Genet Resour Crop Evol (2011) 58:1029–1039
DOI 10.1007/s10722-010-9640-y

RESEARCH ARTICLE

Molecular characterization of LMW glutenin genes from *Taeniatherum* Nevski

Chang Han · Ze-Hong Yan · Shou-Fen Dai · Deng-Cai Liu · Yu-Ming Wei · You-Liang Zheng · Xiu-Jin Lan · Yuan-Ying Peng

Received: 11 July 2010 / Accepted: 15 November 2010 / Published online: 2 December 2010

Abstract We characterized 45 LMW glutenin genes from three diploid species of *Taeniatherum* using 63 primer combinations, designed according to 264 genes reported in wheat and related species. The genes had 909–1,059 bp nucleotides and 301–351 amino acids. The deduced peptides shared similar structures with LMW-m proteins of wheat. The 45 genes shared 77.2–99.7% identities in peptide sequence among each other and 60.0–82.0% identities to proteins from wheat and related species. They were divided into five types according to the N-terminals, starting with METSCIP-, METSRVP-, METGRIP-, METGSIP- and VETSCIP-. The last three and some other structural domain variations were not reported previously in the Triticeae. Thirty-three genes encoded full mature proteins with intact ORFs, whereas the other 12 were pseudogenes with incomplete ORFs, in-frame stop codons or frame-shift mutations. Phylogenetic analysis showed that orthologous genes from *Taeniatherum* were more similar to those in the B and D genomes than in the A genome.

Keywords Gene sequencing · LMW glutenin · Sequence analysis · *Taeniatherum*

Abbreviations

INDELS	Insertions and deletions
ORF	Open reading frame
LMW	Low molecular weight

Introduction

Glutenins and gliadins are the major storage proteins determining end-use quality of wheat flours (Payne et al. 1987; D'Ovidio and Masci 2004). The glutenins can be divided into high molecular weight (HMW) glutenins of 70,000–90,000 Da and low molecular weight (LMW) glutenins of 20,000–45,000 Da (D'Ovidio and Masci 2004). LMW glutenins are

Electronic supplementary material The online version of this article (doi:10.1007/s10722-010-9640-y) contains supplementary material, which is available to authorized users.

C. Han · Z.-H. Yan (✉) · S.-F. Dai · D.-C. Liu · Y.-M. Wei · Y.-L. Zheng · X.-J. Lan · Y.-Y. Peng
Triticeae Research Institute, Sichuan Agricultural University, Wenjiang District, No. 555, Northeast Road, 611130 Chengdu, Sichuan, China
e-mail: zhyan104@163.com

Z.-H. Yan · S.-F. Dai · D.-C. Liu · Y.-M. Wei · Y.-L. Zheng · X.-J. Lan · Y.-Y. Peng
Key Laboratory of Crop Genetic Resources and Improvement, Ministry of Education, Sichuan Agricultural University, 625014 Yaan, Sichuan, China

D.-C. Liu
Northwest Plateau Institute of Biology, Chinese Academy of Science, 810001 Xining, Qinghai, China

grouped as B, C, and D types based on electrophoretic mobilities and isoelectric points. They are also classified as LMW-i (isoleucine), LMW-m (methionine), and LMW-s (serine) types based on the first amino acid residue in the N-terminal of mature proteins (Cloutier et al. 2001; Lew et al. 1992; Masci et al. 1993). In hexaploid wheat (*Triticum aestivum* L. 2n = 6x = 42, AABBDD), LMW glutenins, encoded by the orthologues *Glu-A3*, *Glu-B3,* and *Glu-D3* on chromosome arms 1AS, 1BS and 1DS, respectively (Gupta and Shepherd 1990), account for ~60% of the total endosperm storage proteins. The number of LMW glutenin genes in bread wheat was estimated to be as high as 35–40 (Harberd et al. 1985; Sabelli and Shewry 1991; Cassidy et al. 1998).

The LMW glutenins are important quality determinants of wheat flours and the genes involved are well studied. Moreover, some progress has been made in molecular characterization of LMW glutenin genes from various Triticeae species, including *Aegilops* spp. (Johal et al. 2004; Li et al. 2008), *Agropyron elongatum* (Luo et al. 2005), *Secale sylvestre* (Shang et al. 2005), *Crithopsis delileana* (Guo et al. 2008), *Hordeum chilense,* and *H. brevisubulatum* (Pistón et al. 2005; Hou et al. 2006). Sequence comparisons of these genes revealed some differences between wheat and its relatives. For example, some genes from *H. chilense* and *A. elongatum* lack the N-terminal regions in the predicted mature proteins (Luo et al. 2005; Pistón et al. 2005). To further understand the evolution of LMW glutenin genes in the Triticeae and to better utilize them in wheat quality improvement, more genes in wild cereals need to be analyzed.

The *Taeniatherum* Nevski (TaTa, 2n = 2x = 14) genus is a member of the Triticeae. Biosystematically, it is very distantly related to wheat (Frederiksen 1986; Frederiksen and Bothmer 1986). It contains three diploid species, *T. caput-medusae, T. crinitum* and *T. asperum*. However, the LMW glutenin genes in these species are not reported. To exploit potentially new sources of LMW glutenins that can be used for wheat end-use quality improvement and for understanding the relationships among orthologous genes among Triticeae species, we firstly describe the isolation and characterization of LMW glutenin genes from *Taeniatherum* spp.

Materials and methods

Materials and DNA extraction

Three *T. caput-medusae* (PI 598389, PI 577708, and PI 577710), three *T. crinitum* (PI 561094, PI 204577, and PI 220590), and two *T. asperum* accessions (PI 561091 and PI 561092) used in this study were derived from Turkey (except for PI 220590 from Afghanistan) and provided by the USDA-ARS (http://www.ars-grin.gov/) germplasm bank. Seeds were germinated in darkness at 23°C for one week before planting in pots. Young leaves were harvested and crushed into powder after freezing in liquid nitrogen. Total genomic DNA was extracted using 2 × CTAB method (Yan et al. 2002).

