ORIGINAL PAPER

Characterization of three low-molecular-weight Glu-D3 subunit genes in common wheat

X. L. Zhao · X. C. Xia · Z. H. He · K. R. Gale · Z. S. Lei · R. Appels · W. Ma

Received: 17 October 2005 / Accepted: 25 July 2006 / Published online: 29 August 2006 © Springer-Verlag 2006

Abstract Low-molecular-weight glutenins (LMW-GS) in common wheat (*Triticum aestivum* L.) are of great importance for processing quality of pan bread and noodles. The objectives of this study are to identify LMW-GS coding genes at *GluD3* locus on chromosome 1D and to establish relationships between these genes and *GluD3* alleles (a, b, c, d, and e) defined by protein elec-

Communicated by M. Morgante

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0379-y and is accessible for authorized users.

X. L. Zhao · X. C. Xia · Z. H. He (⊠) Institute of Crop Science/National Wheat Improvement Center, Chinese Academy of Agricultural Sciences (CAAS), Zhongguancun South Street 12, Beijing 100081, China

R. Appels · W. Ma (⊠) Western Australia Department of Agriculture and Molecular Plant Breeding CRC, State Agriculture Biotechnology Centre, Murdoch University, Murdoch, WA 1650, Australia

X. L. Zhao · Z. S. Lei Wheat Research Institute, Henan Academy of Agricultural Sciences, Zhengzhou, 450002 Henan, China

K. R. Gale · W. Ma CSIRO Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia e-mail: w.ma@murdoch.edu.au

Z. H. He

International Maize and Wheat Improvement Center (CIMMYT) China Office, c/o CAAS, Zhongguancun South Street 12, Beijing 100081, China e-mail: zhhe@public3.bta.net.cn trophoretic mobility. Specific primer sets were designed to amplify each of the three LMW-GS chromosome 1D gene regions including upstream, coding and downstream regions of eight wheat cultivars containing GluD3 a, b, c, d and e alleles. Three LMW-GS genes, designated as GluD3-1, GluD3-2 and GluD3-3, were amplified from the eight wheat cultivars. The allelic variants of these three genes were analysed at the DNA and protein level. GluD3-1 showed two allelic variants or haplotypes, one common to cultivars containing protein alleles a, d and e (designated GluD3-11) and the other was present in cultivars with alleles b and c (designated GluD3-12). Comparing with GluD3-12, a 3-bp deletion was found in the coding region of the N-terminal repetitive domain of GluD3-11, leading to a glutamine deletion at the 116th position. GluD3-2 had three variants at the DNA level in the eight cultivars, which were designated as GluD3-21, GluD3-22 and GluD3-23. In comparison to GluD3-21, a single nucleotide polymorphism (SNP) was detected for GluD3-22 in the signal peptide region, resulting in an amino acid change from alanine to threonine at the 11th position; and 11 mutations were found at GluD3-23, with five in upstream region, four in coding region and two in downstream region, respectively. GluD3-3 had two haplotypes, designated as GluD3-31 and GluD3-32, both belonging to LMW-s glutenin subunits though their first amino acids in N-terminal region are different. Compared with the GenBank GluD3 genes, nucleotide sequences of GluD3-21 and GluD3-23 were the same as X13306 and AB062875, respectively. GluD3-22 and GluD3-11 had only one-base difference from U86027 and AB062865. GluD3-12 was not found in the Gen-Bank database, indicating a newly identified GluD3 gene variation. GluD3-3 was a new gene different from

any other known *GluD3* genes. Analyses of the relationship between *Glu-D3* alleles defined by protein electrophoretic mobility and different *GluD3* gene variations at the DNA or protein level provided molecular basis for DNA based identification of glutenin alleles.

Introduction

Flour from common wheat, Triticum aestivum L., possesses unique processing properties in forming doughs that are viscoelastic and thus suitable for the production of breads and noodles (Shewry et al. 2002). Gluten networks provide the basis for these properties, which are developed when flour is mixed with water (Lindsay and Skerritt 1999; Shewry et al. 2002). The major functional components of the gluten network are high and low molecular weight glutenin subunit proteins (abbreviated as HMW-GS and LMW-GS, respectively), which form disulphide-bonded gluten macropolymer (Gras et al. 2001). These glutenin proteins are highly polymorphic, with different alleles varying in their contribution to the fundamental aspects of dough quality, for example, dough strength and extensibility (Payne 1987; Wesley et al. 1999, 2001; Brites and Carrillo 2001; Luo et al. 2001). HMW-GS and LMW-GS alleles are therefore important targets for marker-assisted selection in breeding for grain quality in wheat cultivars (Gupta et al. 1999; Eagles et al. 2001; Gale 2005; Ma et al. 2005). Due to their clear resolution by gel electrophoresis and low gene copy number, the allelic variation of the HMW-GS and their relationship with wheat quality has been studied extensively (Payne et al. 1981; Payne and Lawrence 1983; Anderson and Green 1989; Shewry et al. 1992), and DNA markers based on polymerase chain reaction (PCR) are available to discriminate the most important Glu-1 alleles (Ma et al. 2003). Resolution of the LMW-GS and the scoring of alleles by direct analysis of proteins have been more difficult due to the larger number of expressed subunits and their overlapping mobility with the abundant gliadin proteins (Singh and Shepherd 1988). For these reasons, the role of individual LMW-GS in the determination of wheat quality is less clear, although some alleles or subunits are clearly beneficial or detrimental (Lee et al. 1999; D'Ovidio and Masci 2004; He et al. 2005). Selection of LMW-GS is not commonly used in breeding programs although LMW-GS often confers balanced wheat qualities (Ma et al. 2005). Compared with Glu-A3 and Glu-B3, alleles at Glu-D3 are more difficult to separate by electrophoresis and thus less information is available (Liu et al. 2005; He et al. 2005). Therefore, characterization of LMW-GS genes and their alleles at *Glu-D3* locus could potentially lead to the practical application of both HMW-GS and LMW-GS in wheat breeding programs.

LMW-GS is encoded by the complex Glu-3 loci (Glu-A3, Glu-B3 and Glu-D3) on the short arms of group 1 chromosomes (Singh and Shepherd 1988; Gupta and Shepherd 1990). Gupta and Shepherd (1990) carried out an extensive survey of LMW glutenin proteins by SDS polyacrylamide gel electrophoresis (SDS-PAGE) in bread wheat cultivars and detected 20 different banding patterns, six controlled by Glu-A3, nine by Glu-B3 and five by Glu-D3. Consequently, six alleles existed for Glu-A3 locus (a, b, c, d, e, f), nine for Glu-B3 (a, b, c, d, e, f, g, h, i), and five for Glu-D3 (a, b, c, d, e), respectively. Typically, Glu-A3 allele contains 1-3 subunits or protein bands on the gel, Glu-B3 allele contains 5-8 subunits, and Glu-D3 allele contains 4-5 subunits. Furthermore, by analysing N-terminal amino acid sequences, 39 different LMW subunits/ alleles were identified in one bread wheat cultivar (Lew et al. 1992). Based on the first N-terminal amino acid of the mature protein, LMW-GS was divided into three types: LMW-m, LMW-s and LMW-i, which correspond to methionine, serine and isoleucine, respectively (Kasarda et al. 1988; Tao and Kasarda 1989; Lew et al. 1992; Cloutier et al. 2001). LMW-GS was further classified into 12 groups by Ikeda et al. (2002) according to deduced amino acid sequences and in particular the number and position of cysteine residues available for inter-molecular disulphide bond formation (Shewry and Tatham 1997). More than 70 genes, or pseudogenes of the LMW-GS family have been cloned and sequenced from several common wheat cultivars (Cloutier et al. 2001; Ikeda et al. 2002; Zhang et al. 2004). D'Ovidio and Masci (2004) constructed a dendrogram comparison between the coding regions of LMW-GS genes and grouped them into five clusters containing members that are encoded by different genomes, but share more than 90% identity. Up to now, more than 20 different Glu-D3 genes from common wheat and Triticum tauschii have been released in Genbank (Ikeda et al. 2002; Johal et al. 2004; Zhao et al 2004). However, the allelic relationship of these genes, and their association with glutenin subunits/ alleles were still unclear. In this study, Glu-D3 genespecific primers were designed to amplify different Glu-D3 genes in eight common wheat cultivars containing five common Glu-D3 alleles (a, b, c, d and e; Gupta and Shepherd 1990) for the purpose of defining allelic variations and establishing the relationship between protein allele mobility variants and the GluD3 genes at the DNA level.

Materials and methods

Wheat stocks

Eight common wheat cultivars (Tasman, Chinese Spring, Silverstar, Sunco, Aroona, Norin61, Hartog, and BT2288A) carrying five common *Glu-D3* alleles (McIntosh et al. 1998) were used to amplify *Glu-D3* genes in this study (Table 1). Chinese Spring and its nulli-tetrasomic lines N1AT1B (nullisomic 1A-tetrasomic 1B), N1BT1D (nullisomic 1B-tetrasomic 1D) and N1DT1B (nullisomic 1D-tetrasomic 1B) provided by R. A. McIntosh at the Plant Breeding Institute, University of Sydney, were used to confirm chromosomal locations of identified genes.

Development of PCR primers

Gene-specific primers were developed based on the DNA sequence alignment of Zhang et al. (2003, 2004). Eight reference *Glu-D3* genes *X13306*, *AB062851*, *AB062864*, *AB062865*, *AB062872*, *AB062873*, *AB062875* and *AY223396* available in GenBank were used for primer development (http://www.ncbi.nlm.nih.gov). Each gene was divided into three parts (upstream, coding and downstream region) to design corresponding primers in order to obtain sequence information beyond the coding region by combining three PCR products of normal length. The forward and reverse

Table 1 Primers used for amplifying the three GluD3 genes

primers for coding region were first designed. The amplified coding regions were then used to design the reverse primers of upstream region. The forward primers of downstream region, and the upstream forward and downstream reverse primers were all designed based on the sequences of *X13306* (Colot et al. 1989).

Primers' screening

A total of 71 primers were designed and 334 primer sets (140 for upstream, 128 for coding, and 66 for downstream region) were tested. Primer screening was conducted in four steps: (1) confirm the chromosomal location of PCR fragments amplified with different primer sets using Chinese Spring and its nulli-tetrasomic lines of N1AT1B, N1BT1D and N1DT1B (Fig. 1a–c), (2) use the primer sets for chromosome 1D to amplify the eight wheat cultivars with protein mobility alleles a, b, c, d and e, (3) sequence PCR fragments amplified from these cultivars, and (4) sequence alignment of the PCR fragments with the Glu-D3 genes available in GenBank using the software DNAMAN (http://www.lynnon.com). Based on the result of sequence alignments, nine pairs of primer sets for three Glu-D3 genes were confirmed with annealing temperature ranging from 58 to 62° C. Primer sequences (5'-3')and their relative locations at reference genes are shown in Table 1.

