Theor Appl Genet (2011) 122:1411–1425
DOI 10.1007/s00122-011-1541-8

ORIGINAL PAPER

Phylogenetic relationship of a new class of LMW-GS genes in the M genome of *Aegilops comosa*

Ke Wang · Liyan Gao · Shunli Wang · Yanzhen Zhang · Xiaohui Li · Miaoyi Zhang · Zhenze Xie · Yueming Yan · Matthew Belgard · Wujun Ma

Received: 27 October 2010 / Accepted: 14 January 2011 / Published online: 8 February 2011

Abstract A new class of low molecular weight glutenin subunit (LMW-GS) genes was isolated and characterized from *Aegilops comosa* ($2n = 2x = 14$, MM). Although their DNA structure displayed high similarity to LMW-i type genes, there are some key differences. The deduced amino acid sequences of their mature proteins showed that the first amino acid residue of each gene was leucine and therefore they were designated as LMW-l type subunits. An extra cysteine residue was present in the signal peptide and the first cysteine residue of mature proteins located at the end of repetitive domain. Additionally, a long insertion of 10–22 residues ($LGQQPQ_{5-17}$) occurred in the end of the C-terminal II. Comparative analysis demonstrated that LMW-l type glutenin genes possessed a great number of single-nucleotide polymorphisms and insertions/deletions. A new classification system was proposed according to the gene structure and phylogenetic analysis. In this new system, LMW-GS is classified into two major classes, LMW-M and LMW-I, with each including two subclasses. The former included LMW-m and LMW-s types while the latter contained LMW-l and LMW-i types. Analysis of their evolutionary origin showed that the LMW-l genes diverged from the group 2 of LMW-m type genes at about 12–14 million years ago (MYA) while LMW-i type evolved from LMW-l type at approximately 8–12 MYA. The LMW-s type was a variant form of group 1 of LMW-m type and their divergence occurred about 4–6 MYA. In addition to homologous recombination, non-homologous illegitimate recombination could be an important molecular mechanism for the origin and evolution of LMW-GS gene family. The secondary structure prediction suggested that the novel LMW-l type subunits, such as AcLMW-L1 and AcLMW-L2, may have positive effects on dough properties.

Communicated by B. Friebe.

K. Wang and L. Gao contributed equally to this work.

K. Wang · L. Gao · S. Wang · Y. Zhang · X. Li · M. Zhang · Z. Xie · Y. Yan (✉)
Key Laboratory of Genetics and Biotechnology, College of Life Science, Capital Normal University, 100048 Beijing, China
e-mail: yanym@hotmail.com

W. Ma (✉)
Western Australian Department of Agriculture and Food, State Agriculture Biotechnology Centre, Murdoch University, Perth, WA 6150, Australia
e-mail: w.ma@murdoch.edu.au

Y. Zhang
College of Applied Sciences and Humanities of Beijing Union University, 100083 Beijing, China

M. Belgard · W. Ma
Centre for Comparative Genomics, Murdoch University, Perth, WA 6150, Australia

Introduction

Seed storage proteins, mainly consisting of polymeric glutenins and monomeric gliadins, play a key role in determining wheat dough quality. The polymeric glutenins can be subdivided into high molecular weight glutenin subunits (HMW-GS) and low molecular weight glutenin subunits (LMW-GS) according to their mobilities on SDS–PAGE gel. It is generally accepted that HMW-GSs are important determinants of wheat dough elasticity while LMW-GSs are significantly correlated with dough extensibility and strength (Cornish et al. 2001; Maruyama-Funatsuki et al. 2004; Tanaka et al. 2005; Ma et al. 2005).

So far, HMW-GSs have been extensively studied by means of gene cloning, molecular genetics, genetic transformation, etc. (Shewry and Halford 2002).

There is significant variation in both the gene sequences and genomic organisation of the gene region that encodes LMW-GS among bread wheats and related species. Such complexity has made it difficult to accurately characterize this region from the evolutionary point of view (D'Ovidio and Masci 2004; Huang and Cloutier 2008). To date, it is well documented that the LMW-GSs can be separated into three groups: B, C and D based on their electrophoretic mobilities on SDS–PAGE gel. They are encoded by *Glu-A3*, *Glu-B3* and *Glu-D3* loci on the short arms of chromosomes 1A, 1B and 1D, respectively (Jackson et al. 1983; Yan et al. 1999). Based on their N-terminal amino acid sequences, LMW-GSs are traditionally classified into three subclasses: LMW-m, LMW-s and LMW-i types, which corresponding to the first amino acid residue of the mature protein, methionine, serine, and isoleucine, respectively (D'Ovidio and Masci 2004). Typically, the N-terminal amino acid sequence of LMW-s type subunits is *SHIPGL*, while these for the LMW-m type subunits can vary, including *METSHIGPL-*, *METSRIPGL-* and *METSCIPGL-* (Kasarda et al. 1988; Tao and Kasarda 1989; Lew et al. 1992; Masci et al. 1995). The LMW-i type, first reported by Pitts et al. (1988), lacks the N-terminal domain and starts directly with the repetitive region after the signal sequence, namely *ISQQQQ-* which is the deduced N-terminal sequence for this subclass. Although the typical N-terminal domain was absent, LMW-i type subunits can express in wheat endosperm (Cloutier et al. 2001; Ikeda et al. 2002). In addition to this classification system, Ikeda et al. (2002) classified the LMW-GSs into six types on the basis of the locations of cysteine residues, as the cysteine residues play an important role in the formation of intra- and intermolecular disulphide bonds in the gluten macropolymer. Despite these discoveries, little is known about the genomic organization and evolutionary mechanisms of the *Glu-3* loci (Huang and Cloutier 2008).

Aegilops genus is closely related to *Triticum*. The progenitor of the *Triticum* D genome has been confirmed to be *Aegilops tauschii* (Yan et al. 2003, 2004). The progenitor of B is *Aegilops speltoides* in all polyploid wheats (Dvorak 1998). Of 13 diploid species of *Triticum* and *Aegilops* genera, there are eight distinct genomes: A, D, S, M, C, U, N and T. There are extensive allelic variations of storage proteins in *Triticum* and *Aegilops* species. This variation can provide not only potential gene candidates for wheat quality improvement, but also much valuable information to elucidate the origin and evolution of wheat and related species. As highlighted by Huang et al. (2002), the diploid progenitors and origin of genomes such as B and G and the evolutionary relationships of different genomes among *Triticum* and *Aegilops* species still need to be further investigated.