Primer design, PCR and sequencing

For designing DNA primers for the isolation of LMW glutenin genes from *Taeniatherum* spp., the nucleotide and amino acid sequences of LMW glutenin genes from wheat and related species were collected from the GenBank database (http://www.ncbi.nih.nlm.gov/). Based on the amino acid at the beginning signal peptide or the end of the C-terminals, these genes were classified into different groups. Within a group, the nucleotides were further used for polymorphism site screening. PCR primers were then designed according to the amino acids in each group. For polymorphic nucleotide sites within a group, degenerate primers were designed. Seven forward and nine reverse primers were designed based on the gene types (Table 1). These primers formed 63 combinations. They were used for amplifying the genomic DNA of the eight accessions *Taeniatherum* spp.

PCR amplifications of LMW glutenin gene fragments were conducted in total volumes of 50 μl in a PTC-200 DNA Cycler (MJ Research, USA). The PCR ingredients were 1.25 U high fidelity *ExTaq* polymerase (Takara, China), 0.2 mM of each dNTP and 1 μM of each primer, and 200–300 ng template DNA. The PCR parameters were 94°C for 4 min to denature the template DNA, followed by 35 cycles at 94°C for 1 min, 58°C for 1 min, and 72°C for 2 min, then a final extension at 72°C for 5 min. The PCR products were separated in 0.8% agrose gels and the

Table 1 PCR primers used for cloning LMW glutenin genes from *Taeniatherum* spp. Stop codons are shown by asterisks (*)

Primer name	Primer sequence	Amino acid sequence
Forward		
PF1	5′-atg, aag, acc, ttc, ctc, att, tgt, g-3′	MKTFLIC
PF2	5′-atg, aag, acc, ttc, ctc, atc/a, ttc/t, g-3′	MKTFLIF
PF3	5′-atg, aag, acc/a, ttc, ctc/t, gtc, ttt, g-3′	MKTFLVF
PF4	5′-atg, aaa, acc, ttc, ctc, gtc, tgt, g-3′	MKTFLVC
PF5	5′-atg, agg, acc, ttc, ctt, gtc, ttt, g-3′	MRTFLVF
PF6	5′-atg, aag, acc, ttc, ccc, gtc, ttt, g-3′	MKTFPVF
RF7	5′-atg, aag, aaa, aac, ctc, gtc, ttt, g-3′	MKKNLVF
Reverse		
PR1	5′-tta, tca, gta, ggc, acc, aac, t-3′	RVGAY**; QVGAY**; GVGAY**
PR2	5′-tta, tca, gta, gac, acc, c/aac, tc-3′	GVGVY**
PR3	5′-tta, ggc, acc, aac, tcc, ggt, gc-3′	TGVGA*
PR4	5′-tta, tca, gta, gca, cca, ctc, cg-3′	PEWCY**
PR5	5′-tta, tt/ca, gta, gcc, acc, aac, tc-3′	GVGGY**
PR6	5′-tta, tca, gta, ggc, act, aac, tc-3′	GVSAY**
PR7	5′-tta, tca, gta, ggc, agc, aac, tc-3′	GVAAY**
PR8	5′-tta, tca, gta, ggg, gcc, aac, tc-3′	GVGPY**
PR9	5′-tta, tta, gta, gga, acc, aac, tc-3′	GVGSY**

targeted DNA fragments were recovered and ligated into pMD18-T vectors (Takara, China).

The ligated products were transformed into *E. coli* DH10B cells and positive clones were selected. At least one DNA fragment derived by each primer pair was used for cloning and sequencing. Three clones for each candidate DNA fragment were sequenced. The sequnence was determined by sequnecing of three clones at two directions.

Sequence alignment and phylogenetic analysis

Sequence alignments were conducted by Clustal W (Thompson et al. 1994). The deduced amino acid sequences of 45 genes from *Taeniatherum* and 26 homologous genes from four diploid species of wheat relatives, including eight from *Ae. tauschii* (Johal et al. 2004; Pei et al. 2007; Huang and Cloutier 2008), nine from *Ae. longissima* (Jiang et al. 2008), and nine from *T. monococcum* and *T. urartu* (An et al. 2006), were used to construct a topology tree to elucidate the evolutionary relationships among them. The phylogenetic tree was established using the deduced protein sequences by MEGA 4.0 (Tamura et al. 2007). For the analysis, the Neighbor-Joining (NJ) method and the complete deletion option were used with respect to gaps in the aligned sequences. Bootstrap values were estimated based on 1,000 replications. At the same time, evolutionary distances were measured by calculating *p*-distances for each pair of aligned sequences.

Results

LMW glutenin genes in GenBank

At January 24, 2009, 264 LMW glutenin gene sequences (including complete genes, partial genes and pseudogenes) were in the GenBank database. The sequences were derived from species of five *Triticeae* genera, including *Triticum* (168 sequences), *Aegilops* (42), *Lophopyrum* (25), *Secale* (3), and *Hordeum* (26) (Supplementary Table 1).

According to the seven amino acid residues at the beginning of the signal peptides and the ends of the C-terminals, these genes were divided into 15 and 17 groups, respectively (Table 2). Among them, the signal peptides MKTFLVF and MKTFLIF and the C-terminal GVGAY** and RVGAY** were predominant types, accounting for 45.70% (117/256), 33.6%

Table 2 Signal peptide and C-terminal sequences in 264 LMW glutenin genes from wheat and its relative species

	Signal peptide	No. of sequences	C-terminal	No. of sequences
1	MKTFLVF	117	GVGAY**	126
2	MKTFLIF	86	RVGAY**	30
3	MKTLLIL	12	GVGGY**	16
4	MKTLLIF	11	GVSAY**	14
5	MKNFLVF	6	GVGVY**	13
6	MKTFLIC	6	QVGAY**	13
7	MKTFLVC	5	GVGPY**	8
8	MKTLFIL	4	GVGSY**	6
9	MKTFVVF	2	PPDFWH*	6
10	MKTVLVC	2	PVDFWH*	2
11	MKKNLVF	1	TRVGV**	2
12	MKTFLTF	1	GFGAY**	1
13	MKTFPVF	1	GVAAY**	1
14	MKTLLVF	1	PEWCY**	1
15	MRTFLVF	1	PSVGV**	1
16			TGVGA*	1
17			TVGAYL*	1
Total		256		242

Some partial sequences lacking signal peptide, N-terminal or C-terminal sequences were included among the 264 sequences. The underlined sequences were used for primer design. Asterisks (*) indicate stop codons

(86/256), 52.1% (126/242) and 12.4% (30/242) of the total sequences, respectively.