Primer	Sequence $(5' \rightarrow 3')$	Primer location ^a	Reference gene ^b	Expected size	Ann. temp. (°C)
D3F62	TATAGCCCAGTTCGAATTGGT	125–145	X13306	901 ^c	62
D3R66	CGCCGCAATGGCAAGGAGA	1,350-1,368	GluD3–1		
D3F621	GCTACAACCACTCAACCACA	130-150	GluD3-1	850	58
D3R661	GAAAGACGAAACTTGTGTGGA	932-952	GluD3-1		
D3F3	AAGATCATCACAGGCACAATC	1,269–1,289	GluD3-1	941	58
D3R3	CTGCTGACCCAATTGTTGTAG	2,182-2,202	GluD3-1		
D3F71	GCAAGAACAACAACAGGGTTT	2,043-2,063	GluD3-1	721	58
D3R8	AACTGGTTGTGATTGTCACTG	2,312-2,332	X13306		
D3F63	GCTTGTGACTTTGAGGCCT	98–116	X13306	982	58
D3R64	ATGGTCTCTCCAAACCAGGGA	926-946	GluD3-2		
D3F36	GAAGGAAAAGAGGTGGTTCCT	743-763	GluD3-2	904	62
D3R37	CTATCTGGTGTGGGCTGCAAA	1,626-1,645	GluD3-2		
D3F8	TCCTGCAAGAACAACAACAG	1,475-1,494	GluD3-2	702	58
D3R9	TAGACAGTGGTACTAGTTGGT	2,292-2,311	X13306		
D3F67	GGGCTGGGCTTGTGACTTT	91-109	X13306	989	59
D3R62	TGGTTTCTCCAAACCAGGGAT	948-968	GluD3-3		
D3F22	CCAAACGGAATAATTAAAGCTA	738–759	GluD3-3	974	59
D3R31	GTTGGGGTTGGGAAACACA	1,710-1,728	GluD3-3		
D3F45	CTATTTGTTCATCCATCTATC	1,422–1,442	GluD3-3	941	60
D3R8	AACTGGTTGTGATTGTCACTG	2,312-2,332	X13306		
	Primer D3F62 D3R66 D3F621 D3R661 D3F3 D3R3 D3F71 D3R8 D3F63 D3F63 D3R64 D3F36 D3R37 D3F8 D3R9 D3F67 D3R62 D3F67 D3R62 D3F22 D3R31 D3F45 D3R8	PrimerSequence $(5' \rightarrow 3')$ D3F62TATAGCCCAGTTCGAATTGGTD3R66CGCCGCAATGGCAAGGAGAD3F621GCTACAACCACTCAACCACAD3R661GAAAGACGAAACTTGTGTGGGAD3F3AAGATCATCACAGGCACAATCD3R3CTGCTGACCCAATTGTTGTAGD3F71GCAAGAACAACAACAGGGTTTD3R8AACTGGTTGTGATTGTCACTGD3F63GCTTGTGACTTTGAGGCCTD3R64ATGGTCTCTCCAAACCAGGGAD3F36GAAGGAAAAAGAGGTGGTTCCTD3R37CTATCTGGTGTGGCTGCAAAD3F8TCCTGCAAGAACAACAACAGD3R9TAGACAGTGGTACTAGTTGGTD3F67GGGCTGGGCTTGTGACTTTD3R62TGGTTTCTCCAAACCAGGGATD3F22CCAAACGGAATAATTAAAGCTAD3R31GTTGGGGTTGGGAAACACAD3F45CTATTTGTTCATCCATCTATCD3R8AACTGGTTGTGATTGTCACTG	PrimerSequence $(5' \rightarrow 3')$ Primer location ^a D3F62TATAGCCCAGTTCGAATTGGT125–145D3R66CGCCGCAATGGCAAGGAGA1,350–1,368D3F621GCTACAACCACTCAACCACA130–150D3R661GAAAGACGAAACTTGTGTGGGA932–952D3F3AAGATCATCACAGGCACAATC1,269–1,289D3R3CTGCTGACCCAATTGTTGTAG2,182–2,202D3F71GCAAGAACAACAACAGGGTTT2,043–2,063D3R63GCTTGTGACTTTGAGGCCT98–116D3R64ATGGTCTCTCCAAACCAGGGA926–946D3F36GAAGGAAAAAGAGGTGGTTCCT743–763D3F8TCCTGCAAGAACAACAACAACAG1,475–1,494D3R9TAGACAGTGGTTGTGACTTT91–109D3R62TGGTTTCTCCAAACCAGGGAT948–968D3F22CCAAACGGAATAATTAAAGCTA738–759D3R31GTTGGGGTTGGGAAACAACAA1,710–1,728D3F45CTATTTGTCATCCATCATC1,422–1,442D3R8AACTGGTTGTGATTGTCACTG2,312–2,332	PrimerSequence $(5' \rightarrow 3')$ Primer locationaReference genebD3F62TATAGCCCAGTTCGAATTGGT125–145X13306D3R66CGCCGCAATGGCAAGGAGA1,350–1,368GluD3-1D3F621GCTACAACCACTCAACCACA130–150GluD3-1D3R661GAAAGACGAAACTTGTGTGGGA932–952GluD3-1D3R3CTGCTGACCCAATTGTTGTAG2,182–2,202GluD3-1D3R3CTGCTGACCCAATTGTTGTAG2,312–2,332X13306D3F63GCTTGTGACTTTGAGGCCT98–116X13306D3R64ATGGTCTCTCCAAACCAGGGA926–946GluD3-2D3R37CTATCTGGTGTGGCTGCAAA1,626–1,645GluD3-2D3R37CTATCTGGTGTGGCTGCAAA1,626–1,645GluD3-2D3F8TCCTGCAAGAACAACAACAACAG1,475–1,494GluD3-2D3R9TAGACAGTGGTACTATGTGACTTT91–109X13306D3F62TGGTTTCTCCAAACCAGGGAT948–968GluD3-3D3R31GTTGGGGTTGGGAAACAACAA1,710–1,728GluD3-3D3R8AACTGGTTGTGGATTGTGAACTATC1,422–1,442GluD3-3D3R8AACTGGTTGTGAATTGTCATC2,312–2,332X13306	PrimerSequence $(5' \rightarrow 3')$ Primer locationaReference genebExpected sizeD3F62TATAGCCCAGTTCGAATTGGT125–145X13306901°D3R66CGCCGCAATGGCAAGGAGA1,350–1,368GluD3-1D3F621GCTACAACCACTCAACCAAC130–150GluD3-1D3F63GAAGACGAAACTTGTGTGGA932–952GluD3-1D3F3AAGATCATCACAGGCACAATC1,269–1,289GluD3-1D3F71GCAAGAACAACAACAGGGGTTT2,043–2,063GluD3-1D3F63GCTTGTGACCTAATCGTCACTG2,312–2,332X13306D3F63GCTTGTGACTTTGAGGCCT98–116X13306982D3F36GAAGGAAAAGAGGGTGGTTCCT743–763GluD3-2904D3F37CTATCTGGTGTGGCTGCAAA1,626–1,645GluD3-2904D3F8TCCTGCAAGAACAACAACAG1,475–1,494GluD3-2702D3F67GGGCTGGGCTTGTGACTTTGGGT2,292–2,311X13306989D3F67GGGCTGGGCTTGTGACTTT91–109X13306989D3F67GGGCTGGGCTTGTGACTTT91–109X13306989D3F67GGGCTGGGCTTGTGACTTT91–109X13306989D3F67GGGCTGGGCTTGTGACTTT91–109X13306989D3F63GTTGTGGGGTTGGACAACAACAA738–759GluD3-3974D3F31GTTGGGGTTGGGAACACA1,710–1,728GluD3-3974D3F31GTTGGGGTTGGAACAACAA2,312–2,332X13306941

^a The location was counted from the first nucleotide of the available gene fragments

^b Reference genes refer to the genes used for marking the primer locations

a <u>D3F62/D3R66</u> <u>D3F3/D3R3</u> <u>D3F71/D3R8</u>



b <u>D3F63/D3R64 D3F36/D3R37</u> <u>D3F8/D3R9</u>



c <u>D3F67/D3R62</u> <u>D3F22/D3R31</u> <u>D3F45/D3R8</u>



Fig. 1 Electrophoresis of PCR products amplified from Chinese Spring and its nulli-tetrasomic lines using nine specific primer sets in agarose gel. *1* Chinese Spring, *2* N1DT1B, *3* N1BT1A and *4* N1AT1B. **a** *GluD3-1* gene. **b** *GluD3-2* gene. **c** *GluD3-3* gene. *M* DNA Ladder 2000 (100, 250, 500, 750, 1,000, 1,600 and 2,000 bp)

DNA extraction and PCR amplification

Genomic DNA was extracted from seedlings or seeds using modified CTAB procedure (Gale et al. 2001). PCR was performed in a total volume of 40 μ l containing 1.5 mM MgCl₂, 60 ng of genomic DNA, 3U of *Taq* DNA polymerase, 200 μ M of each of dNTPs and 10 pmol of each PCR primer. PCR cycling was 94°C for 5 min followed by 38 cycles of 94°C for 40 s, 58–62°C for 40 s, 72°C for 90 s, and a final extension at 72°C for 5 min.

Sequencing of PCR products

PCR fragments were sequenced by the Invitrogen Biotechnology Co. Ltd (Beijing). In order to eliminate sequence errors, each PCR and sequence procedure was repeated 2–4 times. Sequence analysis and characterization were performed using software DNAMAN (http://www.lynnon.com).

Results

GluD3 genes amplified from Glu-D3 locus

In all, three different GluD3 genes including seven allelic variants at the DNA level were identified at the Glu-D3 locus on 1D chromosome in eight common wheat cultivars. The first gene (designated as *GluD3-1*) was amplified with primer sets D3F62/D3R66 for an upstream region, D3F3/D3R3 for a coding region and D3F71/D3R8 for a downstream region (Table 1), which generated 1,336, 855 and 713 bp products, respectively. The complete sequence of GluD3-1 was assembled with a total length of 2,792 bp. Likewise, the second gene, designated as GluD3-2 with 2,144 bp, was amplified using primer sets D3F63/D3R64 for the upstream region, D3F36/D3R37 for the coding region and D3F8/D3R9 for the downstream region with fragment sizes of 912, 834 and 631 bp, respectively. The third gene, GluD3-3 in 2,326 bp, was amplified with primer sets D3F67/D3R62 for the upstream region, D3F22/D3R31 for the coding region and D3F45/D3R8 for the downstream region, which resulted in 930-, 918and 858-bp products, respectively.