Aegilops comosa ($2n = 2x = 14$, MM) was considered to be the progenitor of the M genome (Kimber and Feldman 1987). It is known that the M genome widely existed in *Aegilops* genus and possessed significant candidate genes such as resistance genes against diverse pests (Riley et al. 1968). However, the molecular characteristics and genomic organization of LMW-GS genes in *Aegilops comosa* have not been studied to date. This study reports ten LMW-GS genes isolated from different accessions of *Aegilops comosa*. The origin and phylogenetic relationships among M and other related genomes of *Triticum* and *Aegilops* species have been investigated. Through this analysis, a novel subclass of LMW-GS is revealed and its likely significance is discussed.

Materials and methods

Plant materials

Two accessions of *Aegilops comosa* ($2n = 2x = 14$, MM) numbered PI 551017 and PI 551019 (originated from Greece) were used that were kindly provided by CIMMYT.

DNA extraction and PCR amplification

Total genomic DNA of dry seeds was extracted following the method outlined in An et al. (2006). A pair of AS-PCR primers (LMW-1 and LMW-2) was designed to amplify the upstream coding regions and downstream regions based on previously cloned LMW glutenin gene sequences (Cassidy et al. 1998; Cloutier et al. 2001). The sequences of primers were LMW1 5′-ATCATCACAAGCACAAGCATC-3′ and LMW2 5′-TTCTTATCAGTAGGCACCAAC-3′ (synthesized by Sangong), PCR amplifications were performed in a total volume of 50 µl containing 2.5 U LA Taq polymerase (TaKaRa), 60 ng of template DNA, 25 µl 2× GC buffer I ($MgCl^{2+}$ plus), 0.4 mM dNTP, 0.5 µM of each primer, and double distilled H_2O added to 50 µl. The reaction was carried out in a PTC-100 (MJ Research) according to the following protocol: heat lid turned on, initial denaturation at 94°C for 4 min, cycled 35 times at 94°C for 1 min, 56°C for 1 min and 72°C for 1 min, and a final extension at 72°C for 10 min. PCR products were analyzed by agarose electrophoresis with 1% gel in Tris–acetic acid–EDTA buffer.

Molecular cloning and sequencing

The PCR fragments of expected size were purified from the gel using the Gel Extraction Kit (Omega). Purified products were then ligated into pGEM-T Easy vector (Tiangen,

Beijing, China) and transformed into cells of *Escherchia coli* TOP 10. DNA sequencing from three clones of each PCR fragment was carried out by Sunbiotech.

Expression of the cloned LMW-GS genes in *E. coli*

The *AcLMW-L2* and *AcLMW-L3* genes cloned were re-amplified to remove the signal peptides by designing a new pair of primers Lbd-LF3-2 (5′-AAGCCATGGTACC AATC/TTCAC-3′) and Lbd-LR (5′-AAACTCGAGGT AGGCACCAACTC-3′). *NcoI* and *XhoI* sites (underlined) were incorporated into the 5′ ends of the Lbd-LF3-2 and Lbd-LR, respectively. After purification, the PCR products were ligated into the expression vector pET-28a (Novagen), and transformed into *E. coli* BL21 (DE3) plysS cells. The expressed protein extraction and separation were carried out by SDS–PAGE according to Li et al. (2007).

MALDI-TOF–MS

The heterologous expressing proteins and their corresponding subunits in the seed endosperm were detected by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF–MS) according to previous reported methods (An et al. 2006; Pei et al. 2007). Shimadzu corporation AXIMA-CFRTM Plus MS apparatus (Japan) and the matrix of sinapinic acid (SA, α-cyano-4-hydroxycinnamic acid) were used. International two-dot calibration with standard sample Albumin-Aldrase at masses of 39212.88 and 66431.08 Da was used.

SNPs and InDels identification and secondary structure prediction

Identification of single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) present in LMW glutenin genes was based on multiple alignments and performed using Bioedit 7.0. Prediction of secondary structure of deduced amino acid sequences was carried out by PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html; McGuffin et al. 2000; Bryson et al. 2005; Li et al. 2007).

Phylogenetic analysis

Phylogenetic tree was constructed with software DNAMAN5.2.2 using the complete coding regions. Neighbor joining with Kimura two parameter correction method and bootstrapping of 1,000 replicates were selected as working parameters. The divergent time of different genes or genomes were calculated by MEGA3 (Gaut et al. 1996; Kumar et al. 2004; Zhang et al. 2006; An et al. 2006).

Results

Isolation and molecular characterization of a new type of LMW-GS genes

In this study, a total of 10 complete coding genes of LMW-GS from *Aegilops comosa* were cloned and designated as *AcLMW-L1*, *AcLMW-L2*, *AcLMW-L3*, *AcLMW-L4*, *AcLMW-I1*, *AcLMW-I2*, *AcLMW-I3*, *AcLMW-M1*, *AcLMW-M2*, *AcLMW-M3* based on sequencing results (Fig. 1). Since three genes, *AcLMW-L1*, *AcLMW-I2* and *AcLMW-M3* had higher sequence identity to *AcLMW-L2*, *AcLMW-I3* and *AcLMW-M2*, respectively, 7 out of the 10 genes were used for comparative analysis and their structural characterizations were showed in Fig. 1. The deduced amino acid sequences of all aligned genes shared four main domains including a signal peptide of 20 amino acid residues, a short N-terminal (I), a repetitive domain (II) and a C-terminal region. As suggested by Cassidy et al. (1998), the C-terminal domain could be further divided into three regions: the cysteine-rich domain (C-terminal I), glutamine-rich region (C-terminal II) and highly conserved domain (C-terminal III).

Of the 10 LMW-GS, AcLMW-M2 and AcLMW-M3 had the same N-terminal sequence of *METSCIPG*, and therefore belonged to typical LMW-m type subunits. Their positions of eight cysteine residues were as expected and conserved as LMW-m subunits. AcLMW-I2 and AcLMW-I3 lack the typical N-terminus and it appears that the first amino acid residue is isoleucine which is typical for mature protein of the LMW-i type subunits. Interestingly, there is another set of novel genes including four genes sequenced from this study (*AcLMW-L1*, *AcLMW-L2*, *AcLMW-L3*, *AcLMW-L4*). Unlike the known subtypes of LMW proteins which are characterized by either met, iso, or ser start amino acid, these aligned proteins all commence with a leucine. They also possess different N-terminal sequences. The other typical characters of this new type gene included an extra cysteine residue located at position 7 in the signal peptide and the first cysteine residue in the repetitive region presented in the end of domain II. Both cysteine residues resulted from a point mutation, namely T → G transversion, and therefore generated cysteine (TGT) from phenylalanine (TTT). In addition, a long insertion of 10–22 residues (consensus $LGQQPQ_{5-17}$) was present in the end of the C-terminal II as shown in Fig. 1. This appears to be a novel LMW subunit type.