LMW glutenin sequences from *Taeniatherum* spp.

In a total of 504 PCR reactions, 136 produced candidate LMW glutenin fragments of 0.9 to 1.1 Kb (Supplementary Table 2). Among 63 primer combinations, 34 gave positive amplifications in the eight accessions of *Taeniatherum*. Twenty eight, 26, 8, 20, 18, 16, 15, and 5 of these primer combinations produced candidate LMW glutenins in PI 561094, PI220590, PI204577, PI598389, PI577710, PI577708, PI561092, and PI561091, respectively. We chose 45 DNA fragments (including all DNA fragments from PI 561094 and PI 561091 and 1, 2, 1 and 1 from PI 220590, PI 598389, PI 577710 and PI 577708, respectively) for further use in cloning and sequencing (Table 3, Supplementary Table 2).

Forty-five different LMW glutenin genes, designated *Ta-1* to *Ta-45* (GenBank accessions FJ481524 to FJ481568), were obtained (Table 3). Thirty-three genes encoded complete mature proteins with intact open reading frames (ORFs). The remaining 12 were pseudogenes with incomplete ORFs, caused by in-frame stop codons or frame-shift mutations.

Characterization of LMW glutenin genes from *Taeniatherum* spp.

The sequences of the 45 LMW glutenin genes ranged from 909 to 1,059 bp at the nucleotide level and from 301 to 351 amino acids (Table 3; Fig. 1, Supplementary Fig. 1). Gene length differences were caused by INDELs in the glutamine rich repetitive domains. However, the genes shared a similar primary structure with those of homologous genes in wheat and other relatives in four structural regions: viz. a signal peptide region with 20 residues, a N-terminal region with 13 residues, a repetitive domain rich in glutamine and proline residues and characterized by tandem repeat units, and a C-terminal domain consisting of three subregions that are cysteine rich (I), glutamine rich (II), and a final conserved domain (III). Five signal peptide types occurred among the 45 genes, such as MKKNLVF (4 genes), MKTFLIF (11), MKTFLVF (11), MKTFPVF (8), and MRTFLVF (11). Five different N-terminal sequences were also identified, including METGRIP (1 gene), VETSCIP (1), METGSIP (1), METSRVP (14), and METSCIP (28). In addition, there were 9 C-terminal peptides, viz. GTGVGA* (5 genes), GVAAY** (8), GVGAY** (7), GVGGY** (3), GVGPY** (3), GVGSY** (2), GVGVY** (5), GVSAY** (7), and PEWCY** (5).