GluD3-1 had 2 haplotypes or allelic variants at the DNA level in the eight wheat cultivars, one common to cultivars containing alleles a, d and e (designated GluD3-11) and another variant was presented in cultivars with alleles b and c (designated GluD3-12). GluD3-1 was a relatively conservative gene and only one mutation was detected between its two corresponding allelic forms, of which GluD3-11 had a 3-bp (CAA) deletion at the position 1,672–1,674 in gene coding region (Fig. 2), leading to a glutamine deletion at the 116th position in N-terminal repetitive domain of the deduced protein (Fig. 5).

GluD3-2 had three allelic forms at the DNA level in the eight wheat cultivars, designated as GluD3-21, GluD3-22 and GluD3-23, respectively. Compared with GluD3-21 (Fig. 3), GluD3-22 had a single nucleotide polymorphism (SNP) with a base G-A substitution at the 880th position in the coding region, leading to an amino-acid mutation from alanine to threonine at 11th position in signal region of deduced peptide (Fig. 6). GluD3-23 had 11 mutations, five SNPs in upstream region, two SNPs in downstream region, and two SNPs and two deletions (a 6-bp and a 3-bp, respectively) in the coding region. In comparison to GluD3-21, the deduced amino acid sequences of GluD3-23 had a single glutamine deletion at 127th position and a doubleglutamine deletion at position 56-57 in N-terminal repetitive domain, an amino acid mutation from valine to leucine at position 152 in C-terminal cysteine-rich

D3-11 D3-12	AGCGCTTACGTCGGCTAAATCCTATTTGGCGCCCTTCAGCGCCCGCC	90 90
D3-11 D3-12	$ccaaaaaggcaaaaagggcgctagagctgggtatcgaaaacgctacaacaccactccaccccttggtgtgattggttacatgttttct \\ ccaaaaggcaaaaagggcgctagagctgggtatcgaaaacgctacaccactcaaccacccttggtgtgtgt$	180 180
D3-11 D3-12	${\tt CGTCTTTTCTTATGCTTCTTTTTTTTTTTTTTTTTTTTT$	270 270
D3-11 D3-12	AAATCGATGAACTTTTCTTTTTGAATTTGGATGATTTTTTCGATGAAAATTTGATGAACGTTTTTTCGGATGAACTTTTTTTT	360 360
D3-11 D3-12	${\tt tccatgaactttttccaattcgatgaactattttcgaattcgatgaactttttccgatgaactttttccgattcgatgaactttttccgattcgatgaactattcgatgaactattttccgatgaactatttttccgatgaactatttttccgatgaacttttttccgatgaacttttttccgatgaacttttttcgaattcgatgaacttttttcgaattcgatgaacttttttcgattcgatgaacttttttcgatgaacttttttcgatgaacttttttcgatgaacttcaatgaacttttttcgatgaacttttttcgatgaacttttttcgatgaacttcaatgaactttttttt$	450 450
D3-11 D3-12	${\tt CTTTTGATCGAATTCGATGAACTTTTTTAAATTTGATGAACTTTTTCAAATTCTGATGAACTTTTTGAAAGGACAAGAGTACATACCGTACTTTTGATCGAATTCGATGAACTTTTTTTAAATTTGATGAACTTTTTCAAATTCTGATGAACTTTTTGAAAGGACAAGAGTACATACCGTA$	540 540
D3-11 D3-12	${\tt GCAAAACGGTTTGTTTTTCCAAGAAAATCAGTACACACTGTAGCATTTTTTGAGCGAGAGAAATAAAAAAGTGATCGAGCGGTAGCTAGGCAAAACGGTTTGTTT$	630 630
D3-11 D3-12	${\tt CGACCGACGGCGACAGCGACGGCGACGCGATGCTCGGCCAGGCCCATGCAAGCGACGCAGTGGGCGCCGTGTTCACTCGCTCCACCTCTGCATCGACCGAC$	720 720
D3-11 D3-12	${\tt TGATGTCTTAGCTTGTAGCAATTGCCATCCTTT} \\ {\tt ACATGTAAAGAGGATTTGATGATGATGATGCCATCCTATAGGCGTCAGTTCATCTT} \\ {\tt GATGTCTCTAGCTTGTAGCAATTGCCATCCTTT} \\ \\ {\tt ACATGTAAAGAGGATTTGATGATGATGATGCTCATGGTCATGGTCATGGTCCATGGTCCATGGTCCATCTT} \\ \\ {\tt ACATGTAAGAGGATTGCATCTT} \\ \\ {\tt ACATGTAAGAGGATTGCATGTCATGTAAGAGGATTTGATGATGATGATGATGCATGGTCATATGGCGTCAGTTCATCTT} \\ \\ {\tt ACATGTAAGAGGTCATGTCATGTCATGTAAGAGGATTTGATGATGATGATGATGATGCATGTCATGGTCATGTCATGTCATGTTCATGTTGATGGTCATGTCATGTCATGTTCATCTT} \\ \\ {\tt ACATGTAAGAGTTGCATGTTGTAGGCATTGGTCATGTTGATGGTCATGTCATGTCATGTCATGTTCATGTTGATGTCATGTTGATGGTCATGTTCATGTTGATGGTCATGTTCATGTTGATGGTCATGTTGATGGTCATGTTCATGTTGATGTTGATGTTCATGTTGTAGGGTCAGTTCATGTTGTTGTTGTGTGTCATGTTGATGGTCATGTTGATGTTGTTGTTGTGTGTG$	810 810
D3-11 D3-12	ATCACCTTAGAGGAAAATACAAAGTTAGTTTTTCGAAAAGCAACCGAGCCTAGAAGAACCGTACCCTCGACACGCAAGGCTTTAGCATAT ATCACCTTAGAGGAAAATACAAAGTTAGTTTTTCGAAAAGCAACCGAGCCTAGAAGAACCGTACCCTCGACACGCAAGGCTTAGCATAT	900 900
D3-11		990
D3-12 D3-11 D3-12	GGTTGCA AAAGTATACTATCCTGATAAGTGCGTG <u>ACATGTAAAGTGAATAAGGTGATCAT</u> CTATAGCAAACTTTAGGATTCCTATAG GGTTGCA AAAGTGATACTATCCTGATAAGTGCGTG <u>ACATGTAAAGTGAATAAGGTGATTCAT</u> CTATAGCAAACTTTAGGATTCTATAACT	990 1080 1080
D3-11 D3-12	TTGTGTATAATCATATGCACAACTAAAAAGCAACTTTGATGATCAATCCAAAAGTACGCTTGTAGTGCAACCATAACAACAAATGTACC TTGTGTATAATCATATGCACAACTAAAAAGCAACTTTGATGATCAATCCAAAAGTACGCTTGTAGTGCAACCTAACAACAATGTACC	1170 1170
D3-11 D3-12	AAAAATCCATTTCAGAAACATCCAAACATAATTATTAAAGCTGATGGAAAGAAGAAGAAGAGATGGTGCCCGGGCTACTATAAATAGGCAT AAAAATCCATTTCAGAAACATCCAAACATAATTATTAAAGCTGATGGAAAGAAGAAGAAGAGATGGTGCCCGGGCTACTATAAATAGGCAT	1260 1260
D3-11	D3F3 GAAGTATC <u>AAGATCATCAAGGCAAAAACCAAGCAATACTAGTTAACACCAATCCACCATGAAGACCTTCCTCATCTTTGCT</u>	1350
D3-12	GAAGTATCAAGATCATCACAGGCACAATCATCAAAACCAAGCAATACTAGTTAACACCAATCCAACCAA	1350
D3-11 D3-12	<u>CTCCTTGCCATTGCGCG</u> ACAAGTGCCATTGCACAAATGGAGACTAGCCGCGTCCCTGGTTTGGAGAAACCATGCCGCAACAACCATTA CTCCTTGCCATTGCGGCGACAAGTGCCATTGCACAAATGGAGACTAGCCGCGTCCCTGGTTTGGAGAAACCATGGCAGCAACAACCATTA	1440 1440
D3-11 D3-12	eq:caccaccaccaccaccaccaccaccaccaccaccaccac	1530 1530
D3-11 D3-12	${\tt tcgcagcaacaaccagcttctgccgcaacagcaaccagctattattatctgcaaccaac$	1620 1620
D3-11 D3-12	eq:cacaacaaccaaccaacaacaacaacaacaacaacaaca	1707 1710
D3-11 D3-12	eq:caccaccatttcacaacaacaaccaccatttcccacaaca	1797 1800
D3-11 D3-12	${\tt caacagtttccacaacaaatccctgttgttcaaccatccgttttgcagcagctaaacccatgcaaggtgttcctccaacagcagtgttcaacagtgttccacaacaaatccctgttgttcaaccatccgttttgcagcagctaaacccatgcaaggtgttcctccaacagcagtgttcctccaacagcagtgtttcaaccatgcagcagtgttcctccaacagcagtgttcctcctccaacagcagtgtttcccaacagcagtgtttcctccaacagcagtgtttcctcaacagcagtgtttcctcc$	1887 1890
D3-11 D3-12	$\label{eq:catcher} A GCCATGTGGCAATGTCGCAACAGGTCTTGCTAGGTCACCAAGTGTGGCAACAGTGTGGCAACAACAACAATGTTGCCAACAGTGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCCAACAGTGTGGCAACAGTGGTGGTGGCAACAGTGGTGGGTG$	1977 1980
D3-11	D3F71 CTGCCGCAAATCCCCGAACAATCCCGCTCTGAGGCAATCCGTGCCATCGTCTACTCCATCATCCT <u>GCAAGAACAACAACAACAAGGGTTT</u> TGTC	2067
D3-12		2070
D3-12		2160
D3-11	TCTTTCCAACAACCTCAACAACTACCAACAACTAGGGTCAGCAGCCTCAACAACAACAACAAGATACCACAGGGTATATTCTTGCAGCCACCACCAG	2247
D3-12	TCTTTCCAACAACCTCAACAACTACAACAACTGGGTCAGCAGCCTCAACAACAACAACAAGATACCACAGGGTATATTCTTGCAGCCACACCAG	2250
D3-11 D3-12	ATATCTCAACTTGAGGTGATGACTTCCATTGCACTCCGTACCTTGCCAACGATGTGCGGTGTCAACGTGCGGTGTAAGGTGGGGTGTACAGCTGGACGACGACGATGTACGACGGGGGGGG	2337 2340
D3-11 D3-12	$attatgccattcagcattggcactggagttggtggctac \\ \begin{tabular}{c} \end{tabular} \end{tabular} gaaaagatttctagtaatatatatgtcgttagatcaccgttgttta \\ attatgccattcagcattggcactggagttggtggctac \\ \begin{tabular}{c} \end{tabular} tabula$	2427 2430
D3-11 D3-12	eq:gatgatagtcgatgatgatgatgatgatgatgatgatgatgatgatgatg	2517 2520
D3-11 D3-12	$\underline{AA} \underline{AA} \underline{AA} \underline{AA} \underline{AA} \underline{AA} \underline{G} \underline{T} \underline{T} \underline{C} \underline{C} \underline{A} \underline{A} \underline{A} \underline{C} \underline{A} \underline{A} \underline{C} \underline{A} \underline{A} \underline{A} \underline{A} \underline{A} \underline{A} \underline{A} A$	2607 2610
D3-11 D3-12	${\tt tcttgtctggtcacaaacctggtgcgttaatttattattattattattattattaataataataa$	2697 2700
D3-11 D3-12	${\tt TGGTTGAGTTGAAGCAAGGACTAACATACAACTCTTTAATGGTTGGAGATTCAAGATCTTATTTTATTGGTTTGTATCTGGTTCGACTTGGTTGAGGACTAACAACATACAACTCTTTATGGTTGGAGATCTAAGATCTTATTTTATTGGTTTGTTCTGGTTCGACACT$	2787 2790
D3-11 D3-12	AC AC	2789 2792