According to these characteristics showed above, we designated them as a new LMW-l type subunit (Table 1). The consensus sequence in the beginning of N-terminal of the LMW-l subunits was *LPIQQQQ* and had higher similarity with *ISQQQQQ* of LMW-i type subunit encoded by the *Glu-A3* locus, which also lacks N-terminal and

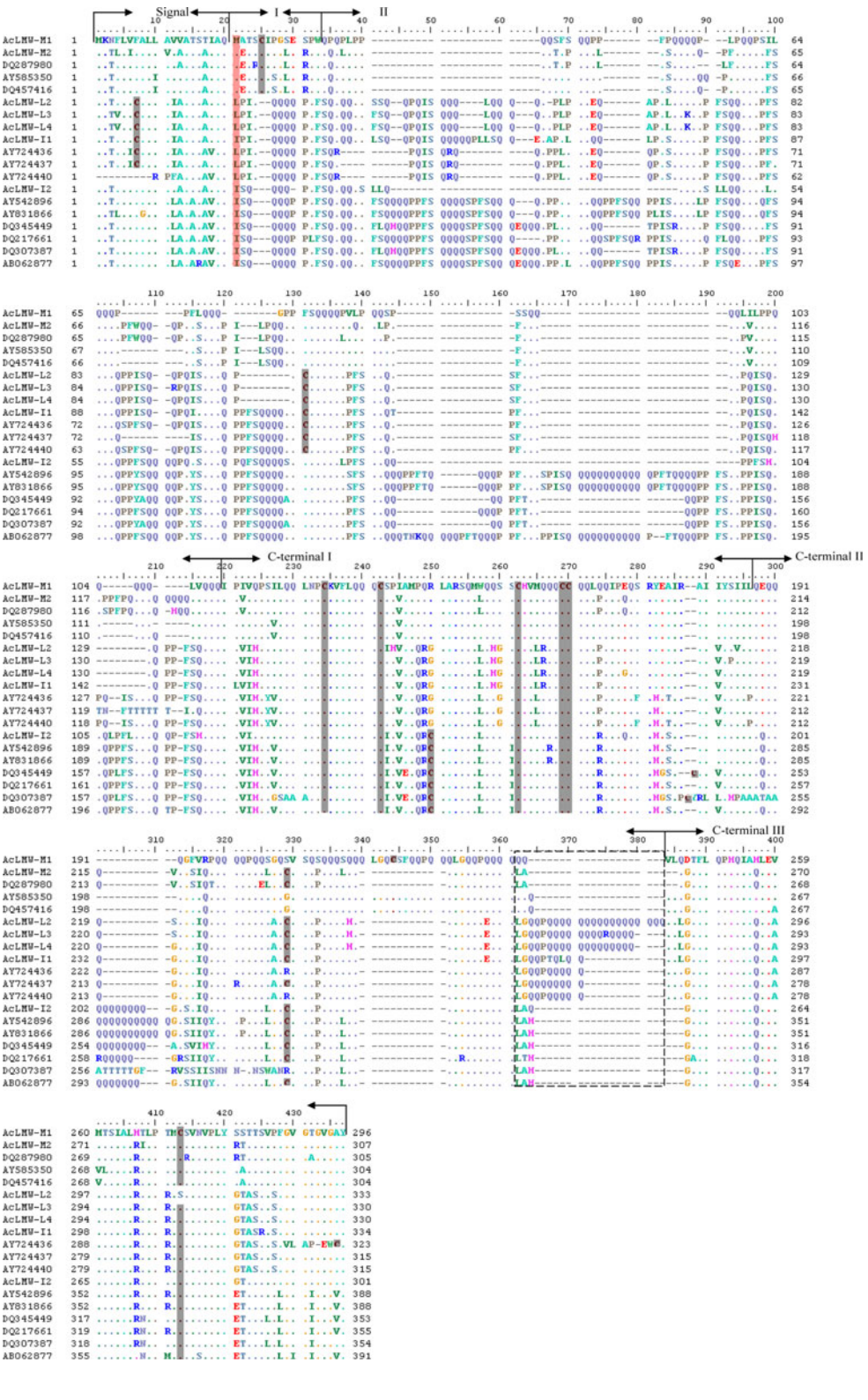

Fig. 1 Multiple alignment of the deduced amino acid sequences of 19 LMW glutenin genes including LMW-i, LMW-m and LMW-l type subunits. Signal represents signal peptide, N-terminal domain (I), repetitive domain (II) and three sub-regions of C-terminal domain were indicated, respectively. The first amino acid residue of the mature proteins and cysteine residues were highlighted by *red* and *gray* shading, respectively. The *box* indicated a long insertion. Deletions were indicated by *dots* and *dashes*

begins directly with repetitive domain after signal peptide. This suggests that M and A genomes could have close evolutionary relationships. Our results showed that two kinds of LMW-GS genes, namely LMW-m and LMW-l type in PI 551017 and LMW-i and LMW-l type in PI 551019 were isolated, suggesting that *Glu-M3* locus encoded at least two different subclasses of LMW-GS in an accession.

Table 1 Comparison of the mature protein sequences of four types of LMW-GS

Subunit types	GenBank accession	Amino acid number	N-terminal sequence	Number of cysteine residues							
				Signal	N-terminal domain	Repetitive domain	C-ter domain I	C-ter domain II	Insert amino acid number	C-ter domain III	Total
LMW-l	EU594331	334	LPISQQQQ	1	0	1	5	1	21	1	9
LMW-l	EU594332	333	LPISQQQQ	1	0	1	5	1	20	0	8
LMW-l	EU594333	330	LPISQQQQ	1	0	1	5	1	16	1	9
LMW-l	EU594334	330	LPISQQQQ	1	0	1	5	1	16	1	9
LMW-l	AY724437	315	LPISQQQQ	1	0	1	5	1	8	1	9
LMW-i	EU594336	301	ISQQQQPP	0	0	0	6	1	0	1	8
LMW-i	EU594337	301	ISQQQQPP	0	0	0	6	1	0	1	8
LMW-i	AY542896	388	ISQQQQPP	0	0	0	6	1	0	1	8
LMW-m	EU594338	296	MATSCIPG	0	1	0	5	1	0	1	8
LMW-m	EU594339	307	METSCIPG	0	1	0	5	1	0	1	8
LMW-m	EU594340	308	METSCIPG	0	1	0	5	1	0	1	8
LMW-m	AY287980	305	METRCISG	0	1	0	5	1	0	1	8
LMW-s	AB164415	392	MENSHIPG	0	0	1	5	1	0	1	8
LMW-s	DQ357057	354	MENSHIPG	0	0	1	5	1	0	1	8
LMW-s	EU369708	364	MENSHIPG	0	0	1	5	1	0	1	8

SNPs and InDels analysis of LMW-l genes

The coding sequences of LMW-l type subunits were compared with 15 LMW-i type subunit genes reported previously or deposited in GenBank, and variations in SNPs and InDels were identified (Table 2). A total of 51 SNPs were detected at different positions. The numbers of SNPs in signal peptide, N-terminal (I), repetitive domain (II) and C-terminal region were 2, 5, 4 and 40, respectively. Moreover, there were 16, 19 and 5 SNPs in the C-terminal I, II and III, respectively. Two-, one- and three-base insertions were present at the positions 131–132, 139 and 915–919, and only a three-base deletion at 71–73 was found in LMW-l genes.