Table 3 Characteristics of 45 LMW glutenin genes of *Taeniatherum*

Sequence	NCBI accession	Primer combination	Source	DNA length (bp)	Signal peptide	N-terminal	C-terminal	Protein similarity sequence
Ta-1p	FJ481524	PF2 + PR7	*T. asperum* PI 561091	1,044	MKTFLIF	METGRIP	GVAAY**	81% CAA74550 *T. durum*
Ta-2p	FJ481525	PF3 + RR3	*T. asperum* PI 561091	909	MKTFLVF	METSCIP	GTGVGA*	80% BAB78760 *T. aestivum*
Ta-3	FJ481526	PF6 + PR1	*T. asperum* PI 561091	915	MKTFPVF	METSCIP	GVGAY**	80% BAB78760 *T. aestivum*
Ta-4	FJ481527	PF6 + PR7	*T. asperum* PI 561091	915	MKTFPVF	METSCIP	GVAAY**	79% BAB78760 *T. aestivum*
Ta-5	FJ481528	PF7 + PR7	*T. asperum* PI 561091	915	MKKNLVF	METSCIP	GVAAY**	82% ABM66823 *Ae. geniculata*
Ta-6p	FJ481529	PF3 + PR4	*T. crinitum* PI 220590	1,033	MKTFLVF	METSRVP	PEWCY**@	61% ABM66823 *Ae. geniculata*
Ta-7	FJ481530	PF6 + PR2	*T. crinitum* PI 220590	1035	MKTFPVF	METSRVP	GVGVY**	60% ABM66823 *Ae. geniculata*
Ta-8p	FJ481531	PF2 + PR4	*T. caput-medusae* PI 598389	1,045	MKTFLIF	METSRVP	PEWCY**@	79% CAA74550 *T. durum*
Ta-9p	FJ481532	PF7 + PR1	*T. caput-medusae* PI 577710	915	MKKNLVF	METSCIP	GVGAY**	77% BAB78760 *T. aestivum*
Ta-10	FJ481533	PF2 + PR7	*T. caput-medusae* PI 577708	1,050	MKTFLIF	METSRVP	GVAAY**	80% CAA74550 *T. durum*
Ta-11	FJ481534	PF3 + PR7	*T. caput-medusae* PI 577708	1,050	MKTFLVF	METSRVP	GVAAY**	80% CAA74550 *T. durum*
Ta-12p	FJ481535	PF5 + PR4	*T. caput-medusae* PI 577708	1,048	MRTFLVF	METSRVP	PEWCY**@	80% CAA74550 *T. durum*
Ta-13	FJ481536	PF5 + PR6	*T. caput-medusae* PI 577708	1,050	MRTFLVF	METSRVP	GVSAY**	80% CAA74550 *T. durum*
Ta-14p	FJ481537	PF6 + PR4	*T. caput-medusae* PI 577708	1,048	MKTFPVF	METSRVP	PEWCY**@	80% CAA74550 *T. durum*
Ta-15p	FJ481538	PF2 + PR1	*T. crinitum* PI 561094	915	MKTFLIF	METGSIP	GVGAY**	68% BAB78760 *T. aestivum*
Ta-16	FJ481539	PF2 + PR2	*T. crinitum* PI 561094	1,059	MKTFLIF	METSRVP	GVGVY**	80% CAA74550 *T. durum*
Ta-17	FJ481540	PF2 + PR3	*T. crinitum* PI 561094	909	MKTFLIF	METSCIP	GTGVGA*	80% BAB78760 *T. aestivum*
Ta-18	FJ481541	PF2 + PR5	*T. crinitum* PI 561094	1,059	MKTFLIF	METSRVP	GVGGY**	80% CAA74550 *T. durum*
Ta-19	FJ481542	PF2 + PR6	*T. crinitum* PI 561094	915	MKTFLIF	METSCIP	GVSAY**	79% BAB78760 *T. aestivum*
Ta-20	FJ481543	PF2 + PR6	*T. crinitum* PI 561094	1,059	MKTFLIF	METSRVP	GVSAY**	78% CAA74550 *T. durum*
Ta-21	FJ481544	PF2 + PR7	*T. crinitum* PI 561094	915	MKTFLIF	METSCIP	GVAAY**	79% BAB78760 *T. aestivum*
Ta-22	FJ481545	PF2 + PR8	*T. crinitum* PI 561094	915	MKTFLIF	METSCIP	GVGPY**	78% BAB78760 *T. aestivum*
Ta-23p	FJ481546	PF3 + PR1	*T. crinitum* PI 561094	1,059	MKTFLVF	METSRVP	GVGAY**	81% CAA74550 *T. durum*
Ta-24	FJ481547	PF3 + PR2	*T. crinitum* PI 561094	915	MKTFLVF	METSCIP	GVGVY**	80% BAB78760 *T. aestivum*
Ta-25	FJ481548	PF3 + PR2	*T. crinitum* PI 561094	1,059	MKTFLVF	METSRVP	GVGVY**	80% CAA74550 *T. durum*
Ta-26	FJ481549	PF3 + PR3	*T. crinitum* PI 561094	909	MKTFLVF	VETSCIP	GTGVGA*	80% BAB78760 *T. aestivum*
Ta-27	FJ481550	PF3 + PR6	*T. crinitum* PI 561094	915	MKTFLVF	METSCIP	GVSAY**	82% ABM66823 *Ae. geniculata*
Ta-28	FJ481551	PF3 + PR7	*T. crinitum* PI 561094	915	MKTFLVF	METSCIP	GVAAY**	80% BAB78760 *T. aestivum*
Ta-29	FJ481552	PF3 + PR8	*T. crinitum* PI 561094	915	MKTFLVF	METSCIP	GVGPY**	79% BAB78760 *T. aestivum*
Ta-30	FJ481553	PF3 + PR9	*T. crinitum* PI 561094	915	MKTFLVF	METSCIP	GVGSY**	82% ABM66823 *Ae. geniculata*
Ta-31	FJ481554	PF5 + PR1	*T. crinitum* PI 561094	915	MRTFLVF	METSCIP	GVGAY**	79% BAB78760 *T. aestivum*
Ta-32	FJ481555	PF5 + PR2	*T. crinitum* PI 561094	915	MRTFLVF	METSCIP	GVGVY**	79% BAB78760 *T. aestivum*
Ta-33	FJ481556	PF5 + PR3	*T. crinitum* PI 561094	909	MRTFLVF	METSCIP	GTGVGA*	82% ABM66823 *Ae. geniculata*
Ta-34p	FJ481557	PF5 + PR4	*T. crinitum* PI 561094	913	MRTFLVF	METSCIP	PEWCY**@	82% ABM66823 *Ae. geniculata*

Table 3 continued

Sequence	NCBI accession	Primer combination	Source	DNA length (bp)	Signal peptide	N-terminal	C-terminal	Protein similarity sequence
Ta-35p	FJ481558	PF5 + PR5	*T. crinitum* PI 561094	915	MRTFLVF	METSCIP	GVGGY**	82% ABM66823 *Ae. geniculata*
Ta-36	FJ481559	PF5 + PR5	*T. crinitum* PI 561094	1,059	MRTFLVF	METSRVP	GVGGY**	80% CAA74550 *T. durum*
Ta-37	FJ481560	PF5 + PR6	*T. crinitum* PI 561094	915	MRTFLVF	METSCIP	GVSAY**	82% ABM66823 *Ae. geniculata*
Ta-38	FJ481561	PF5 + PR7	*T. crinitum* PI 561094	915	MRTFLVF	METSCIP	GVAAY**	78% BAB78760 *T. aestivum*
Ta-39p	FJ481562	PF5 + PR8	*T. crinitum* PI 561094	915	MRTFLVF	METSCIP	GVGPY**	82% ABM66823 *Ae. geniculata*
Ta-40	FJ481563	PF6 + PR1	*T. crinitum* PI 561094	915	MKTFPVF	METSCIP	GVGAY**	82% ABM66823 *Ae. geniculata*
Ta-41	FJ481564	PF6 + PR3	*T. crinitum* PI 561094	909	MKTFPVF	METSCIP	GTGVGA*	82% ABM66823 *Ae. geniculata*
Ta-42	FJ481565	PF6 + PR6	*T. crinitum* PI 561094	915	MKTFPVF	METSCIP	GVSAY**	82% ABM66823 *Ae. geniculata*
Ta-43	FJ481566	PF6 + PR9	*T. crinitum* PI 561094	915	MKTFPVF	METSCIP	GVGSY**	82% ABM66823 *Ae. geniculata*
Ta-44	FJ481567	PF7 + PR1	*T. crinitum* PI 561094	915	MKKNLVF	METSCIP	GVGAY**	82% ABM66823 *Ae. geniculata*
Ta-45	FJ481568	PF7 + PR6	*T. crinitum* PI 561094	915	MKKNLVF	METSCIP	GVSAY**	82% ABM66823 *Ae. geniculata*

Pseudogenes and stop codons are indicated by p and asterisks (*), respectively. @ indicates that C-terminal sequences were obtained by ignoring a single base insertion in front of the primer-binding site. The protein sequences CAA74550, BAB78670 and ABM66823 were reported by D'Ovidio et al. (1997); Ikeda et al. (2002), and unpublished data, respectively