Fig. 2 Alignment of two *GluD3-1* gene haplotypes. The nucleotide mutation locus is *shadowed*. The endosperm boxes, CAAT box, TATA motif and AATAAA polyadenylation signals are *underlined*. The initiation codon and terminator codon are *bold and underlined*. The primer positions are indicated with *arrows*

90 90 90 180 180 180 D3-21 ATCAATCCACCTACGCCTCGAAAAAAGAAATCTATCACTCCACCTCAGCATTGATGTCTCTAGCTTGTAGAAACTGCCATCCTTTACATG 270 270 D3-23 ATCAATCCACCTACGCCTCGAAAAAAGAAATCTATCACTCCACCTCAGCATTGATGTCTCTAGCTTGTAGAAACTGCCATCCTTTACATG 360 D3-22 360 D3-23 TAAAgCGGATTCGATGAGTCATGTCTATAGACGTCAGTTLATCTTATCATCTTACAGGAAAGTACAAAGTTAGTTTTCTGAAAA 360 D3-21 GCAACCGAATATAGAAGAACACCCCACACCCAAGGCTTTACTAATCGAGCATATCCTAACAGCCCCACACATGATTGCAAAACTTAGTCATA 450 D3-22 GCAACCGAATATAGAAGAACACTCCACACTCAAGGCTTTACTAATCGAGCATATCCTAACAGCCCACACATGATTGCAAACTTAGTCATA 450 D3-23 GCAACCGAATATAGAAGAACACTCCACACCTCAAGGCTTTACTAATCGAGCATATCCTAACAGCCCCACACATGATTGCAAACTTAGTCATA 450 D3-21 CACAAGTTTTGCCTTTCTTGTTTACGGCTGACAGCCTATACAAGGTTCCAAACTCGGTTGTAAAAGTGATACTATCTTGATAAGTGTGTG 540 D3-22 CACAAGTTTTGCCTTTCTTGTTTACGGCTGACAGCCTATACAAGGTTCCAAACTCGGTTGTAAAAGTGATACTATCTTGATAAGTGTGTG 540 D3-23 CACAAGTTTTGCCTTTCTTGTTTACGGCTGACAGCCTATACAAGGTTCCAAACTCGGTTGTAAAAGTGATACTATCTTGATAAGTGTGTG 540 630 $\underline{\textbf{ACATGTAAAGTTAATAAGGTGAGTCAT}} \textbf{ATATAGCAAATATCGGGGTTTCTGTACTTTGTGTGTGTACGTATGCACAACTAAAAAATCAACT$ D3-22 630 630 D3-21 TTGATGATCAATATATCCAAAAGTACGCTTGTAGCTAGTGCAAACCTAACCCAATGTAACAAAATAATTCATTTCAGATGGAGCCAAACA 720 D3-22 TTGATGATCAATATATCCAAAAGTACGCTTGTAGCTAGTGCAAACCTAACCCAATGTAACAAAATAATTCATTTCAGATGGAGCCAAACA D3-23 TTGATGATCAATATATCCAAAAGTACGCTTGTAGCTAGTGCAAACCTAACaCAAAATAATTCATTCAGATGGAGCCAAACA 720 D3F36 D3-21 GAATTATTAAAGCTGATGCAAA<u>GAAGGAAAAGAGGTGGTTCCT</u>GGGCTAC<u>TATAA</u>ATAGGCATGAAGTATAAAGATCATCACAAGCACAA 810 D3-22 GAATTATTAAAGCTGATGCAAA<u>GAAGGAAAAGAGGTGGTTCCT</u>GGGCTAC<u>TATAA</u>ATAGGCATGAAGTATAAAGATCATCACAAGCACAA 810 810 D3-21 GCATCAGAACCAAGCAACACTAGTTAACACCAATCCACCATGCAGGACCTTCCTCGTCTTTGCCCTCCTCGCCGTTGCGGCGACAAGTGCA 900 D3-22 GCATCAGAACCAAGCAACCATGTTAACACCAATCCACCATGAAGACCTTCCTCGTCTTTGCCCTCCCCCGCCGTTGCGGCGACAAGTGCA 900 D3-23 GCATCAGAACCAAGCAACACTAGTTAACACCAATCCACCATGCAGGACCTTCCTCGTCTTTGCCCTCCTCGCCGTTGCGGCGACAAGTGCA 900 D3R64 -D3-21 ATTGCGCAGATGGAGACTAGATGCA<u>TCCCTGGTTTGGAGAGAGCACT</u>GGCAGCAGCATCACCATCACCACAACAGACATTTCCACAACAA 990 D3-22 ATTGCGCAGATGGAGACTAGATGCA<u>TCCCTGGTTTGGAGAGACCAT</u>GGCAGCAGCAACCATTACCACCACAACAACAACAACAA 990 D3-23 ATTGCGCAGATGGAGACTAGATGCATCCCTGGTTTGGAGAGACCATGGCAGCAGCAACCATTACCACCACAACAACAACAACAA 990 D3-22 GTTGTTCAGCCATCCATTTTGCAGCAGCTAAACCCATGCAAGGTATTCCTCCAGCAGCGCGGCAGCCCTGTGGCAATGCCACAACGTCTT 1350 D3-23 GTTGTTCAGCCATCCATTTTGCAGCAGCTAAACCCATGCAAGcTATTCCTCCAGCAGCAGCGCGGCGAGCCCTGTGGCAATGCCACAACGTCTT 1341 D3-21 GCTAGGTCGCAAATGTTGCAGCAGAGCAGTTGCCATGTGATGCAACAACAATGTTGCCAGCAGTTGCCGCAAATCCCCCCAGCAATCCCCGC 1440 D3-23 GCTAGGTCGCAAATGTTGCAGCAGGAGCAGTTGCCATGTGGATGCAACAACAATGTTGCCAGCAGTTGCCGCAAATCCCCCCAGCAATCCCCCC 1431 D3F8 _ D3-21 TATGAGGCAATCCGTGCTATCATCTACTCCATCATCCTGCAAGAACAACAACAACAGGGTTCGGGGTTCCCATCCAACATCAGGGCAACCAC 1530 D3-21 CAACAGTTGGGCCAATGTGTTTCCCAACCCCAACAGCAGCAGCAGCAGCAGCAACACCTCCAACAACAACAACAACTGGCACAGGGT 1620 D3-22 CAACAGTTGGGCCAATGTGTTTCCCAACCCCAACAGCAGTCGCAGCAACCAGCGAGCAACCACCTCAACAACAACAACAACTGGCACAGGGT 1620 D3-23 CAACAGTTGGGCCAATGTGTTTCCCCAACCCCAACAGCAGTCGCAGCAGCCACCTCGGGCAACAACCACCTCAACAACAACAATTGGCACAGGGT 1611 D3R37 D3-21 ACCTT<u>TTTGCAGCCACACCAGATAG</u>CTCAGGTTGAGGTGATGACTTCCATTGCGGCTCCGTATCCTGCCAACGATGTGCAGTGTTAATGTG 1710 D3-22 ACCTTTTTGCAGCCACCAGATAGCTCAGCTTGAGGTGATGACTTCCATTGCGCTCCGTATCCTGCCAACGATGTGCAGTGTTAATGTG 1710 D3-23 ACCTTTTTGCAGCCACACCAGATAGCTCAGGCTGAGGTGATGACTTCCATTGCGCTCCGTATCCTGCCAACGATGTGCAGTGTTAATGTG 1701 D3-21 CCGTTGTACAGAACCACCACTAGTGTGCCATTCGCCGTGGCACCGGAGTTGGTGCCTAC<u>TGATAA</u>GGAAAGATCTCTAGTAATATAAA 1800 $\texttt{D3-23} \ \texttt{CGGTTGTACAGAACCACCAGTGTGTGCCATTCGACGTTGGCACCGGAGTTGGTGCCTAC<math>\overline{\texttt{TGATAA}}$ $\texttt{GGAAAGATCTCTAGTAATAATAAA} \ \texttt{1791}$ D3-21 TTGGGTCACCGTTGTTTAGTCGATGGATATGTCGATGCGGTGACAAATAAAGTGTCACACAATGTCATGTGTGACCCGCCCAAACTA 1890 D3-22 TTGGGTCACCGTTGTTTAGTCGATGGATATGTCGATGCAGCGGTGACAAATAAAGTGTCACACAATGTCATGTGTGACCCGCCCAAACTA 1890 D3-23 TTGGGTCACCGTTGTTTAGTCGATGGATATGTCGATGCAGCGGTGACAAATAAAGTGTCACACAATGTCATGTGTGACCCGCCCAAACTA 1881 $\texttt{D3-21} \ \texttt{GTTGTTTAAATTCTGAAATAAAAATAAAAATAAAAGTTGTATCAAGACAATGTTCATATTGGCATTGTGTGGATGTCAATCTGATTGCCATG 1980$ D3-22 GTTGTTTAAATTCTGAAATAAAAATAAAAATAAAATGTTGTATCAAGACAATGTTCATATTGGCATTGTGTGGATGTCAATCTGATTGCCATG 1980 D3-23 GTTGTTTAAATTCTGA<u>AATAAAATAAAATAAA</u>GTTGTATCAAGACAATGTTCATATTGGCATTGTGGGATGTCAATCTGATTGCCATG 1971 2144 2144 2135

Fig. 3 Alignment of three *GluD3-2* haplotypes. The nucleotide mutation loci are *shadowed*. The endosperm boxes, CAAT box, TATA motif and AATAAA polyadenylation signals are *under*-

lined. The initiation codon and terminator codon are *bold and underlined*. The primer positions are indicated with *arrows*

region and a mutation from glycine to asparaginic acid at the position 299 in the final conserved part of the protein (Fig. 6).