Expression and identification of LMW-l type genes in *E. coli*

The coding regions (without signal peptides) of two typical LMW-l genes *AcLMW-L2* and *AcLMW-L3* were expressed in *E. coli*. Because his-tag encoding sequences were incorporated into the end of the expressed genes, the hybrid protein was about 1,083 Da larger than the former proteins themselves. The expressed fusion proteins were identified by both SDS–PAGE and MALDI-TOF–MS. As shown in Fig. 2a, lanes 1 and 4 had an extra protein band comparing the negative lane 2. Moreover, the accurate M_r of two fusion proteins determined by MALDI-TOF–MS were 36,465 and 36,416 Da (Fig. 2b), which was consistent to the deduced 36,765 and 36,471 Da (including 1,083 Da of His-tag) of hybrid AcLMW-L2 and AcLMW-L3 subunits, respectively. Thus, the extra bands in lanes 1 and 4 are likely the fusion products of *AcLMW-L2* and *AcLMW-L3* genes and confirmed that the LMW-l type genes without N-terminal could express normally as LMW-i type genes (Cloutier et al. 2001).

Prediction of secondary structure

Protein secondary structure is the base of high complex spatial conformation. In this study, PSIPRED method was used to predict secondary structures of the new LMW-l type subunits and the results were showed in Table 3. Previous reports indicated that β-strand was generally considered to confer the protein with high elasticity and to improve the capability to resist distortion, indicating that the content of β-strands in LMW-GS may have positive effects on dough quality (Tatham et al. 1985, 1987, 1990; Jiang et al. 2008). Comparing to the reported XYGluD3-LMWGS1 subunit (AY263369), which was considered to have a positive effect on dough quality (Zhao et al. 2004), AcLMW-L1 had 4 β-strands, being twice of that of AY263369. The percent of β-strand in AcLMW-L1 was also the highest among all the predicted LMW glutenin subunits, slightly higher than AlLMW-m2 reported by Jiang et al. (2008). AcLMW-L2 ranked the second, slightly exceeding AY263369. These results implied that AcLMW-L1 and AcLMW-L2 were likely to have positive effects on dough properties. As far as the dispersal of every secondary motif, the results showed that α-helix and β-strand

Table 2 The positions of SNPs and InDels identified between LMW-l and LMW-i type genes

	20	34	61	63	64	68	71–73	75	123	131–132	139	315	320	430	608	610	611	613	615
LMW-GS																			
LMW-l	G	A	C	A	C	T	–	G	A	AA	C	A	C	G	C	C	A	A	T
Fourteen other LMW-i Genes	T	C/G/T	A	G/T	G/T	A/C	TCA	A/–	G	–	–	C/G/–	A/–	C/T/–	T	G/T	C/T	C/T	A/G
LMW-GS	616	618	736	764	794	809	826	833	837	847	915–919	923	926	927	1000	1001	1005	1017	
LMW-l	T	A	A	G	G	T	T	C	C	T	AGTC	T	C	A	C	C	T/C	C	
Fourteen other LMW-i Genes	C	G/T	G	C/T	A	A	C	T	A/G	C	–	C/–	T/–	G/–	G/T	T/–	A/G	G/T	
LMW-GS	1,046	1,050	1,051	1,052	1,053	1,057	1,058	1,066	1,178	1,099	1,100	1,154	1,186	1,208	1,254	1,258	1,266		
LMW-l	G	C	A	A	C	A	A	A/–	A	A/–	A/–	C	C	A	G	G	T		
Fourteen other LMW-i Genes	A/C	T	C/G/–	G/T	T	T	C	C	C/G	C/T	C/T	T	T	T	A	C	C		

Horizontal dashes indicated the deletions of nucleotide. Other 14 LMW-i genes included AB062877 (Ikeda et al. 2002), X07747 (Pitts et al. 1988), AY585350 (Johal et al. 2004), AY542896 (Cloutier et al. 2001), AY831866 (Ozdemir et al. 2005), AY453154, AJ293097, AY217661, AY217662, DQ307387, DQ307388, DQ307389, DQ345449 and DQ857249

dispersed usually in the C-terminal I and highly conserved domain (C-terminal III). AcLMW-L1 had an extra β-strand in C-terminal II while AcLMW-L2 had an extra α-helix in repetitive domain (II).

Phylogenetic analysis of LMW-l type genes

To investigate the phylogenetic relationships among LMW-GS genes from different genomes, the complete coding sequences of 34 genes were used to construct a phylogenetic tree, which includes the 10 new LMW-GS genes obtained in this study and other 24 genes from GenBank, namely AB062872 (Ikeda et al. 2002), AY263369 (Zhao et al. 2004), AY585350 (Johal et al. 2004), AY542896 (Cloutier et al. 2001), AY453157 and AY453160 (Zhang et al. 2004), AY606257, AY646285, AY724436, AY724437 and AY724440 (Luo et al. 2005), DQ357052, DQ357057 and DQ357058 (Zhao et al. 2006), DQ457416, AJ293099, DQ287977, AY841014, AB164415, AB164416, DQ217663,