The 45 genes shared 77.2–99.7% identities in peptide sequences among each other (data not shown) and 60.0–82.0% identities with those in wheat and other relatives. Thirteen genes (*Ta-1, 8, 10, 11, 12, 13, 14, 16, 18, 20, 23, 25* and *36*) showed a high similarity of 78.0 to 81.0% to CAA74550 (D'Ovidio et al. 1997), a durum wheat *Glu-B3* encoded LMW glutenin. Sixteen genes (*Ta-2, 3, 4, 9, 15, 17, 19, 21, 22, 24, 26, 28, 29, 31, 32* and *38*) showed a high similarity of 68.0 to 80.0% to BAB78760, a common wheat (*T. aestivum*) *Glu-D3* encoded LMW glutenin (Ikeda et al. 2002; D'Ovidio and Masci 2004). The remaining 16 genes (*Ta-5, 6, 7, 27, 30, 33, 34, 35, 37, 39, 40, 41, 42, 43, 44* and *45*) showed a high similarity of 60.0 to 82.0% to an *Ae. geniculata* LMW glutenin ABM66823 (unpublished data).

Forty-four genes started with methionine in the N-terminal and were therefore LMW-m types. Among them, *Ta-1* and *Ta-15* had unique N-terminal sequences of METGRIP and METGSIP, respectively. However, gene *Ta-26* started with valine (Val) at the N-terminal and had a unique N-terminal sequence of VETSCIP. These three N-terminal structures were not reported previously in Triticeae species. The genes with N-terminal METSRVP or METGSIP were longer at the amino acid level than those with M(V)ETSCIP or METGRIP.

Genes *Ta-18* and *Ta-36* should produce the same mature proteins after signal peptide removal (Supplementary Fig. 2a). However, their signal peptide regions differed by four single base mutations. Similarly, there were three, three, and four single base differences between *Ta-24* and *Ta-32* (Supplementary Fig. 2b), *Ta-17* and *Ta-41* (Supplementary Fig. 2c), and among *Ta-4, Ta-21* and *Ta-28* (Supplementary Fig. 2d), respectively.

Twelve pseudogenes (Fig. 1) were caused by in-frame stop codons (Fig. 1a) or frame-shift mutations (Fig. 1b). The single base transition of C/T in CAA or CAG (glutamine, Gln) in *Ta-1, 9, 23, 39* and *40* led to in-frame stop codons TAA or TAG at amino acid residues 56, 75, 129, 53 and 102, respectively, in the repetitive domain (Fig. 1a). The pseudogene *Ta-15* was caused by an in-frame stop codon, a single base transverse T/G in TTA at amino acid residue 44. The single base transverse of A/T in AAG (lysine, Lys) in *Ta-2* led to the in-frame stop codon TAG at residue 243 in the glutamine-rich domain (Fig. 1a). Five genes (*Ta-6, 8, 12, 14* and *34*) had frame-shift

mutations within their ORFs (Fig. 1b) because of single base G or A insertions upstream the binding site of primer PR4 (Fig. 1c).

Phylogenetic analysis of LMW glutenin genes from *Taeniatherum* spp

LMW glutenin genes from *Taeniatherum* spp. and four diploid relative species of wheat formed two separating branches (Fig. 2). The genes from the A genome aggregated in one branch, whereas the remaining genes formed a parallel branch. *Taeniatherum* genes were dispersed in the two subclades and all the *Taeniatherum* LMW genes except *Ta-15* aggregated in the branch formed by four genes from *Ae. tauschii* and three genes from *Ae. longissima*. However, *Ta-15* clustered with one *Ae. tauschii* gene and six genes from *Ae. longissima*. The results suggested that the LMW glutenin genes from *Taeniatherum* were more similar to those in the B and D genome diploids than the A genome diploids.