GluD3-3 gene also had two allelic forms at the DNA level, one common to cultivars with protein mobility alleles a, b and e and the other from cultivars with c and d (designated as GluD3-31 and GluD3-32, respectively). Sixteen SNPs were detected between the two haplotypes with seven in the upstream region, eight in the coding region and one in the downstream region (Fig. 4). The deduced amino acid sequences showed that the two variants of GluD3-3 belong to LMW-s glutenin subunits even though the N-terminal amino acids of GluD3-32 has been mutated from methionine to isoleucine (Fig. 7) (Ikeda et al 2002). Between the two variants, there were seven amino-acid substitutions, and a synonymous mutation at the 52nd position in the gene-coding region (Fig. 7).

Characterization of GluD3-1, GluD3-2 and GluD3-3 genes and their deduced amino acid sequences

The seven allelic variants of the three GluD3 sequences identified in this study all contain a complete gene-coding sequence, including promoter sequences with the endosperm boxes, CAAT boxes, TATA motif, start codon, terminator sequence with double-stop codons, and AATAAA polyadenylation signals (Figs. 2, 3, 4). Sequence alignments indicated that the homology of DNA sequences was 81.5–88.5% among the three gene sequences, and 99.3–99.9% among different allelic forms within each of the three gene sequences (Table 2, below diagonal). In addition, the upstream sequence of GluD3-1 was about 500 bp longer than those of GluD3-2 and GluD3-3.

The deduced amino-acid sequences of the three genes showed that they all had a single open reading frame (ORF) (Figs. 5, 6, 7). Each allele encodes a highly conserved signal peptide of 20 amino acids and a short N-terminal conserved region with 13 amino acids, which was followed by an N-terminal repetitive domain and a C-terminal conserved domain involving three sub-regions of cysteine-rich, glutamine-rich and final conserved domain. All deduced LMW glutenin subunits showed a typical set of eight conserved cysteine (Cys) residues. Based on the position of the first and the seventh cysteine (Ikeda et al. 2002), the GluD3-1, GluD3-2 and GluD3-3 deduced subunits could be classified into types III, IV and I, respectively. Multiple sequence alignments showed that the amino acid identity was 77.5-88.9% among three genes and 98.0-99.9% among allelic forms of each gene sequences (Table 2). The results indicated that *GluD3-1*, GluD3-2 and GluD3-3 are typical Glu-D3 genes.

The relationship between GluD3 gene haplotypes and Glu-D3 mobility alleles

PCR amplification of GluD3 genes indicated that each of the eight cultivars with five different LMW-GS electrophoretic mobility alleles at GluD3 locus contained at least three different GluD3 genes, all varying across the traditionally recognized alleles defined by protein electrophoretic mobility, i.e., a, b, c, d, e (Table 3). The cultivars Tasman and Chinese Spring (Glu-D3a allele) possessed the allelic forms GluD3-11, GluD3-21, and *GluD3-31*, while Silverstar and Sunco (*Glu-D3b* allele) was associated with haplotypes of GluD3-12, GluD3-22, and GluD3-31. Hartog and Bt2288A (Glu-D3e allele) contained common allelic forms of *GluD3-11*, GluD3-31, and GluD3-21 or GluD3-22 (the latter two should be same at the protein level). Aroona with allele c and Norin 61 with allele d carried two common alleles GluD3-23 and GluD3-32, and, displayed some difference for the gene GluD3-1 with GluD3-12 and GluD3-11, respectively. The results indicated that the cultivars with same protein mobility allele always display the same alleles at protein level.

Discussion

Complexity of primer development

Bread wheat is an allohexaploid species, consisting of three subgenomes (A, B, and D) with high sequence identity (Gu et al. 2004), which makes it difficult to design gene-specific primers. In the development of PCR primers, the homology and specificity among different LMW-GS genes were considered simultaneously. For the 71 primers designed, both the forward primers of the upstream region and reverse primers of the downstream region of the three genes were all designed based on X13306, indicating the high nucleotide identity among LMW-GS genes (Ikeda et al. 2002; D'Ovidio and Masci 2004). Though other primers were developed according to the specific SNPs of relevant genes, the rate of useful primers were much less than expected. In the 334 pairs of primer set, over 86% was not specific to Glu-D3 locus or could not amplify PCR products that could be interpreted. Only 46 pairs (13.8%) were clearly assigned to the *Glu-D3* locus, but most of them could not produce acceptable PCR products from the eight cultivars for sequencing. Even for the primer sets with successful sequencing results, there were also some exceptions where several primer sets designed according to SNPs of one gene also amplified another gene product. A typical example was

D3-31	ATTTGGTGCCCTAACACACACACTCTTACGTTGGGCCTAATCACTCCCTCC	90
D3 - 32	atttggtgccctaacacacacactcttacgttgggcctaatcactccctcc	90
D3-31 D3-32	${\tt TAGATGGAACTGATGATTCTCAAAAAGAAAAAGAGATGGAACTGATTGTCTAAGAAAAAAAA$	180 180
D3-31 D3-32	a catagatggaactgatgattatctcccgaaaaaaaagatggaactgatgacgtgtctagagttaggacccggatatatgtaggaactgatgatgatgatgatgatgatgatgatgatgatgatgat	270 270
D3-31 D3-32	${\tt GTAGCTTGTAAGAAATTACCATCCTTTACATATAACGGGAATTCGATGAGTCATGTCATGCTCTAGAGGCGTCAGTTCATCTTATCATCTTGTAGGAAATTACCATCCTTTACATATAACGGGAATTCGATGAGGCCATGTCATGCTCTAGAGGCGTCAGTTCATCTTATCATCTTATCATCTTATCATCTTATCATCTTATCATC$	360 360
D3-31 D3-32	${\tt TAGAGGAAACTACAAAGTTATTTTTTTTGAAAAGCAACCAAC$	450 450
D3-31 D3-32	$\label{eq:construct} a cagc construct a can be constructed a construct a con$	540 540
D3-31 D3-32	${\tt TGCAAAAGTGATACTATCCTGATAAGTGTGTGTGACATGTAAAGTGAATAAGGTGAGTCATATATAGCAAACATCGGGGTTTCTGTACTTTG TGCAAAAGTGATACTATCCTGATAAGTGTGTGTGACATAGGTGAATAAGGTGAGTCATATAGCAAACATCaGGGTTTCTGTACTTTG TACTTTG TACTTG TACTTG TACTTG TACTTG TACTTTG TACTTG TACTG TACTTG TA$	630 630
D3-31 D3-32	TGTGTGATCATATGCACAATTAAAAAGCAACTTTGATGATCAATCCAAAAGTACGCGTGTAGCTAGTGCAACCTAACGCATTGTACCAAA TGTGTGGATCATATGCACAATTAAAAAGCAACTTTGATGATCAATCCAAAAGTACGtGTGTAGCTAGTGCAACCTAACGCATTGTACCAAA	720 720
D3-31		810
D3-32		810
D3-31 D3-32	TATAAAGATCATCACAAGCACAAGCATCAAAAACCAAGCAACACTAGTTAACACTAATCCACC ATG AAGACCTTCCTCATCTTTGCCCTCC TATAAAGATCATCACAAGCACAAGCATCAAAAACCAAGCAACACTAGTTAACACTAATCCACCA TG AAGACCTTCCTCATCTTTGCCCTCC	900 900
D2 21		0.00
D3-31	TCGCCGTTGCGGCAACAAGTGCCATTGCACAAATtGGAGAATAGCCACATCCCTGGTTTGGAGAAAACCATCGCAGCAACAACCATTACCAC	990 990
D3-31 D3-32	TGCAACAAACATTATCGCACCAACAACAACAACCAGCTGTCCAACAACAACCACAACCACATTTCCACAACAGCAACCATGTTCACAGCAAC TGCAACAAACATTATCGCACCAACAACAACCAGCCGTCCAACAACAACCACCATTTCCACAACAGCAACCATGTTCACAGCAAC	1080 1080
D3-31 D3-32	AACAACCACCATTATCGCAGCAACAACAACCACCATTTTCACAACAACAACCACCATTTTCGCAGCAACAACAACCATCATTTTCGCAGC AACAACCACCATTATCGCAGCAACAACCACCACCATTTTCACAACAACCACCACCATTTTCGCAGCAACAACCACCATCATTTTCGCAGC	1170 1170
D3-31 D3-32	aacaaccaaccattttcaccaccatcattttcgcagcaaccaac	1260 1260
D3-31 D3-32	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$	1350 1350
	D3F45	
D3-31 D3-32	eq:accalcalcalcalcalcalcalcalcalcalcalcalca	1440 1440
D3-31 D3-32	$\underline{\mathbf{TC}} \mathtt{TC} \mathtt{TG} \mathtt{CAGCAGCCATG} \mathtt{CAAGGCTATTCCAGCAGCACATG} \mathtt{CAGCACAAGGCCTTGCAAGGCCCTGCGAAAGT} \mathtt{TC} \mathtt{TC} \mathtt{TG} \mathtt{CAGCAGCAAGGCCATG} \mathtt{CAGCAAGGCCACAAGGCCCTG} \mathtt{CAGCAAGGCC} \mathtt{CAGCAGCAAGGCC} \mathtt{CAGCAGCAAGGCC} \mathtt{CAGCAGCAAGGCC} \mathtt{CAGCAGCAAGGCC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAAGGCC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAAGGC} \mathtt{CAGCAGCAGCAAGGC} \mathtt{CAGCAGCAGCAAGGC} \mathtt{CG} C$	1530 1530
D3-31 D3-32	${\tt TGCAGCAGAGCAGTTGCCATGTGATGCAACAACAATGTTGCCAGCAGTTGCTGCAAAATCCCCCAGCAATCCCGCTATGAGGCAATCCCGTGTGCAGCAGAGCAGTTGCCATGTGATGCAACAACAATGTTGCCAGCAGTTGCCGCAAATCCCCCAGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCCCAGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCCCCAGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCCCCAGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCGCAGTGCAGTGCAGCAATCCCGCAGTGCAGTGCAGTGCAGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCGCCAGCAATCCCGCTATGAGGCAATCCCGGCAATCCCGCTATGAGGCAATCCCGTGTGTGCAGCAATCCCGCAGTGGCAGTGCAGTGGCAGTGCAGTGCAGTGCAGTGGCAGTGCAGTGGCAGTGCAGTGGCAGTGCAGTGGCAGTGCAGTGCAGTGGCAGTGCAGTGGCAGTGCAGTGGCAGTGCAGTGGCAGTGCAGTGGCAGTGCAGTGGCAGTGGCAGTGCAGTGGCAGTGGCAGTGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGCAGTGGCAGGCA$	1620 1620
D3-31 D3-32	$\tt CTATCATCTACTCCATCATCCTGCAAGAACAACAACAAGGTTCAGGGTTCCATCCA$	1710 1710
D2_21	\sim D3K31 \sim D3K51 \sim D3	1000
D3-31		1800
D3-31 D3-32		1890 1890
D3-31 D3-32	CCACTAGTGTGCCCATTCGGCGTTGGCGCCGGAGTTGGTGCCTAC TGATAA GGAAAGGTCTCTAGTAATATATAGTTGGATCACCGTTGTT CCACTAGTGTCCCATTCGGCGTTGGCCCCGGAGTTGGTGCCTAC TGATAA GGAAAGGTCTCTAGTAATATATAGTTGGATCACCGTTGTT	1980 1980
D3-31		2070
D3-31 D3-32	AATAAAATACAAATAAAGTTGTATCAAGACAATGTTCATATTGGCATTGTGGGATGTCCATATGCATGC	2160
D3-31 D3-32	TTTGCTTTCCTGGTCACAAGGCCAACCTGGTGCCTTAATTAA	2250
D3-31 D3-32	GTAAAATTTGGGTTGAACTCTTTATCGGTTGGAGATTTGAGATCTTGTTTTTTATTGGTTTGTATCTAGTTCCAAC GTAAAATTTGGGTTGAACTCTTTATCGGTTGGAGATTTGAGATCTTGTTTTTTTT	2326 2326