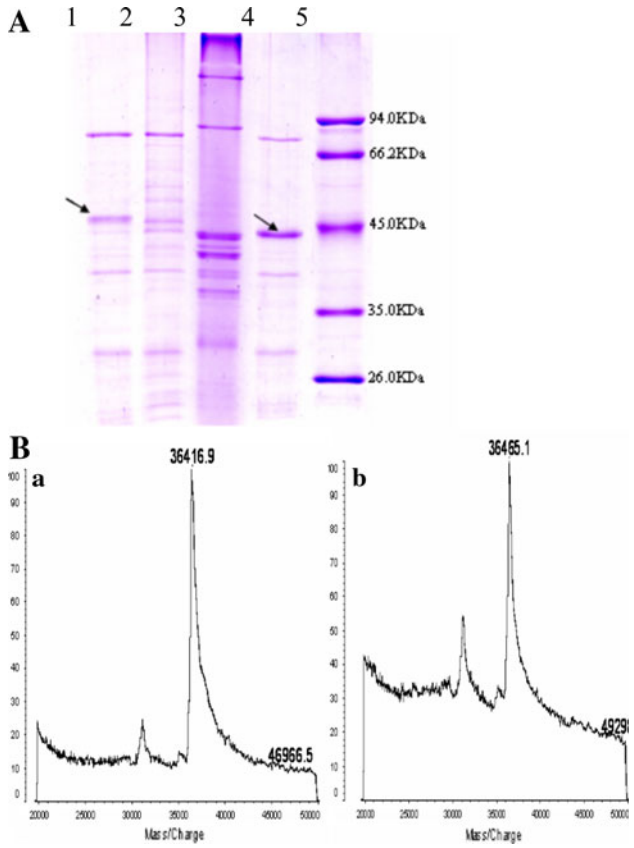

Fig. 2 Expression and identification of two LMW-l type genes in *E. coli*. **a** SDS analysis of the expressed proteins in *E. coli*. The expressed proteins of *AcLMW-L2* (*Lane 1*), 28a vector (*Lane 2*), the glutenin subunit of the seed of PI551019 (*Lane 3*), the expressed protein of *AcLMW-L3* (*Lane 4*) and the protein marker (*Lane 5*) were showed. **b** Determination of M_rs of the expressed fusion proteins of *AcLMW-L2* (*a*), *AcLMW-L3* (*b*) by MALDI-TOF–MS

Table 3 The secondary structure prediction of the nine deduced LMW-GSs

LMW-GS	Structure motifs	Content %	Total	Dispersal in every region				
				N-terminal domain	Repetitive domain	C-ter domain I	C-ter domain II	C-ter domain III
AcLMW-L1	α-helix	7.64	2	–	–	1	–	1
	β-strand	2.87	4	–	–	2	1	1
AcLMW-L2	α-helix	1.92	3	–	1	1	–	1
	β-strand	1.28	2	–	–	1	–	1
AcLMW-I1	α-helix	6.69	3	–	–	2	–	1
	β-strand	1.59	2	–	–	1	–	1
AcLMW-L3	α-helix	10.97	4	–	–	2	–	2
	β-strand	1.29	2	–	–	1	–	1
AY724436	α-helix	–	–	–	–	–	–	–
	β-strand	1.32	2	–	–	–	–	2
AY724437	α-helix	10.51	3	–	–	2	–	1
	β-strand	0.68	1	–	–	–	–	1
AY263369	α-helix	11.6	3	–	–	2	–	1
	β-strand	1.4	2	–	–	1	–	1
AY831866	α-helix	5.98	3	–	–	2	–	1
	β-strand	1.09	2	–	–	–	1	1
AY542896	α-helix	2.72	1	–	–	–	–	1
	β-strand	1.09	2	–	–	1	–	1

DQ217661, DQ307387 and DQ345449. The results were showed in Fig. 3.

The phylogenetic tree clustered into two clear branches with 75% homology. Among each cluster, two subgroups were apparently separated, namely LMW-m and LMW-s, LMW-l and LMW-i, with identities of 79 and 77%, respectively. Previous studied indicated that the LMW-i genes showed greater divergences to LMW-s and LMW-m genes (An et al. 2006; Jiang et al. 2008). In this study, the new LMW-l type genes appeared to be more homologous to LMW-i than to LMW-m and LMW-s type genes.

In order to further investigate the evolutionary relationships among LMW-GS genes, the neighbor-joining tree was constructed based on complete sequences of 28 LMW-GS genes, and their divergent times were calculated with software MEG3.0. According the previous reports (Gaut et al. 1996; Kumar et al. 2004; An et al. 2006; Zhang et al. 2006, 2008; Li et al. 2007), the evolutionary rate of 6.5×10^{-9} was used. As shown in Fig. 4, it was discovered that LMW-l type genes were closer related to LMW-i than LMW-m and LMW-s genes, which was agreement with the result of homology tree. As shown in Table 4, LMW-m type genes could be divided into two subgroups: group 1 (*AcLMW-m2*, *AY263369* and *DQ287977*) and group 2 (*AcLMW-m1*, *AY841014* and *AJ293099*). Group 1 was more related to LMW-s-type than group 2, and their divergent time was estimated about 4–6 million years ago (MYA). Group 2 was more similar with LMW-l type than group 1, and their divergence occurred about 12–14 MYA. Furthermore, the divergent time of LMW-l type and LMW-i type was about 8–12 MYA. These results were generally in agreement with those by the neighbor-joining tree analysis (Fig. 4).

Alignments among glutenin loci of different genomes MYA provide useful information about the origin and divergent times of chromosome sets and the evolution of wild sources and modern cultivars. Thus, the phylogenesis of different genomes was analyzed based on 26 LMW-GS genes which were located on specific genomes, and the divergent times were calculated. The neighbor-joining tree (Fig. 5) demonstrated that M and A genomes were estimated to diverge at about 7.51 MYA while M and D genomes diverged much later, at about 2–3 MYA, suggesting that the divergence of the *Glu-3* loci was much earlier than genomes.

Discussion

In this study, we isolated, for the first time, 10 LMW-GS genes from M genome of *Aegilops comosa* and a new LMW-l type was characterized. Our results confirmed that there existed *Glu-M3* locus on the M genome and its molecular characteristics could provide new evidence for further insight into the origin and evolution of LMW-GS gene family.

