stop).

Fig. 3. Example of puroindoline gene expression analysis used to characterize 246 common wheat varieties from Asia. Puroindolines a and b were resolved using SDS–PAGE into two bands located around Mr 15k. Both puroindolines a and b are absent in the very hard durum variety, Stewart (lane 1), but present in the soft wheat variety, Norin 61 (lane 2). The hard wheat variety, Haruyutaka (*Pina-D1b/Pinb-D1a*), is an 'a-null' variety with no puroindoline a expression (lane 3). The hard wheat variety Haruminori (*Pina-D1a/Pinb-D1b*) has a mutation in puroindoline b, which generally reduces puroindoline b expression relative to puroindoline a (lane 4). The hard wheat variety Haruhikari (*Pina-D1a/Pinb-D1c*) has a mutation in puroindoline b, which either interferes with the expression or extraction such that puroindoline b is not detected (lane 5). Arrowheads indicate puroindolines a (upper) and b (lower).

SDS–PAGE was used to attempt to 'diagnose' the four varieties (Norin 61, Haruyutaka, Haruminori and Haruhikari) that were still requiring puroindoline haplotype classification, and to provide further information on some of the varieties in the range of HI where the soft and hard classes may overlap. Figure 4 presents the results of this work and is worthy of a brief explanation. Alpowa near-isogenic lines (Morris and King, 2007) and durum cv. Renville were used as controls. Alpowa is a soft variety and expresses puroindolines a and b. Two of its near-isogenic sister lines possess relevant puroindoline mutations [PI 644084 carries *Pinb-D1e* (lane 1) and PI 644080 carries *Pina-D1b* (lane 2)] representing 'b-null' and 'a-null' controls, respectively. The SDS–PAGE analysis supports the DNA sequence data and the designation
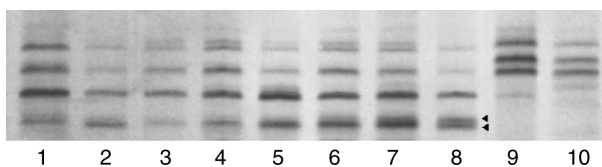
Fig. 4. SDS–PAGE of Triton X-114 proteins from individual kernels of Alpowa near-isogenic line (NIL) PI 644084 puroindoline 'b-null' *Pinb-D1e* (lane 1), PI 644080 puroindoline 'a-null' *Pina-D1b* (lane 2), KU4706 (HI 74.9) puroindoline 'a-null' *Pina-D1b* (lane 3), KT020-531 (HI 71.2) *Pina-D1l* (lane 4), KU4739 (HI 61.8) *Pina-D1a/Pinb-D1a* (lane 5), Sadabouzu (HI 61.5) unknown/*Pinb-D1a* (lane 6), KU7001 (HI 55.4) *Pina-D1a/Pinb-D1a* (lane 7), Alpowa PI 566596 *Pina-D1a/Pinb-D1a* (lane 8), durum cv. Renville (lane 9) and KT020-584 (HI 97.8) (lane 10). Arrowheads indicate puroindoline a (upper) and puroindoline b (lower).

of KU4706 (HI 74.9) as a puroindoline 'a-null' (*Pina-D1b*) (lane 3). The SNP deletion and ORF shift present in KT020-531 (HI 71.2, *Pina-D1l*) appears to interfere with the translation or extraction of puroindoline a (lane 4). KU4739 was the hardest 'soft' wheat (HI 61.8). Even though its puroindoline haplotype was determined to be *Pina-D1a/Pinb-D1a*, it appears to lack or have a very low relative level of puroindoline a expression (lane 5). Sadabouzu with hardness similar to KU4739 (HI 61.5) was one of three varieties for which repeated efforts at sequencing puroindoline a were unsuccessful or produced ambiguous results. Sadabouzu appeared to produce a heavy puroindoline b band relative to puroindoline a (lane 6). KU7001 with an HI of 55.4 and a puroindoline haplotype of *Pina-D1a/Pinb-D1a* expressed puroindoline b at a relatively higher abundance compared with puroindoline a (lane 7). Lastly, KT020-584 appeared to possess no puroindoline a or b, consistent with its failure to amplify either gene in PCR and its hardest HI (97.8) of any of the 246 varieties included here (lane 10). We refer to it simply as a 'double null' (Table 2).