Fig. 4 Alignment of two *GluD3-3* haplotypes. The nucleotide mutation loci are *shadowed*. The endosperm boxes, CAAT box, TATA motif and AATAAA polyadenylation signals are

underlined. The initiation codon and terminator codon are *bold* and *underlined*. The primer positions are indicated with *arrows*

the upstream reverse primer D3R61 (TGGCACT-TGTCGCCGCAAT), designed for GluD3-1 gene, but the primer set D3F68/D3F61 amplified the same nucleotide sequence of GluD3-2 as D3F63/D3R64, though the second base (G \rightarrow T) and last base (T \rightarrow C) of D3R61 could not match to the GluD3-2 gene. These results indicated that the so-called conserved sequences of 3'-terminal were relative and the gene-specific primers were difficult to obtain (Ikeda et al. 2002; Zhang et al. 2004). In the present study, the primers

h-aci	id	sequences of two <i>nabla</i> the beginning of N-terminal repetitive domain	n dou
P-1	L1	GIFLQPHQISQLEVMTSIALRTLPTM \mathbf{c} GVNVPLYSSTTIMPFSIGTGVGGY	350
P-1	L2	GIFLQPHQISQLEVMTSIALRTLPTM \mathbf{c} GVNVPLYSSTTIMPFSIGTGVGGY	351
P-1 P-1	L1 L2	L QEQQQGFVQPQQQQPQQSGQGVSQHQQQSQQQQQLGQ C SFQQPQQLQQLGQQPQQQQIPQ QEQQQGFVQPQQQQPQQSGQGVSQHQQQSQQQQLGQ <u>C</u> SFQQPQQLQQLGQQPQQQQIPQ	299 300
P-1	L1	$\label{eq:constraint} \begin{split} \texttt{KVFLQQQ} \texttt{C}\texttt{S}\texttt{HV}\texttt{A}\texttt{M}\texttt{S}\texttt{Q}\texttt{R}\texttt{L}\texttt{A}\texttt{R}\texttt{S}\texttt{Q}\texttt{M}\texttt{W}\texttt{Q}\texttt{Q}\texttt{S}\texttt{C}\texttt{H}\texttt{V}\texttt{M}\texttt{Q}\texttt{Q}\texttt{Q}\texttt{C}\texttt{C}\texttt{Q}\texttt{Q}\texttt{L}\texttt{P}\texttt{Q}\texttt{I}\texttt{P}\texttt{E}\texttt{Q}\texttt{S}\texttt{R}\texttt{S}\texttt{A}\texttt{I}\texttt{R}\texttt{I}\texttt{V}\texttt{S}\texttt{I}\texttt{I}\texttt{I}\\ \texttt{K}\texttt{V}\texttt{F}\texttt{L}\texttt{Q}\texttt{Q}\texttt{C}\texttt{S}\texttt{H}\texttt{V}\texttt{A}\texttt{M}\texttt{S}\texttt{Q}\texttt{R}\texttt{L}\texttt{A}\texttt{S}\texttt{S}\texttt{M}\texttt{M}\texttt{Q}\texttt{Q}\texttt{S}\texttt{S}\texttt{C}\texttt{H}\texttt{V}\texttt{M}\texttt{Q}\texttt{Q}\texttt{Q}\texttt{C}\texttt{C}\texttt{Q}\texttt{Q}\texttt{L}\texttt{P}\texttt{Q}\texttt{I}\texttt{P}\texttt{E}\texttt{Q}\texttt{S}\texttt{R}\texttt{S}\texttt{A}\texttt{I}\texttt{R}\texttt{I}\texttt{V}\texttt{S}\texttt{I}\texttt{I} \\ \texttt{I}\texttt{H}\texttt{M}\texttt{M}\texttt{Q}\texttt{G}\texttt{C}\texttt{G}\texttt{M}\texttt{M}\texttt{M}\texttt{Q}\texttt{G}\texttt{S}\texttt{C}\texttt{H}\texttt{M}\texttt{M}\texttt{Q}\texttt{Q}\texttt{C}\texttt{C}\texttt{Q}\texttt{Q}\texttt{L}\texttt{P}\texttt{Q}\texttt{I}\texttt{P}\texttt{G}\texttt{S}\texttt{S}\texttt{S}\texttt{A}\texttt{I}\texttt{R}\texttt{I}\texttt{V}\texttt{S}\texttt{I}\texttt{I} \\ \texttt{I}\texttt{H}\texttt{H}\texttt{H}\texttt{H}\texttt{H}\texttt{H}\texttt{H}\texttt{H}\texttt{H}H$	239
P-1	L2		240
P-1	L1	QQQQPVLPQQPPFSQQQQPPFSQQQQPSSQQPPFPQQHQQFPQQQ [†] IPVVQPSVLQQLNP C	179
P-1	L2	QQQQPVLPQQPPFSQQQQPPFSQQQQPSSQQPPFPQQHQQFPQQQIPVVQPSVLQQLNP C	180
P-1	L1	IILQQSPFSQQQQPVLPQQQPVIILQQPPFSQQQQPVLPQQPPFSQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ	119
P-1	L2		120
P-1	L1	MKTFLIFALLAIAATSAIAQMETSRVPGLEKPWQQQPLPPQQQPP C SQQQQPFPQQQQPI	60
P-1	L2	MKTFLIFALLAIAATSAIAQMETSRVPGLEKPWQQQPLPPQQQPP C SQQQQPFPQQQQPI	60

Fig. 5 Alignment of deduced amino-acid sequences of two GluD3-1 haplotypes. The positions of the cysteine residues are bold and underlined, and a mutation locus is shadowed. Filled nabla the beginning of short N-terminal conserved region; open

nabla the beginning of N-terminal repetitive domain; down arrow the beginning of the three subregions of C-terminal part of the protein which indicate a cysteine-rich region, a glutamine-rich region and the final conserved part of the protein, respectively

Fig. 6 Alignment of deduced amino-acid sequences of three GluD3-2 haplotypes. The positions of the cysteine residues are bold and underlined, and five mutation loci are shadowed. Filled nabla the beginning of short N-terminal conserved region; open nabla the beginning of N-terminal repetitive domain; down arrow the beginning of the three subregions of C-terminal part of the protein which indicate a cysteine-rich region, a glutamine-rich region and the final conserved part of the protein, respectively

P-21	MKTFLVFALLAVAATSAIAQMETRCIPGLERPWQQQPLPPQQTFPQQPLFSQQQQQQLFP	60
P-22	$MKTFLVFALLtVAATSAIAQMETR{\textbf{C}}IPGLERPWQQQPLPPQQTFPQQPLFSQQQQQLFP$	60
P - 23	MKTFLVFALLAVAATSAIAQMETR C IPGLERPWQQQPLPPQQTFPQQPLFSQQQQLFP	58
P-21	QQPSFSQQQPPFWQQQPPFSQQQPILPQQPPFSQQQQLVLPQQPPFSQQQQPVLPPQQSP	120
P-22	QQPSFSQQQPPFWQQQPPFSQQQPILPQQPPFSQQQQLVLPQQPPFSQQQQPVLPPQQSP	120
P - 23	QQPSFSQQQPPFWQQQPPFSQQQPILPQQPPFSQQQQLVLPQQPPFSQQQQPVLPPQQSP	118
P-21	FPQQQQQHQQLVQQQ IPVVQPSILQQLNPCKVFLQQQCSPVAMPQRLARSQMLQQSSCHV	180
P-22	FPQQQQQHQQLVQQQIPVVQPSILQQLNP C KVFLQQQ C SPVAMPQRLARSQMLQQSS C HV	180
P - 23	$\texttt{FPQQQQ}.\texttt{HQQLVQQQIPVVQPSILQQLNP}{\underline{C}}\texttt{Klflqqq}{\underline{C}}\texttt{SPVAMPQRLARSQMLQQSS}{\underline{C}}\texttt{HV}$	177
P-21	MOOOCCOOLPOIPOOSRYEAIRAIIYSIILUEOOOVOGSIOSOOOOPOOLGOCVSOPOOO	240
P-22	MOOOCCOOLPOIPOOSRYEAIRAIIYSIILOEQOOVOGSIQSOOOOPOOLGOCVSOPOOO	240
P - 23	$\texttt{MQQQ} \underline{\textbf{CC}} \texttt{QQLPQIPQQSRYEAIRAIIYSIILQEQQQVQGSIQSQQQPQQLGQ} \texttt{VSQPQQQ}$	237
P-21	SOOOLGOOPOOOOLAOGTFLOPHOIAOLEVMTSIALRILPTMCSVNVPLYRTTTSVPFGV	300
P-22	SOOOLGOOPOOOOLAOGTFLOPHQIAQLEVMTSIALRILPTMCSVNVPLYRTTTSVPFGV	300
P - 23	$\texttt{SQQQLGQQPQQQQLAQGTFLQPHQIAQLEVMTSIALRILPTM} \underline{\texttt{C}}\texttt{SVNVPLYRTTSVPFdV}$	297
P - 21	GTGVGAY	307
P-22	GTGVGAY	307
P-23	GTGVGAY	304