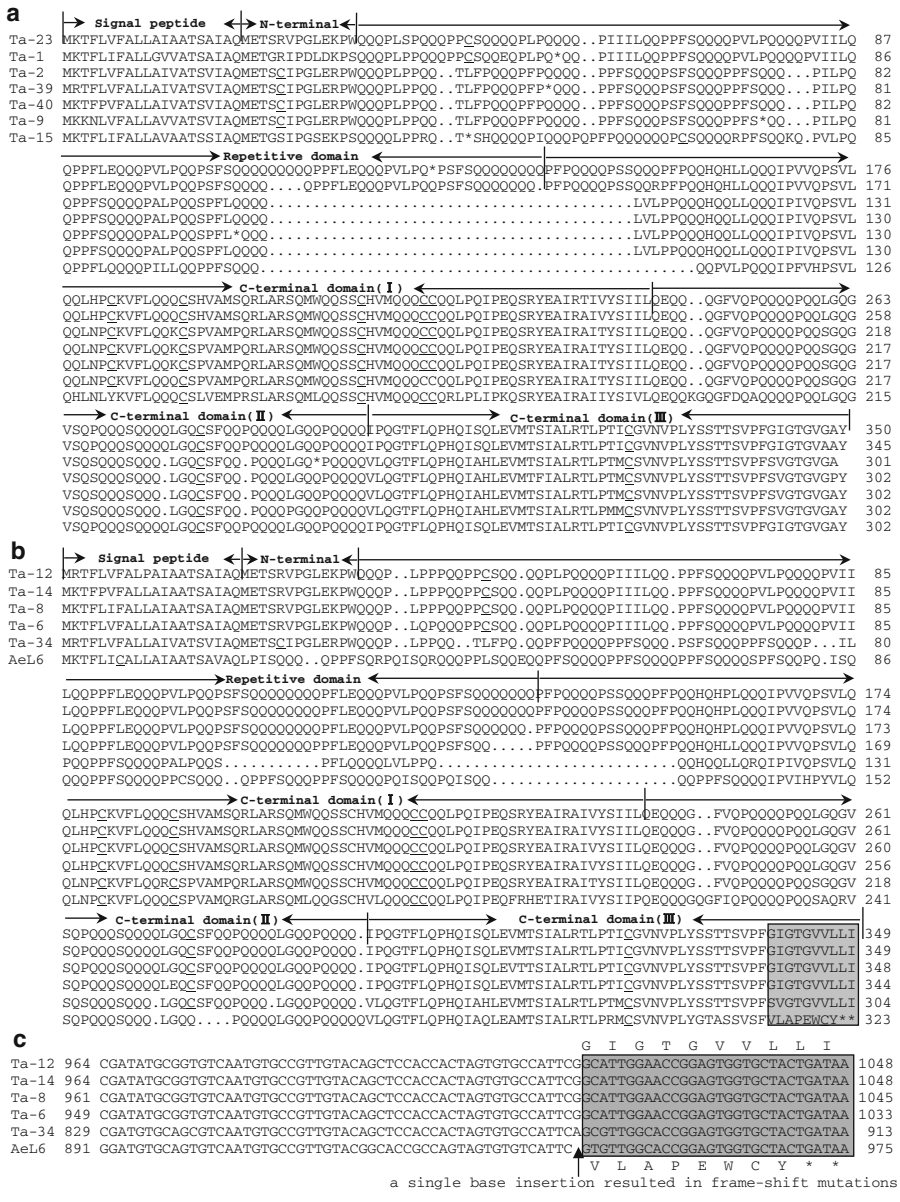

Fig. 1 Pseudogenes were caused by in-frame stop codons (**a**) and frame shift mutations (**b**) by single base insertions at the DNA level (**c**). The mutation regions are *boxed*. The GenBank accession number for AeL6 was AY724436

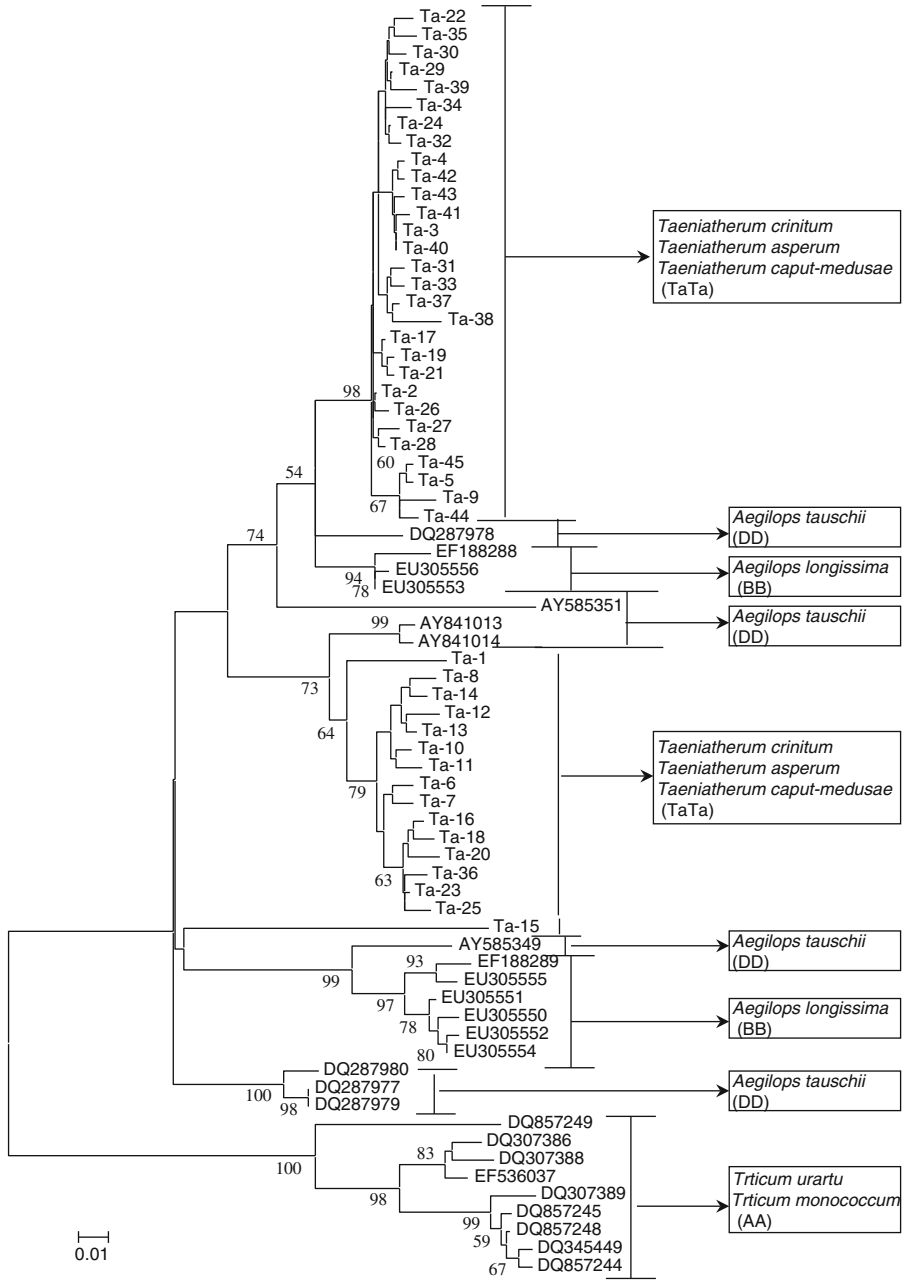

Fig. 2 Phylogenetic analysis of LMW glutenin genes of *Taeniatherum* spp. with those of orthologous genes from the three diploid donor species of bread wheat

Discussion

PCR primer design is critical for cloning homologous genes by PCR. Because researchers often select one or a few genes as reference sequences for PCR primer design, only one or a few genes in alien species may be obtained from such a strategy (D'Ovidio et al. 1997; An et al. 2006; Pei et al. 2007). In the present study, seven forward and nine reverse primers representing 63 primer combinations, were designed from conserved sequences in signal peptides or the C-terminals of 264 genes previously reported in wheat

and its wild relatives. Using these primer combinations, we produced 136 candidate LMW glutenin gene fragments in a total of 504 PCR reactions in eight accessions of *Taeniatherum*, resulting in positive PCR amplifications of ~27% of them. Of all the primer combinations, 34 produced candidate LMW glutenin DNA fragments, resulting in positive PCR amplifications of ~54%. After sequencing the DNA fragments produced by selected primer combinations (Table 3), we obtained 45 genes from *Taeniatherum*. Theoretically, these primer combinations should be capable of isolating LMW glutenin genes from other Triticeae cereals. Because not all conserved sequences in signal peptides, or the C-terminals of 264 genes, were used for primer design in the present study, more PCR primers should be designed and used for isolating LMW glutenin genes from *Taeniatherum* spp. as well as other species.