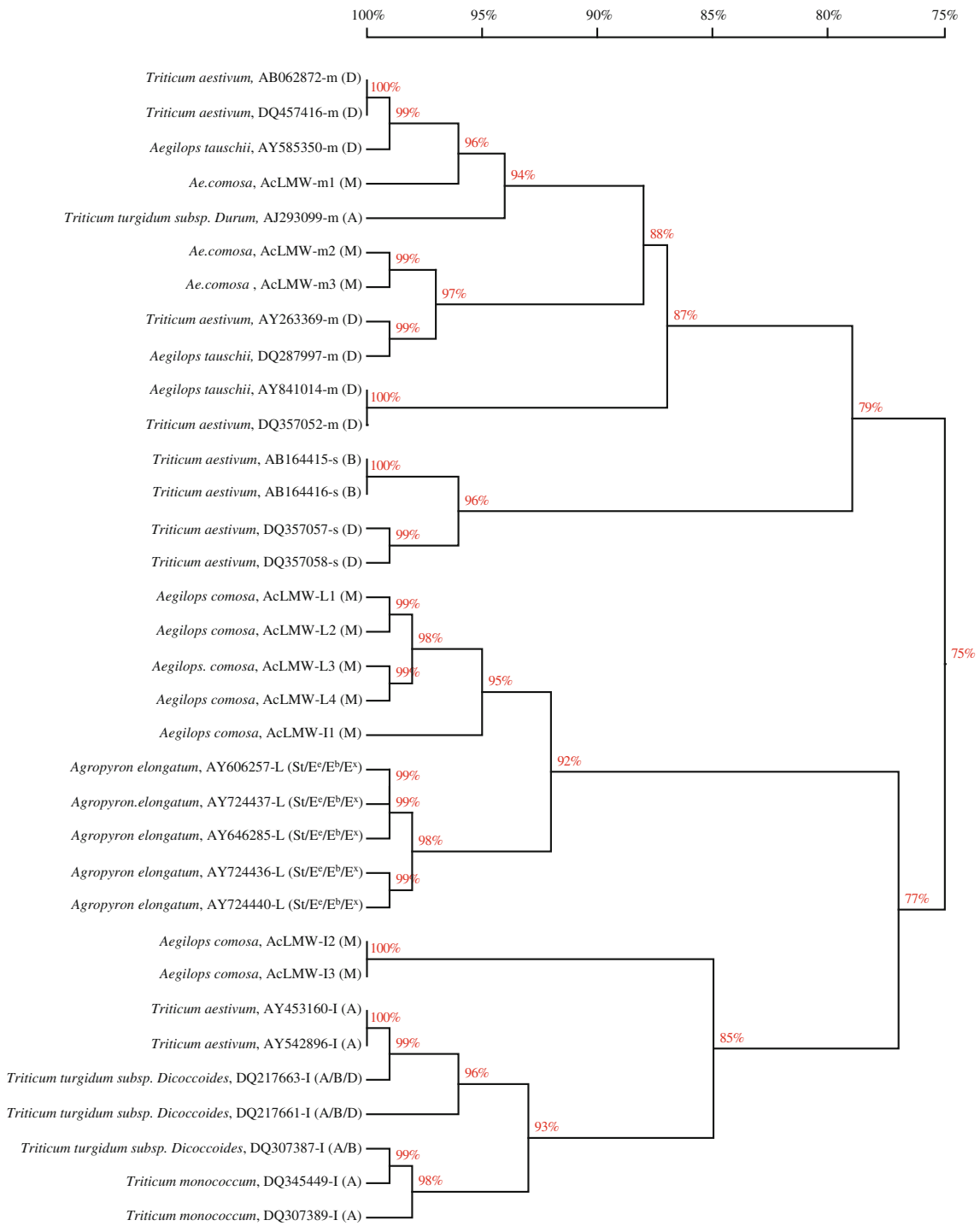

Fig. 3 Homology tree based on the coding regions of 34 LMW-GS genes. The suffixes of GenBank accession numbers indicated the different types of the genes

Allelic variations and genomic organization of LMW-GS genes at *Glu-3* loci

It is known that the *Glu-3* loci encoded LMW-GS are widely presented in different genomes of common wheat and related cereals (D'Ovidio and Masci 2004). The extensive allelic variations at the *Glu-3* loci were considered to be responsible for the difference in quality properties of wheat dough. The alleles at the *Glu-3* loci were first named by Gupta and Shepherd (1990) when 20 alleles were identified. Nieto-Taladriz et al. (1997) proposed a nomenclature for 19 alleles detected in durum wheat. The

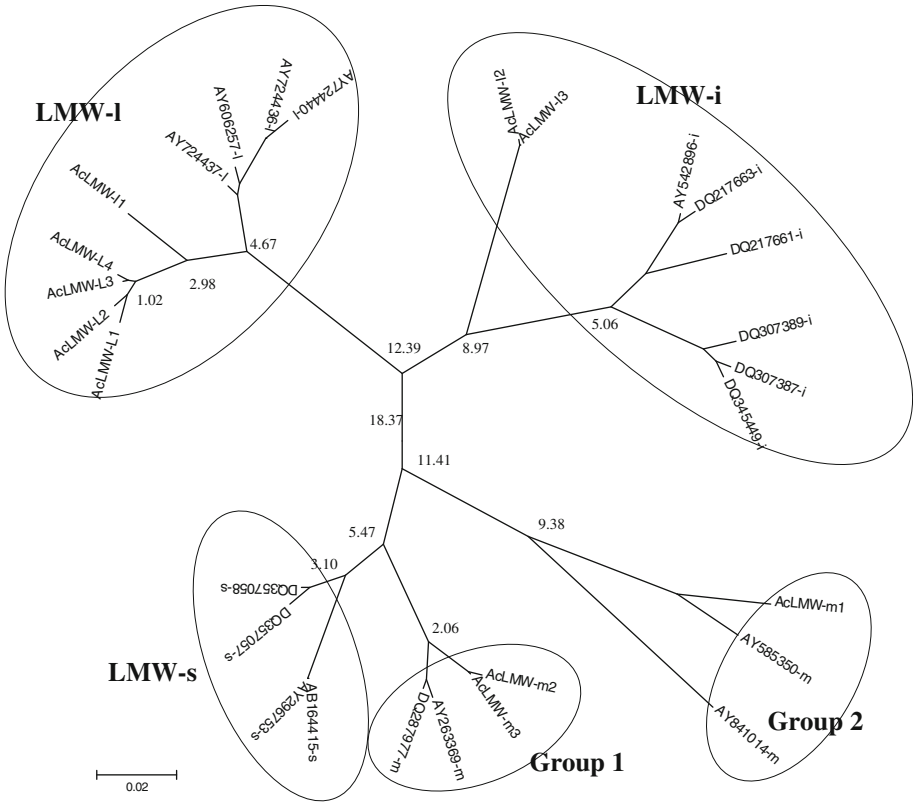

Fig. 4 Neighbor-joining tree based on the complete sequences of 28 LMW-GS genes including four types of LMW-GS genes. The suffixes of GenBank accession numbers indicated the different types of LMW-GS genes. The divergent times were specified by the numbers, and the unit was million years ago (MYA)

copy numbers of the LMW-GS genes were estimated to have about 10–15 (Harberd et al. 1985) and 30–40 (Sabelli and Shewry 1991; Cassidy et al. 1998; Huang and Cloutier 2008) in hexaploid wheat. The physical distance between two adjacent LMW-i-type genes from the A^m genome of *Triticum monococcum* was estimated to be more than 150 kb (Wicker et al. 2003). More recently, Gao et al. (2007) reported that two LMW-GS genes were separated from each other by approximately 100 kb. Huang and Cloutier (2008) isolated 12 unique LMW glutenin genes and 7 pseudogenes from the hexaploid wheat 'Glenlea' BAC library, and the average physical distance between two adjacent LMW-GS genes was 81 kb according to the LMW-GS gene number in the BAC clones. Their work also suggested that *Glu-D3* encodes most of the LMW-GS genes and could be the largest locus among three loci in hexaploid wheat.