Geographical distribution of the puroindoline genotypes

All but three of the 246 varieties included in the study, Sadabouzu, KU4740 and KT020-560, were classified according to their puroindoline haplotypes (Tables 1 and 2). The remaining 243 varieties were summarized according to their puroindoline haplotypes and geographic origin (Table 3). The frequency of soft/hard kernel texture and puroindoline hardness haplotypes varied depending on the origin of the varieties. The lowest frequency of hard varieties occurred in Korea and south-western Japan. Tibet and Pakistan also had low frequencies of hard varieties. The highest frequency of hard varieties appeared in north-east China followed by Afghanistan and Xinjiang. It should be noted that the 13 varieties from China 'region unknown' could significantly shift these frequencies up or down. Also, there was no overt effort made to draw 'representative' samples of varieties from any of the regions, so that the results reported here are purely 'observational'.

Given those caveats, it is nevertheless of interest to examine the frequency of specific puroindoline hardness alleles due to their reflection of germplasm movement and utilization, as well as the interaction of wheat with human culture. The *Pinb-D1b* allele was the most prevalent across eastern Asia, being present in nearly half of all hard varieties (Table 3). Of the less frequent alleles, three of the four *Pinb-D1c* varieties were from north-east Japan. The *Pinb-D1p* allele was found predominantly in

Table 3. Frequency of puroindoline haplotypes of 246 Asian common wheat varieties based on geographic origin, the total number of lines from each region and the percentage of hard kernel lines from each region

Collection/breeding area	Puroindoline haplotype *Pina-D1/Pinb-D1*							Total	% hard[b]
	a/a	a/b	a/c	a/p	b/a	l/a	other[a]		
Afghanistan	2	0	0	1	1	0	2	6	67
Bhutan	5	0	0	0	2	0	0	7	29
China (east)	19	6	0	4	0	2	1	32	41
China (inner Mongolia)	4	2	0	2	0	0	0	8	50
China (north-east)	2	5	0	1	1	0	0	9	78
China (north-west)	6	2	0	6	1	0	0	15	60
China (Sichuan)	9	1	0	0	1	0	0	11	18
China (region unknown)	6	3	1	2	0	1	0	13	53
China (Tibet)	15	0	0	1	0	0	0	16	6
China (Xinjiang)	1	0	0	1	0	0	1	3	67
Japan (north-east)	37	9	3	0	2	0	0	51	30
Japan (south-west)	53	1	0	0	0	0	1	55	4
Korea	4	0	0	0	0	0	0	4	0
Nepal	3	1	0	0	2	0	1	7	57
Pakistan	8	0	0	1	0	0	0	9	11
Total	174	30	4	19	10	3	6	246	29

[a] KU3062 and KU3069 are *Pinb-D1ab* from Afghanistan; KT020-584 is a 'double null' from Xinjiang; KT020-560, Sadabouzu and KU4740 are unknown *Pina/Pinb-D1a* from east China, south-west Japan and Nepal, respectively.
[b] Based on SKCS HI delineation of 55.7 for soft and hard classes, KU4739 (Nepal) with an HI of 61.8 was the only hard wheat that possessed the *Pina-D1a/Pinb-D1a* haplotype.

a geographic band ranging across northern China and Inner Mongolia. Although the numbers were not great, there seemed to be a notable occurrence of *Pina-D1b* among the few varieties from Nepal and Bhutan (two each, 31% of all varieties from these two countries). All three *Pina-D1l* varieties came from eastern China/China 'unknown'. Lastly, the newly discovered SNP was associated with Afghanistan ($n = 2$) and the 'double null' with Xinjiang.

Discussion

Early work on wheat kernel texture and its relationship with puroindoline a and b gene sequence indicated that a limited number of puroindoline haplotypes accounted for the hard texture observed among wheat varieties of North America (Giroux and Morris, 1998; Morris *et al.*, 2001). Subsequent surveys conducted on populations from the United States, Canada and Australia would generally support this conclusion (Cane *et al.*, 2004; Lillemo *et al.*, 2006; Pickering and Bhave, 2007). However, these results also largely reflect the ancestral wheat varieties introduced during European colonization, and germplasm introductions made during the early stages of wheat breeding and hybridization in these countries. Accordingly, these puroindoline haplotypes are quite similar to those encountered in Europe (Lillemo and Morris, 2000; Corona *et al.*, 2001; Pogna *et al.*, 2002; Bagulho *et al.*, 2003; Gazza *et al.*, 2005; Huang and

Röder, 2005; Ravel *et al.*, 2006). Our survey and others (Chen *et al.*, 2005; Ikeda *et al.*, 2005; Xia *et al.*, 2005; Chang *et al.*, 2006) conducted on wheat varieties of eastern Asia indicate some notable differences in both haplotype frequency and richer diversity in mutations/alleles. It is noteworthy that the survey of Ram *et al.* (2002) of Indian wheat varieties bears some relationship with the current puroindoline haplotype composition of the CIMMYT breeding programme (Lillemo *et al.*, 2006).