LMW glutenins belong to multigene families and gene numbers were estimated at 35–40 in bread wheat (Harberd et al. 1985; Sabelli and Shewry 1991; Cassidy et al. 1998). Variation in N-terminal, C-terminal and other regions can result in gene alteration. Based on the first amino acid residue in the N-terminal, the LMW glutenin genes are classified as LMW-m, LMW-i and LMW-s types (Cloutier et al. 2001; Lew et al. 1992). Forty four of 45 genes from *Taeniatherum* spp. were LMW-m since the first amino acid residue was methionine. However, *Ta-26* had a unique N-terminal with the first amino acid residue being valine (Val), probably resulting from a single base mutation from ATG (Met) to GTG (Val). The LMW-m genes can be further divided into METSRVP-, MDTSCIPG-, METSCIP-, MENSHIP-, METSHIPS, METSHIPG-, METRCIP-, and METSCIS- types according to the second or following residues in the N-terminal, and the genes involved were located to specific loci in hexaploid and/or tetraploid wheat (Van Campenhout et al. 1995; D' Ovidio et al. 1997; Zhang et al. 2004; Huang and Cloutier 2008). For example, the genes with N-terminals METSCIP- and METSRVP- were specific for the wheat *Glu-D3* locus (Ikeda et al. 2002; Zhao et al. 2007; Huang and Cloutier 2008). Although *Taeniatherum* is not closely related to wheat (Frederiksen 1986; Frederiksen and Bothmer 1986), 28 and 14 of the 45 LMW glutenin genes shared the same N-terminal METSCIP- and METSRVP- as in wheat. The C-terminal regions of LMW glutenins were also variable. Based on sequence differences in the C and N-terminal domains, LMW glutenin genes from the bread wheat variety Norin 61 were classified into six types and 12 groups (Ikeda et al. 2002). The 45 LMW glutenin genes from *Taeniatherum* possessed 9 different C-terminal peptides, GTGVGA*, GVAAY**, GVGAY**, GVGGY**, GVGPY**, GVGSY**, GVGVY**, GVSAY**, and PEWCY**. The alignment of nucleotide sequences suggested that the variations in N and C terminals were most probably caused by one or more single base mutations. The lengths of LMW glutenin genes are not uniform. Normally, they vary from 909 to 1,167 bp in length and range from ~32,000 to ~42,800 Da in encoded mature protein (D'Ovidio and Masci 2004). Deletion and/or insertion of repeat units in the repetitive domain are largely responsible for the length variation (D'Ovidio et al. 1999). Unequal crossing-over and/or slippage during replication can result in deletion and/or insertion of repeat units in the repetitive domain and has been suggested as one of the mechanisms for wheat prolamin evolution (Shewry et al. 1989). Allelic gene sequence comparisons suggest that deletion and/or insertion of repeat units in the repetitive domains are also responsible for new LMW glutenin genes (D'Ovidio et al. 1999). The DNA lengths of the 45 LMW glutenin genes from *Taeniatherum* spp. were likewise not the same, 30 genes with N-terminals METSCIPG, VETSCIPG and METGSIPG ranged from 909 to 915 bp, whereas the remaining 12 genes with N-terminals METSRVP and METGRIP ranged from 1,033 to 1,059 bp.

A large number of pseudogenes have been reported in cereals, including pseudogenes for high molecular weight glutenins (Forde et al. 1985), γ-gliadins (Anderson and Greene 1997) and LMW glutenins (Johal et al. 2004). Two types of LMW glutenin pseudogenes were observed in this study. Cereal prolamins are characterized by an abundance of glutamine residues. Consequently, single base transition mutations C/T at the first nucleotide position in glutamine codons (CAA or CAG) in these genes result in a high frequency of stop codons (TAA and TAG). Seven pseudogenes were attributed to in-frame stop codons (TAA or TAG) in the repetitive domain (*Ta-23*, *1*, *39*, *40*, *9* and *15*) or in the glutamine-rich C-terminal (*Ta-2*). The insertion or deletion of a single base will also result in a frame shift mutation in the triplet sets for the entire

subsequent sequence. The function of the new protein is likely to be lost because the entire protein sequence is altered and different from the original one beyond the site of mutation. Single base G (*Ta-12, 14, 8* and *6*) or A (*Ta-34*) insertions may result in frame shift mutations and function losses of the putative LMW glutenin pseudogenes in *Taeniatherum*.

It was suggested that the prolamine storage protein genes in the tribe of Triticeae have a common evolutionary origin (Shewry and Tatham 1990) and the abundant variations in LMW glutenin genes in wheat relatives represent potentially new genes for wheat end-use quality improvement (D' Ovidio and Masci 2004). Based on phylogenetic analyses and sequence alignments, the genes from *Taeniatherum* were more similar to those at the *Glu-B3* and *Glu-D3* loci rather than at *Glu-A3* (Fig. 2), suggesting that LMW glutenin gene duplication in *Taeniatherum* probably occurred after separation of the Ta, B and D genomes.

Acknowledgments This work was partially funded by the Natural National Science Foundation of China (grant no. 30671272), China New Century Excellent Talents in Universities (06-0819), the Foundation for Author of National Excellent Doctoral Dissertation of China (200458), Sichuan Youth Science and Technology Foundation (09ZQ026-088), and Education Department of Sichuan Province (09ZA077 and 09ZZ024). We thank Professor Robert A. McIntosh, University of Sydney for critical review of this manuscript.