 ∇

P-31 P-32	$\label{eq:mktflifallavaatsaiaQ} \mbox{mktflifallavaatsaiaQ} \mbox{mktflifallavaatsaiaQ} \mbox{mktflifallavaatsaiaQ} \mbox{ienshipglekpsQQ} \mbox{plplQQ} \mbox{plplQ} $	60 60
P-31 P-32	PQQQP C SQQQQPPLSQQQQPPFSQQQPPFSQQQQPSFSQQQQPPFSQQQQPVL PQQQP C SQQQQPPLSQQQQPPFSQQQQPPFSQQQQPPFSQQQQPPFSQQQQPVi	120 120
P-31 P-32	PQQPSFSQQQLPPFSQQQQPFSQQQQPVLPQQPPFSQQQQPILPQQPPFSQQQQQPVLPQ PQQPSFSQQQLPPFSQQQQPPSQQQQPVLPQQPPFSQQQQQPILPQQPPFSQQQQQPVLPQ	180 180
P-31 P-32	$\begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{l} \end{array} \\ \end{array} $	240 240
P-31 P-32	$\begin{array}{l} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \end{array} \\ \begin{array}{c} \end{array} \\ \begin{array}{c} \end{array} \\ \end{array} $	300 300
P-31 P-32	LAQGTFLQPHQIAQLEVMTSIALRTLPTMCRVNVPLYRTTSVPFGVGAGVGAY	354 354

Fig. 7 Alignment of deduced amino-acid sequences of two GluD3-3 haplotypes. The positions of the eight cysteine residues are marked with bold letter and underlined. Seven mutation loci and a synonymous mutation locus are shadowed. Filled nabla the beginning of short N-terminal conserved region; open nabla the beginning of N-terminal repetitive domain; down arrow the beginning of the three subregions of C-terminal part of the protein which indicate a cysteine-rich region, a glutamine-rich region and the final conserved part of the protein, respectively

Gene	GluD3-11	GluD3-12	GluD3-21	GluD3-22	GluD3-23	GluD3-31	GluD3-32
GluD3-11	100	99.7	79.0	78.7	79.5	77.5	77.8
GluD3-12	99.9	100	79.0	78.7	79.5	77.6	77.9
GluD3-21	81.6	81.6	100	99.7	99.3	88.2	88.9
GluD3-22	81.5	81.5	100	100	99.0	87.9	88.5
GluD3-23	81.6	81.6	99.6	99.5	100	88.1	88.8
GluD3-31	80.3	80.3	88.4	88.4	88.4	100	98.0
GluD3-32	80.4	80.4	88.5	88.5	88.5	99.3	100

 Table 2
 Homology comparison of the seven LMW-GS haplotypes (below diagonal) and their deduced amino-acid sequences (above diagonal) (%, irrespective of the sequence length)

 Table 3
 Relationship between Glu-D3 mobility alleles and GluD3 gene haplotypes

Cultivar	<i>Glu-D3</i> allele ^a	GluD3-11 ^b	GluD3-12 ^b	GluD3-21 ^b	GluD3-22 ^b	GluD3-23 ^b	GluD3-31 ^b	GluD3-32 ^b
Tasman	а	+		+			+	
Chinese Sp.	а	+		+			+	
Silverstar	b		+		+		+	
Sunco	b		+		+		+	
Aroona	с		+			+		+
Norin 61	d	+				+		+
Hartog	е	+			+		+	
BT2288A	е	+		+			+	

These new gene sequences have been submitted to Genbank (The accession numbers of GluD3-11, GluD3-12, GluD3-21, GluD3-22, GluD3-23, GluD3-31, and GluD3-32 are DQ357052, DQ357053, DQ357054, DQ357055, DQ357056, DQ357057, and DQ357058, respectively)

^a Glu-D3 alleles were defined by protein electrophoretic mobility

^b "+" means the gene haplotype is present in the corresponding cultivars

were selected firstly using Chinese Spring nulli-tetrasomic lines, so their specificity of amplification could be confirmed.

Nucleotide mutations of the three genes

Mutations contributing to different Glu-D3 alleles include substitution (transition and transversion), deletion, or insertion of base pairs (Heidenreich 2005). For the DNA sequences of the three genes identified in this study, a total of 29 mutations were detected, including three deletions/insertions and 26 SNPs. In the 26 SNPs, 18 were transitions and eight transversions, indicating that the former mutation type occurs more frequently than the latter at the Glu-D3 locus. It was interesting to note that all three deletions were complete triplet code (two three-base: CAA and one six-base: CAACAA) deletion. This kind of deletion was also found in triticale (Li et al. 2005) and in other common wheat cultivars (Zhao et al. 2004). It was further observed from the multiple alignments of different Glu-3 genes that almost all deletions/insertions were on the basis of the triplet code, i.e. one or multiple triplets except pseudogenes (Ikeda et al. 2002; Johal et al. 2004; Zhang et al. 2004). The triplet deletions/insertions did not cause frame shift and resulted in less variation to its amino acid sequence than frame shift mutation. Furthermore, CAA was the codon of glutamine that is the most abundant amino acid in LMW-GS (Figs. 5, 6), so its deletion might have the least effect on protein properties compared with other amino acid deletion. The average mutation rate of the three genes was about 4.0‰, a little higher than that observed in human genome (approximately one SNP per kilobase, Cooper et al. 1985), but lower than that occurred at *Glu-A3* locus (Zhang et al. 2004).

Comparison of GluD3-1, GluD3-2 and GluD3-3 with other Glu-D3 genes registered in Genbank

Currently, approximately 100 entries related to LMW-GS genes have been registered in Genbank, which include complete genes, partial genes and pseudogenes in common wheat, durum and *Ae. tauschii* (Johal et al. 2004). In total, 63 wheat genes with complete coding sequence including 16 *Glu-A3*, 13 *Glu-B3*, 18 *Glu-D3* and 16 undefined genes (Okita et al. 1985; Pitts et al. 1988; Colot et al. 1989; Van et al. 1995; Masci et al. 1998; Ikeda et al. 2002; D'Ovidio and Masci 2004; Zhao et al. 2004; Ozdemir and Cloutier 2005) were selected to compare with the genes identified in this study. The sequence alignments of the seven allelic forms of three

genes found in this study with 18 Glu-D3 genes (Table 4) from Genbank indicate that *GluD3-1* is highly similar to AB062865, AB062866 and AB062867, of which *GluD3-11* has only 1 bp difference (T to C) at the 1,784th position compared with AB062865 (Table 4). *GluD3-2* is highly identical to X13306, U86027, U86029, AB062875, AY223396 and AY299485, among which the sequences of GluD3-21 and GluD3-23 were the same as X13306 and AB062875, respectively, and shared 99.9% identity with the rest relevant entries (Table 4). GluD3-32 had, in fact, significant differences with AB062863 and AB062864 although they share 99.9-100% identity in the common sequence (Table 4). Compared with GluD3-32, AB062863 had a base substitution from A to G at the 1,416th position and a 42-base deletion from the 1,339th to 1,380th position in its coding region, leading to an amino acid mutation from glutamine to arginine at the181th position and 14-amino-acid deletion from the 160th to 173rd position, respectively (data not shown); AB062864 also revealed a 171-base deletion from the 1,214th to 1,384th position in its coding region, resulting in a deletion of 57-amino-acids from the 117th to 173th position. The above results indicated that *GluD3-3* represents a new gene different from any reported Glu-D3 genes.

It is obvious that there are more than three genes at the *Glu-D3* locus on 1D chromosome because several Identifying *Glu-D3* alleles defined by protein mobility is more difficult than identifying *Glu-A3* and *Glu-B3* alleles by direct protein electrophoresis (Liu et al.

Glu-D3 entries from Genbank do not match the three

genes (GluD3-1, GluD3-2 and GluD3-3) reported in

this study. The three genes characterized in this study

have two, three, and two allelic variations at the DNA level, respectively. Theoretically, each gene defined in

this study encodes a protein subunit and the allelic var-

iation of these genes will result in different protein subunits. Previous studies have all considered *Glu-3*

allele defined by protein mobility as a unit in studying

their genetic effects. There are numerous reports that

different Glu-A3, Glu-B3 and Glu-D3 alleles have

different effects on wheat quality (Gupta et al. 1991;

Morel 1994; Morel et al. 1994; Redaelli et al. 1997; Branlard et al. 2001; He et al. 2005). Due to difficulties

in distinguishing *Glu-D3* alleles by SDS-PAGE, the

genetic study of Glu-D3 alleles is seldom reported (Liu

et al. 2005; He et al. 2005). However, Ma et al (2005)

demonstrated that different Glu-D3 alleles have differ-

ent genetic effects on dough quality. These allelic effects are actually combinations of different genes and

proteins. The genes and their allelic variations

reported in this study make it possible to further dis-

sect genetic effects of Glu-D3 alleles at the DNA or

protein level. This could have potential use in practical

Gene	GluD3-11	GluD3-12	GluD3-21	GluD3-22	GluD3-23	GluD3-31	GluD3-32
AB062865	100^{a}	99.9	86.6	86.5	86.6	85.0	85.0
AB062866	99.3	99.3	85.9	85.9	86.0	84.5	84.5
AB062867	99.7	99.7	86.8	86.8	86.9	85.1	85.2
X13306	78.1	78.1	100^{b}	99.9	99.6	88.4	88.5
U86027	86.7	86.7	99.7	100 ^c	99.4	92.8	93.0
U86029	86.7	86.7	99.9	100 ^c	99.6	92.8	93.0
AB062875	86.6	86.6	99.7	99.6	100 ^d	92.3	92.5
AY223396	86.2	86.2	99.6	99.5	99.9	92.2	92.4
AY299485	86.8	86.8	99.6	99.5	99.9	92.2	92.2
AB062863	83.0	83.0	91.8	91.7	91.8	99.1	99.9
AB062864	85.1	85.1	91.9	91.8	91.7	99.2	100 ^e
U86026	85.1	85.1	88.5	88.4	88.4	86.9	87.1
AB062873	89.1	89.1	90.5	90.4	90.4	86.1	86.2
AB062874	85.7	85.7	89.9	89.8	89.8	88.0	88.2
AB062872	85.1	85.1	91.7	91.7	91.6	88.6	88.6
M11077	83.0	83.0	88.5	88.4	88.3	85.6	85.6
AB062851	86.7	86.7	86.7	86.6	86.9	86.6	86.8
X84961	77.5	77.5	82.1	82.1	82.3	83.5	83.8

breeding program.