In the present study, most (72.5%) of the wheat varieties possessed the wild-type soft puroindoline haplotypes (Table 3). However, this frequency varied considerably depending on geographic origin. For example, of the seven regions with ten or more varieties, the frequency of hard varieties ranged from 2% (south-west Japan) to 60% in north-west China. Obviously, though reasonably close geographically, some other factor is likely involved in the separation of these gene pools. Other contrasts were between north-east China (78% hard) and Korea (0% hard). Naturally, more extensive and systematic surveys will be required to resolve these associations.

Among the 'hard' wheat varieties carrying an identified puroindoline haplotype ($n = 74$, Table 2), the most prevalent haplotype was *Pina-D1a/Pinb-D1b* ($n = 30$, Table 3). Interestingly, this haplotype was also often the most frequent among wheat varieties of the United States, Canada, Australia and Europe.

The second most prevalent haplotype was *Pina-D1a/ Pinb-D1p* ($n = 19$, Table 3). This SNP deletion appears

to be common in eastern Asia but not the Occident. Within Asia, the *Pinb-D1p* allele appears in a geographic region extending from north-eastern China through Inner Mongolia, north-western China, Xinjiang and Tibet, with the greatest frequency in north-western China (Table 3). This allele was also present in Pakistan and Afghanistan but was not found in Japan. As such, its distribution may be associated with the so-called 'Silk Road'. Similarly, Nakamura (2002) reported that the *Glu-D1f* allele of high-molecular-weight glutenin might have been introduced along the Silk Road; it appeared frequently in Japanese varieties, especially those in the south-western part of the country. However, *Pinb-D1p* was not detected in south-western Japan. It may be that the population size and genetic diversity were reduced during the transmission of common wheat to Japan. Tsujimoto *et al.* (1998) argued that the low variation between Japanese and eastern Asian wheat varieties was due to a decreased population size during the transmission of common wheat to these areas, and that plants with less genetic variation became the founders of the present common wheat varieties in the region. Tanaka *et al.* (2003) also reported that the gliadin patterns of Japanese common wheat varieties differed greatly from those of European countries, and proposed that the limited and specific variation in seed storage protein composition of Japanese varieties may simply be attributable to the limited variability in germplasm used in breeding programmes.

A few specific varieties are worthy of further discussion. The first are KU3062 and KU3069 as they possess a new puroindoline mutation in puroindoline b, a C/T SNP at position 382 that confers a premature stop codon. At HIs of 71.0 and 76.5, this mutation is consistent with the 'loss-of-function' model for the kernel softening effect of the two puroindolines. The next variety worth discussing further is KT020-584, which appears to be a 'double null' (Table 2). Tranquilli *et al.* (2002) first identified a puroindoline 'double null' in the substitution line Chinese Spring (Red Egyptian 5D) [CS(RE5D)]. Subsequently, Ikeda *et al.* (2005) discovered three Korean and two American varieties of this type. In the case of Tranquilli *et al.* (2002), their adjusted HI for CS(RE5D) was 95.6, the hardest of all varieties tested. In our study, KT020-584 was also the hardest with an HI of 97.8. It seems that the complete absence of both puroindolines, as it occurs in these varieties and in durum, confers the hardest kernel texture. Lastly, even though KU4739, Sadabouzu and KU7001 expressed both puroindolines a and b, they were classified in the 'hard group'. All of them had a low relative level of puroindoline a expression compared with puroindoline b. Capparelli *et al.* (2003) reported that puroindoline a expression and presence control the abundance of total puroindoline and its association with starch. Hogg *et al.* (2004) also reported that puroindolines a and b interact and together affect wheat grain texture.