References

An X, Zhang Q, Yan Y, Li Q, Zhang Y, Wang A, Pei Y, Tian J, Wang H, Hsam SLK, Zeller FJ (2006) Cloning and molecular characterization of three novel LMW-i glutenin subunit genes from cultivated einkorn (*Triticum monococcum* L.). Theor Appl Genet 113:383–395

Anderson OD, Greene FC (1997) The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet 95:59–65

Cassidy BG, Dvorak J, Anderson OD (1998) The wheat low-molecular-weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. Theor Appl Genet 96:743–750

Cloutier S, Rampitsch C, Penner GA, Lukow OM (2001) Cloning and expression of a LMW-i glutenin gene. J Cereal Sci 33:143–154

D'Ovidio R, Masci S (2004) Review: the low-molecular-weight glutenin subunits of wheat gluten. J Cereal Sci 39:321–339

D'Ovidio R, Simeone M, Masci S, Porceddu E (1997) Molecular characterization of a LMW-GS gene located on chromosome 1B and the development of primers specific for the *Glu-B3* complex locus in durum wheat. Theor Appl Genet 95:1119–1126

D'Ovidio R, Marchitelli C, Ercoli Cardelli L, Porceddu E (1999) Sequence similarity between allelic *Glu-B3* genes related to quality properties of durum wheat. Theor Appl Genet 98:455–461

Forde J, Malpica JM, Halford NG, Shewry PR, Anderson OD, Greene FC, Miflin BJ (1985) The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.). Nucleic Acids Res 13:6817–6832

Frederiksen S (1986) Revision of *Taeniatherum* (Poaceae). Nord J Bot 6:389–397

Frederiksen S, Bothmer RV (1986) Relationships in *Taeniatherum* (Poaceae). Can J Bot 64:2343–2347

Guo ZF, Dong P, Long XY, Wei YM, Zhang LJ, Zheng YL (2008) Molecular characterization of LMW prolamines from *Crithopsis delileana* and the comparative analysis with those from Triticeae. Hereditas 145:204–211

Gupta RB, Shepherd KW (1990) Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutelin.1. Variation and genetic control of the subunits in hexaploid wheats. Theor Appl Genet 80:65–74

Harberd NP, Bartels D, Thompson RD (1985) Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol Gen Genet 198:234–242

Hou YC, Liu Q, Long H, Wei YM, Zheng YL (2006) Characterization of low molecular-weight glutenin subunit genes from *Hordeum brevisubulatum* ssp. *turkestanicum*. Biol Bull 33:35–42

Huang XQ, Cloutier S (2008) Molecular characterization and genomic organization of low molecular weight glutenin subunit genes at the *Glu-3* loci in hexaploid wheat (*Triticum aestivum* L.). Theor Appl Genet 116:953–966

Ikeda TM, Nagamine T, Fukuoka H, Yano H (2002) Identification of new low-molecular-weight glutenin subunit genes in wheat. Theor Appl Genet 104:680–687

Jiang CX, Pei YH, Zhang YZ, Li XH, Yao DN, Yan YM, Ma WJ, Hsam SLK, Zeller FJ (2008) Molecular cloning and characterization of four novel LMW glutenin subunit genes from *Aegilops longissima*, *Triticum dicoccoides* and *T. zhukovskyi*. Hereditas 145:92–98

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

Lew EJL, Kuzmicky DD, Kasarda DD (1992) Characterization of low molecular weight glutenin subunits by reverse-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis and N-terminal amino acid sequencing. Cereal Chem 69: 508–515

Li XH, Ma WJ, Gao LY, Zhang YZ, Wang AL, Ji KM, Wang K, Appels R, Yan YM (2008) A novel chimeric low-molecular-weight glutenin subunit gene from the wild relatives of wheat *Aegilops kotschyi* and *Ae. juvenalis*: evolution at the *Glu-3* loci. Genetics 180:93–101

Luo Z, Chen FG, Feng DS, Xia GM (2005) LMW glutenin genes in *Agropyron elongatum* and their potential value in wheat breeding. Theor Appl Genet 111:272–280

Masci S, Lafiandra D, Porceddu E, Lew EJL, Tao HP, Kasarda DD (1993) D-glutenin subunits: N-terminal sequences and evidence for the presence of cysteine. Cereal Chem 70:581–585

Payne PI, Nightingale MA, Krattiger AF, Holt LM (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. J Sci Food Agr 40:51–65

Pei YH, Wang AL, An XL, Li XH, Zhang YZ, Huang XQ, Yan YM (2007) Characterization and comparative analysis of three low molecular weight glutenin C-subunit genes isolated from *Aegilops tauschii*. Can J Plant Sci 87:273–280

Pistón F, Martín A, Dorado G, Barro F (2005) Cloning and molecular characterization of B-hordeins from *Hordeum chilense* (Roem. et Schult.). Theor Appl Genet 111: 551–560

Sabelli PA, Shewry PR (1991) Characterization and organization of gene families at the *Gli-1* loci of bread and durum wheats by restriction fragment analysis. Theor Appl Genet 83:209–216

Shang HY, Wei YM, Long H, Yan ZH, Zheng YL (2005) Identification of LMW glutenin-like genes from *Secale sylvestre* Host. Russ J Genet 41:1372–1380

Shewry PR, Tatham AS (1990) The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J 267:1–12

Shewry PR, Halford NG, Tatham AS (1989) The high molecular weight subunits of wheat, barley and rye. In: Miflin BJ (ed) Genetics, molecular biology, chemistry and role in wheat gluten structure and functionality. Oxford surveys of plant molecular and cellular biology, vol 6. University Press, New York, pp 163–219

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Van Campenhout S, Vander Stappen J, 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

Yan ZH, Wan YF, Liu KF, Zheng YL, Wang DW (2002) Identification of a novel HMW glutenin subunit and comparison of its amino acid sequence with those of homologous subunits. Chinese Science Bull 47:220–225

Zhang W, Gianibelli MC, Rampling LR, Gale KR (2004) Characterization 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 XL, Xia XC, He ZH, Lei ZS, Appels R, Yang Y, Sun QX, Ma W (2007) Novel DNA variations to characterize low molecular weight glutenin *Glu-D3* genes and develop STS markers in common wheat. Theor Appl Genet 114:451–460