Table 4 Identity comparison with Glu-D3 genes in Genbank (%, without considering the sequence length and deletions)

^a GluD3-11 had 1-bp difference from AB062865 at the 1,784th position (T \rightarrow C)

^b GluD3-21 is the same as X13306

^c Based on *GluD3-22*, *U86027* has 1-base (A) deletion at the fourth position backwards; U86029 has 1-base (A) deletion at the 45th position and three-base (AGC) deletion at the 667–669th position

^d GluD3-23 is the same as AB062875

^e Compared with *GluD3-32*, *AB062863* has a 42-base deletion in the mid-encoding region and a base substitution from A to G at the 503rd position, and *AB062864* has 171-base deletion in the mid-coding region

2005; He et al. 2005). PCR based markers are necessary to readily identify *Glu-D3* alleles and study their genetic effects. Zhang et al (2004) developed a series of PCR markers for discriminating *Glu-A3* alleles based on sequence variations of one single gene among alleles. We thoroughly characterized three genes and their variations across the cultivars with five *Glu-D3* alleles and established the relationship between them. Table 3 indicates that different genes have different polymorphism distribution patterns across cultivars, thus providing molecular bases for designing independent PCR based markers. The reported information provides the molecular basis for developing new markers for not only discriminating *Glu-D3* alleles but also for dissecting *Glu-D3* alleles' effects at protein subunit level.

Acknowledgments This project was funded by the National Basic Research Program (2002CB11300), International Collaboration Project on Wheat Improvement from Ministry Of Agriculture (2003-Q01), and Natural Sciences Foundation of the Beijing Municipal Government (5041001).

References

- Anderson OD, Green FC (1989) The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. Theor Appl Genet 77:689–700
- Branlard G, Dardevet M, Saccomano R, Lagoutte F, Gourdon J (2001) Genetic diversity of wheat storage proteins and bread wheat quality. Euphytica 119:59–67
- Brites C, Carrillo JM (2001) Influence of high molecular weight (HMW) and low molecular weight (LMW) glutenin subunits controlled by *Glu-1* and *Glu-3* loci on durum wheat quality. Cereal Chem 78:59–63
- Cloutier S, Rampitsch C, Penner GA, Lukow OM (2001) Cloning and expression of a LMW-i glutenin gene. J Cereal Sci 33:143–154
- Colot V, Bartels D, Thompson R, Flavell R (1989) Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol Gen Genet 216:81–90
- Cooper DN, Smith BA, Cooke H, Niemann S, Schmidtke J (1985) An estimate of unique sequence heterozygosity in human genome. Hum Genet 69:201–205
- D'Ovidio R, Masci S (2004) The low-molecular weight glutenin subunits of wheat gluten. J Cereal Sci 39:321–339
- Eagles HA, Bariana HS, Ogbonnaya FC, Rebetzke GJ, Hollamby GJ, Henry RJ, Henschke PH, Carter M (2001) Implementation of markers in Australian wheat breeding. Aust J Agric Res 52:1349–1356
- Gale KR (2005) Diagnostic DNA markers for quality traits in wheat. J Cereal Sci 41:181–192
- Gale KR, Ma W, Zhang W, Rampling L, Hill AS, Appels R, Morris P, Morrel M (2001) Simple high-throughput DNA markers for genotyping in wheat. In: Eastwood R et al (eds) 10th Australian wheat breeding assembly proceedings, pp 26–31
- Gras PW, Anderssen RS, Keentock M, Bekes F, Appels R (2001) Gluten protein functionality in wheat flour processing: a review. Aust J Agric Res 52:1311–1323

- Gu YQ, Derr DC, Kong XY, Anderson OD (2004) Rapid genome evaluation revealed by comparative sequence analysis of orthologous regions from four Triticease genomes. Plant Physiol 135:459–470
- Gupta RB, Shepherd KW (1990) Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutenin. I. Variation and genetic control of the subunits in hexaploid wheats. Theor Appl Genet 80:65–74
- Gupta RB, Bekes F, Wrigley CW (1991) Prediction of physical dough properties from glutenin subunit composition in bread wheat. Cereal Chem 68:328–333
- Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390
- He ZH, Liu L, Xia XC, Liu JJ, Pena RJ (2005) Composition of HMW and LMW glutenin subunits and their effects on dough properties, pan bread, and noodle quality of Chinese bread wheats. Cereal Chem 82:345–350
- Heidenreich RA (2005) Medical and molecular genetics: mutation and repair of DNA, published online at http://www.eddev.arizona.edu/courses/mmg/Lectures/ Mutation%20and%20Repair%20of%20DNA.pdf
- Ikeda TM, Nagamine T, Fukuoka H, Yano H (2002) Characterization of new low-molecular-weight glutenin subunit genes in wheat. Theor Appl Genet 104:680–687
- Johal J, Gianibelli MC, Rahman S, Morell MK, Gale KR (2004) Characterization of low molecular weight glutenin genes in Aegilops tauschii. Theor Appl Genet 109:1028– 1040
- Kasarda DD, Tao HP, Evans PK, Adalsteins AE, Yuen SW (1988) Sequencing of protein from a single spot of a 2-D gel pattern: N-terminal sequence of a major wheat LMW-glutenin subunit. J Exp Bot 39:899–906
- Lee Y-K, Giaffi M, Appels R, Morell MK (1999) The low molecular weight glutenin subunit proteins of primitive wheat. IV. Functional properties of individual genes. Theor Appl Genet 98:149–155
- Lew EJL, Kuzmicky DD, Kasarda DD (1992) Characterization of low molecular weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino acid sequencing. Cereal Chem 69:508–515
- Li GY, He ZH, Sun QX, Lillemo M, Morris C, Xia XC (2005) Identification of novel secaloindoline-a and secaloindoline-b alleles in CIMMYT hexaplod triticale lines. J Cereal Sci 43:378–386
- Lindsay MP, Skerritt JH (1999) The glutenin macropolymer of wheat flour doughs: structure-function perspectives. Trends Food Sci Technol 10:247–253
- Liu L, He ZH, Yan J, Zhang Y, Xia XC, Pena RJ (2005) Allelic variation at the *Glu-1* and *Glu-3* loci, presence of the 1B.1R translocation, and their effects on mixographic properties in Chinese bread wheat. Euphytica 142:197–204
- Luo C, Griffin WB, Branlard G, Mcneil DL (2001) Comparison of low- and high molecular-weight wheat glutenin allele effects on flour quality. Theor Appl Genet 102:1088–1098
- Ma W, Zhang W, Gale KR (2003) Multiplex-PCR typing of high molecular weight glutenin alleles in wheat. Euphytica 134:51–60
- Ma W, Appels R, Bekes F, Larroque O, Morell MK, Gale KR (2005) Genetic characterisation of dough rheological properties in a wheat doubled haploid population: additive genetic effects and epistatic interactions. Theor Appl Genet 111:410–422
- Masci S, D'Ovidio R, Lafiandra D, Kasarda DD (1998) Characterization of a low-molecular-weight glutenin subunit gene

from bread wheat and the corresponding protein that represents a major subunit of the glutenin polymer. Plant Physiol 118:1147–1158

- McIntosh RA, Hart GE, Devos KM, Gale MD, Rogers WJ (1998) Catalogue of gene symbols for wheat. In: Proceedings of the 9th international wheat genetics symposium, Saskatoon, Canada, pp 1–235
- Morel MH (1994) Acid-polyacrylamide gel electrophoresis of wheat glutenin: a new tool for the separation of high and low molecular weight subunits. Cereal Chem 71:238–242
- Morel MH, Bonicel J, Melas V, Austran JC (1994) Multiple approach on the composition of LMW-GS and its effect on dough properties. In: Gluten Proteins 1993, Association of Cereal Research, Detmold, Germany, pp 244–254
- Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of the alpha-/beta-type and gamma-type gliadin DNA sequences. J Biol Chem 260:8203–8213
- Ozdemir N, Cloutier S (2005) Expression analysis and physical mapping of low-molecular-weight glutenin loci in hexaploid wheat (*Triticum aestivum* L.). Genome 48:401–410
- Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Ann Rev Plant Physiol 38:141–153
- Payne PI, Lawrence GJ (1983) Catalogue of alleles for the complex gene loci, *Glu-A1*, *Glu-B1*, and *Glu-D1* which code for high-molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun 11:29–35
- Payne PI, Holt LM, Law CN (1981) Structural and genetical studies on the high-molecular-weight subunits of wheat glutenin. Theor Appl Genet 60:229–236
- Pitts EG, Rafalski JA, Hedgcoth C (1988) Nucleotide sequence and encoded amino acid sequence of a genomic gene region for a low molecular weight glutenin. Nucleic Acids Res 16:11376
- Redaelli R, Pogna NE, Ng PKW (1997) Effects of prolamins encoded by chromosomes 1B and 1D on the rheological properties of dough in near-isogenic lines of bread wheat. Cereal Chem 74:102–107

- Shewry PR, Tatham AS (1997) Disulphide bonds in wheat gluten proteins. J Cereal Sci 25:207–227
- Shewry PR, Halford NG, Tatham AS (1992) High-molecularweight subunits of wheat glutenin. J Cereal Sci 15:105–120
- Shewry PR, Halford NG, Belton PS, Tathan AS (2002) The structure and properties of gluten: an elastic protein from wheat grain. Biol Sci 357:133–142
- Singh NK, Shepherd KW (1988) Linkage mapping of genes controlling endosperm storage proteins in wheat. 1. Genes on the short arms of group-1 chromosomes. Theor Appl Genet 75:628–641
- Tao HP, Kasarda DD (1989) Two-dimensional gel mapping and N-terminal sequencing of LMW-Glutenin subunit. J Exp Bot 40:1015–1020
- Van CS, Vander SJ, Sagi L, Volckaert G (1995) Locus-specific primers for LMW glutenin genes on each of the group 1 chromosomes of hexaploid wheat. Theor Appl Genet 91:313–319
- Wesley AS, Lukow OM, Ames N, Kovaks MIP, McKenzie RIH, Brown D (1999) Effect of single substitution of glutenin or gliadin proteins on flour quality of Alpha 16, a Canada Prairie Spring wheat breeder line. Cereal Chem 76:743–747
- Wesley AS, Lukow OM, McKenzie RIH, Ames N, Brown D (2001) Effect of multiple substitution of glutenin or gliadin proteins on flour quality of Canada Prairie Spring wheat. Cereal Chem 78:69–73
- Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of AS-PCR markers for γ-gliadin alleles in *Triticum Aestivum*. Theor Appl Genet 107:130–138
- Zhang W, Gianibelli MC, Rampling L, Gale KR (2004) Characterisation and marker development for low molecular weight glutenin genes from *Glu-A3* alleles of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 108:1409–1419
- Zhao HX, Wang RJ, Guo AG, Hu SW, Sun GL (2004) Development of primers specific for LMW-GS genes located on chromosome 1D and molecular characterization of a gene from *Glu-D3* complex locus in breed wheat. Hereditas 141:193– 198