In closing it is of interest to consider other possible factors affecting the frequencies of the puroindoline haplotypes. One of these is cultural preference for specific food products. In other words, the prevalence of some puroindoline haplotypes may be attributed to selective preferences for food quality/processing traits in a given region. We observed that the *Pina-D1a/Pinb-D1p* genotype appeared most frequently in China. Therefore, this haplotype may have been preferable with respect to Chinese food culture, typified by noodles and dumpling, as opposed to bread in Europe. Although almost all the varieties in south-west Japan are used for noodles, this genotype did not appear in Japan. However, the white firm noodles in China are quite different from the elastic soft udon noodles of Japan. In this regard, it is perhaps equally, if not more, probable that the genetic constitution of wheat in various regions largely reflects the early movement and adaptation to local environments, which then set the stage for influencing the quality attributes and styles of foods that evolved from those particular wheat varieties. The challenge for current wheat improvement is to maintain and improve the culinary end-use qualities of new hybridizations, while advancing yield and resistance to biotic and abiotic stresses. In this regard, wheat and humankind enjoy an interesting interspecific relationship.

Acknowledgements

References

AACC International (2000) *Approved Methods of the American Association of Cereal Chemists*, 10th ed. Method 39-70A. Paul, MN: The Association.

Bagulho AS, Muacho MC, Carrillo JM and Brites C (2003) Influence of glutenin and puroindoline composition on the quality of bread wheat varieties grown in Portugal. In: Lafiandra D, Masci S and D'Ovidio R (eds) *The Gluten Proteins*. Cambridge, UK: The Royal Society of Chemistry, pp. 113–116.

Baker RJ (1977) Inheritance of kernel hardness in spring wheat. *Crop Science* 17: 960–962.

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

Bhave M and Morris CF (2008b) Molecular genetics of puro-indolines and related genes: regulation of expression, membrane binding properties and applications. *Plant Molecular Biology*, 66: 221–231.

Campbell KG, Bergman CJ, Gualberto DG, Anderson JA, Giroux MJ, Hareland G, Fulcher RG, Sorrells ME and Finney PL (1999) Quantitative trait loci associated with kernel traits in soft × hard wheat cross. *Crop Science* 39: 1184–1195.

Cane K, Spackman M and Eagles HA (2004) Puroindoline genes and their effects on grain quality traits in southern Australian wheat cultivars. *Australian Journal of Agricultural Research* 55: 89–95.

Capparelli R, Borriello G, Giroux MJ and Amoroso MG (2003) Puroindoline A-gene expression is involved in association of puroindolines to starch. *Theoretical and Applied Genetics* 107: 1463–1468.

Chang C, Zhang H, Xu J, Li W, Liu G, You M and Li B (2006) Identification of allelic variations of puroindoline genes controlling grain hardness in wheat using a modified denaturing PAGE. *Euphytica* 152: 225–234.

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

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

Fukunaga K and Inagaki M (1985) Genealogical pedigrees of Japanese wheat cultivars. *Japanese Journal of Breeding* 35: 89–92.

Gautier MF, Aleman ME, 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 Molecular Biology* 25: 43–57.

Gautier MF, 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 Science* 153: 81–91.

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

Giroux MJ and Morris CF (1997) A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. *Theoretical and Applied Genetics* 95: 857–864.

Giroux MJ and Morris CF (1998) Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. *Proceedings of the National Academy of Sciences of the United States of America* 95: 6262–6266.

Greenblatt GA, Bettge AD and Morris CF (1995) Relationship between endosperm texture and the occurrence of friabilin and bound polar lipids on wheat starch. *Cereal Chemistry* 72: 172–176.

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

Huang XQ and Röder MS (2005) Development of SNP assays for genotyping the puroindoline b gene for grain hardness in wheat using pyrosequencing. *Journal of Agricultural and Food Chemistry* 53: 2070–2075.

Hogg AC, Sripo T, Beecher B, Martin JM and Giroux MJ (2004) Wheat puroindolines interact to friabilin and control wheat grain hardness. *Theoretical and Applied Genetics* 108: 1089–1097.

Ikeda TM, Ohnishi N, Nagamine T, Oda S, Hisatomi T and Yano H (2005) Identification of new puroindoline genotypes and their relationship to flour texture among wheat cultivars. *Journal of Cereal Science* 41: 1–6.

Jolly CJ, Rahman S, Korrt AA and Higgins TVJ (1993) Characterization of the wheat Mr 15000 'grain softness protein' and analysis of the relationship between its accumulation in the whole seed and grain softness. *Theoretical and Applied Genetics* 86: 589–597.

Krishnamurthy K and Giroux MJ (2001) Expression of wheat puroindoline genes in transgenic rice enhances grain softness. *Nature Biotechnology* 19: 162–166.

Law CN, Young CF, Brown JWS, Snape JW and Worland JW (1978) The study of grain protein control in wheat using whole chromosome substitution lines. *Seed Protein Improvement by Nuclear Techniques*. Vienna, Austria: International Atomic Energy Agency, pp. 483–502.