It was found that the A genome can encode LMW-i and LMW-m subunits (Lee et al. 1999; Ikeda et al. 2002) while both B and D genomes encode two types of LMW-GS: LMW-m and LMW-s (Ikeda et al. 2006). In the current study, we found that the M genome encodes at least three types of LMW-GS. The largest variations were found to occur in the repetitive domain and the C-terminal II. Our results and previous reports showed that these variations were mainly resulted from SNPs and InDels presented in the LMW-GS genes, which occurred frequently in both hexaploid wheat and related species (An et al. 2006; Pei et al. 2007; Li et al. 2008a, b). These variations are generally considered to result from dot variation, unequal crossing-over and/or slip-mismatching during replication (D'Ovidio and Masci 2004). Unequal crossing over and/or slippage could result in striking expansion or contraction of glutenin genes (Anderson and Greene 1989). This mechanism is supported by several unusual HMW and LMW glutenin genes (D'Ovidio et al. 1996; Li et al. 2007). Recently, a new molecular mechanism of nonhomologous illegitimate recombination in HMW-GS and LMW-GS genes was found, which was believed to be an important molecular mechanism responsible for allelic variations of glutenin genes in addition to a basic homologous recombination mechanism (Zhang et al. 2008; Li et al. 2008b). Illegitimate recombination can lead to genome expansion and contraction, and therefore is considered to be a major factor in wheat (Wicker et al. 2003) and Arabidopsis (Devos et al. 2002) genome evolution.

Classification of the LMW-GS genes

LMW-GS are generally divided into three major subgroups, viz. LMW-s, LMW-m and LMW-i types according to the traditional criteria of the first amino acid residue of mature protein (Lew et al. 1992; Cloutier et al. 2001; D'Ovidio and Masci 2004). In addition, Ikeda et al. (2002)

Table 4 Estimation of divergence time (MYA) among 21 LMW-GS genes (below the diagonal: divergent time of LMW-GS gene, above the diagonal: standard error)

LMW-GS gene	Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. *AcLMW-L1*	LMW-l	–	0.26	0.31	0.63	0.66	0.80	0.92	0.95	0.96	0.97	1.04	0.99	0.99	1.00	1.00	1.03	0.95	0.93	0.93	1.00	0.96
2. *AcLMW-L2*	LMW-l	0.88	–	0.30	0.59	0.63	0.81	0.93	0.95	0.96	0.96	1.02	0.99	0.99	0.99	0.99	1.02	0.94	0.92	0.93	0.99	0.94
3. *AcLMW-L3*	LMW-l	1.08	0.99	–	0.59	0.62	0.80	0.91	0.93	0.94	0.95	1.02	0.95	0.96	0.98	0.97	1.01	0.91	0.90	0.90	0.95	0.91
4. *AY606257-l*	LMW-l	4.51	4.42	4.03	–	0.29	0.76	0.86	0.87	0.88	0.88	0.96	0.92	0.90	0.91	0.91	0.99	0.83	0.83	0.80	0.83	0.86
5. *AY724436-l*	LMW-l	5.20	5.11	4.71	0.88	–	0.79	0.85	0.87	0.86	0.87	1.00	0.93	0.93	0.93	0.93	1.01	0.87	0.88	0.84	0.86	0.88
6. *AcLMW-I2*	LMW-i	9.04	9.13	8.93	8.15	8.74	–	0.82	0.87	0.83	0.83	0.93	0.92	0.86	0.97	0.88	0.95	0.80	0.82	0.78	0.82	0.83
7. *AY542896-i*	LMW-i	11.20	11.29	11.09	9.52	9.62	8.35	–	0.38	0.54	0.54	0.96	0.91	0.89	0.97	0.90	0.94	0.86	0.88	0.83	0.87	0.88
8. *DQ217661-i*	LMW-i	11.59	11.68	11.49	9.92	10.41	8.93	1.77	–	0.54	0.54	0.97	0.91	0.90	0.98	0.91	0.97	0.90	0.93	0.88	0.91	0.94
9. *DQ307387-i*	LMW-i	11.68	11.78	11.59	10.21	10.51	8.64	3.24	3.24	–	0.24	1.00	0.96	0.94	0.98	0.95	0.99	0.91	0.94	0.89	0.90	0.94
10. *DQ345449-i*	LMW-i	11.68	11.78	11.59	10.21	10.51	8.64	3.24	3.24	0.59	–	0.97	0.98	0.93	0.98	0.94	0.97	0.91	0.93	0.88	0.89	0.93
11. *AcLMW-m1*	Group2	13.74	13.64	13.25	12.08	12.67	12.28	13.25	13.55	13.85	13.64	–	0.86	0.85	0.80	0.85	0.62	0.90	0.90	0.90	0.92	0.93
12. *AcLMW-m2*	Group1	12.28	12.37	11.78	10.70	11.39	10.21	9.52	9.72	10.51	10.70	9.04	–		0.88	0.51	0.86	0.73	0.75	0.72	0.73	0.71
13. *AY263369-m*	Group1	12.37	12.28	11.88	10.90	11.59	9.82	9.43	9.82	10.41	10.41	9.13	2.65	–	0.92	0.21	0.83	0.67	0.67	0.66	0.68	0.68
14. *AY841014-m*	Group2	13.06	13.16	12.96	11.88	12.47	13.25	13.85	14.14	13.45	13.45	7.56	10.12	10.31	–	0.92	0.72	0.90	0.92	0.90	0.91	0.93
15. *DQ287977-m*	Group1	12.28	12.17	11.78	10.80	11.49	9.92	9.33	9.72	10.31	10.31	9.04	2.55	0.49	10.01	–	0.84	0.66	0.67	0.64	0.68	0.67
16. *AJ293099-m*	Group2	13.75	13.64	13.25	12.08	12.67	12.57	13.06	13.55	13.64	13.45	4.12	8.93	8.64	7.36	8.54	–	0.83	0.86	0.83	0.90	0.86
17. *DQ357057-s*	LMW-s	10.80	10.80	10.40	9.04	9.72	8.44	9.04	9.82	9.62	9.72	9.33	5.20	4.91	9.82	4.62	8.74	–	0.46	0.24	0.46	0.43
18. *AB164415-s*	LMW-s	10.90	10.60	10.31	9.43	10.12	9.33	9.72	10.51	10.60	10.51	10.01	5.89	5.40	10.70	5.30	9.52	2.46	–	0.46	0.56	0.51
19. *DQ357058-s*	LMW-s	10.41	10.40	10.01	8.64	9.33	8.15	8.64	9.43	9.33	9.33	9.23	5.11	4.71	9.62	4.42	8.54	0.59	2.46	–	0.47	0.42
20. *EU369722-s*	LMW-s	10.80	10.60	10.21	8.74	9.43	8.64	9.43	10.21	9.72	9.72	9.52	5.79	5.79	9.43	5.50	9.13	2.65	3.54	2.55	–	0.51
21. *AB164416-s*	LMW-s	10.80	10.80	10.41	9.33	9.82	9.23	9.62	10.60	10.51	10.51	9.62	5.50	5.40	9.82	5.11	9.33	1.86	2.75	1.86	2.95	–