Lillemo M and Morris CF (2000) A leucine to proline mutation in puroindoline b is frequently present in hard wheats from Northern Europe. *Theoretical and Applied Genetics* 100: 1100–1107.

Lillemo M, Chen F, Xia X, William M, Pena RJ, Trethowan R and He Z (2006) Puroindoline grain hardness alleles in CIMMYT bread wheat germplasm. *Journal of Cereal Science* 44: 86–92.

Mattern J, Morris R, Schmidt JW and Johnson VA (1973) Location of genes for kernel properties in the wheat cultivar 'Cheyenne' using chromosome substitution lines. In: Sears ER and Sears LMS (eds) *Proceedings of the 4th International Wheat Genetics Symposium*. Columbia, MO: University of Missouri, pp. 703–707.

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

Morris CF and Bhave M (2007) Reconciliation of D-genome puroindoline allele designations with current DNA sequence data. *Journal of Cereal Science*, doi: 10.1016/jjcs2007.09.012.

Morris CF and King GE (2008) Registration of hard kernel puroindoline allele near-isogenic line hexaploid wheat genetic stocks. *Journal of Plant Registrations* 2: 67–68.

Morris CF and Rose SP (1996) Wheat. In: Henry RJ and Kettlewell PS (eds) *Cereal Grain Quality*. New York: Chapman and Hall, pp. 3–54.

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

Murray HG and Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321–4325.

Nakamura H (2002) The geographical diversity of the frequency of the *Glu-D1f* allele in Asian common wheat, and the transmission route through which the wheat may have reached Japan. *Australian Journal of Agricultural Research* 53: 1265–1269.

Nakamura H, Inazu A and Hirano H (1999) Allelic variation in high-molecular-weight glutenin subunit loci of *Glu-1* in Japanese common wheats. *Euphytica* 106: 131–138.

Nishikawa K, Furuta Y and Wada T (1980) Genetic studies on α-amylase isozymes in wheat. III. Intraspecific variation in *Aegilops squarrosa* and birthplace of hexaploid wheat. *Japanese Journal of Genetics* 55: 325–336.

Pickering PA and Bhave M (2007) Comprehensive analysis of Australian hard wheat cultivars shows limited *puroindoline* allele diversity. *Plant Science* 172: 371–379.

Pogna N, Gazza L, Corona V, Zanier R, Niglio A, Mei E, Palumbo M and Boggini G (2002) Chapter 10: puroindolines and kernel hardness in wheat species. In: Ng PKW and Wrigley CW (eds) *Wheat Quality Elucidation, the Bushuk Legacy.* St. Paul, NY: AACC, Inc., pp. 155–169.

Ram S, Boyko E, Giroux MJ and Gill BS (2002) Null mutation in puroindoline A is prevalent in Indian wheats: puroindoline genes are located in the distal part of 5DS. *Journal of Plant Biochemistry and Biotechnology* 11: 79–83.

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Beckert M and Charmet G (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). *Genome* 49: 1131–1139.

Sourdille P, Perretant MR, Charmet G, Leroy P, Gautier MF, Joudrier P, Nelson JC, Sorrells ME and Bernard M (1996) Linkage between RFLP markers and genes affecting kernel hardness in wheat. *Theoretical and Applied Genetics* 93: 580–586.

Symes KJ (1965) The inheritance of grain hardness in wheat as measured by particle size index. *Australian Journal of Agricultural Research* 16: 113–123.

Tanaka H, Tomita M, Tsujimoto H and Yasumuro Y (2003) Limited but specific variations of seed storage proteins in Japanese common wheat (*Triticum aestivum* L.). *Euphytica* 132: 167–174.

Tanaka H, Toyoda S and Tsujimoto H (2005) Diversity of low-molecular-weight glutenin subunit genes in Asian common wheat (*Triticum aestivum* L.). *Breeding Science* 55: 349–354.

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

Tsujimoto H, Yamada T and Sasakuma T (1998) Pedigree of common wheat in east Asia deduced from distribution of the gametocidal inhibitor gene (*Igc1*) and α-amylase isozymes. *Breeding Science* 48: 287–291.

Tsunewaki K (1966) Comparative gene analysis of common wheat and its ancestral species. II. Waxiness, growth habit and awnedness. *Japanese Journal of Botany* 19: 175–229.

Xia L, Chen F, He Z, Chen X and Morris CF (2005) Occurrence of puroindoline alleles in Chinese winter wheats. *Cereal Chemistry* 82: 38–43.