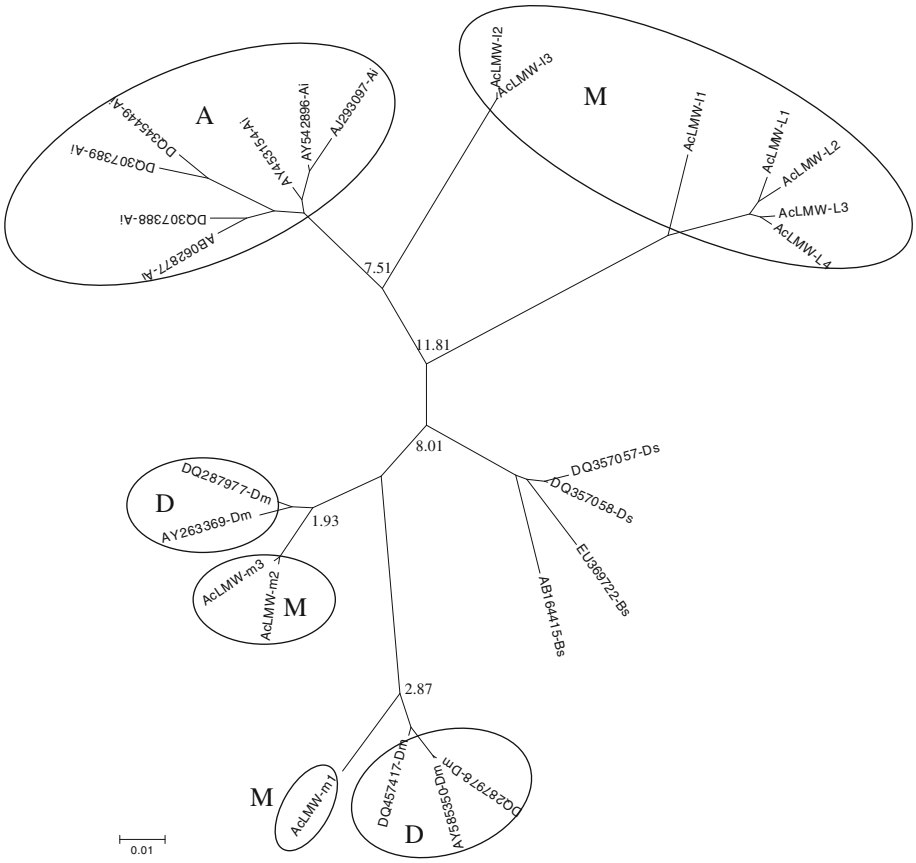

Fig. 5 Neighbor-joining tree based on the complete sequences of 26 LMW-GS genes located in different genomes. The first letter of suffix of GenBank accession number indicated the different genomes of LMW-GS genes, and the second one showed the different types of LMW-GS. The divergent times were specified by the numbers, and the unit was million years ago (MYA)

classified LMW-GS into 6 types and 12 groups based on the dispersal of cysteine residues, N- and C-terminal sequences. Long et al. (2005) analyzed the deduced N-terminal amino acid sequences of more than 90 complete and partial LMW-GS genes from common wheat and related species and divided them into seven groups. Each group could be specifically assigned onto a specific locus. In the present study, a new LMW-l type LMW-GS gene was isolated and characterized from *Aegilops comosa*, which had the N-terminal sequence of LPISQQQQ and significant different structural characteristics from other types. This kind of LMW-GS genes was also present in the wild source *Agropyron elongatum* ($StStE^eE^bE^x$, $2n = 70$) as reported by Luo et al. (2005). The validity of LMW-l type classification was further supported by the presence of specific cysteine residues and phylogenetic analysis of *Glu-3* gene family.

Ikeda et al. (2006) found that a single LMW-GS gene produced two different N-terminal sequences SHIPGLERPS and MENSHIPGL through 2-DE and N-terminal sequencing, which might result from post-translational processing of the LMW-GS gene by an asparaginyl endoprotease. Comparing the N-terminal sequences of LMW-m and LMW-s type subunits, LMW-s-type is likely to be identified as LMW-m-type or derivatives of LMW-m-type subunits if the unit M-E-T (N) of the LMW-m-type was omitted (Huang and Cloutier 2008). LMW-s-type genes lost the unit M-E-N due to post-translational cleavage by an asparaginyl endoprotease or the chemical conditions during sample preparation. According to our analysis, all LMW-s-type genes contained the unit M (I)-E-N; the divergent time analysis between LMW-s and LMW-m type genes indicated that the N-terminal sequences of the LMW-s-type (SHIPGLERPS) evolved from those of LMW-m type (MENSHIPGL). Since the LMW-s type subunit with the N-terminal MENSHIPGL was confirmed to present in the seed endosperm (Ikeda et al. 2006) and most of the LMW-GS genes isolated from common wheat and related species so far were LMW-m type, it could deduce that the possibility of post-translational cleavage by an asparaginyl endoprotease was very low. Furthermore, both LMW-m and LMW-s genes possessed similar structure and very close relationships at the amino acid sequence level (Huang and Cloutier 2008). Therefore, the LMW-s type subunit occurs in a low frequency, and in some cases, LMW-m type subunit was identified as LMW-s type in the past. However, our results revealed that the LMW-s type genes were clustered into a separate subgroup

in the first class and was diverged in a much later stage (Figs. 3, 4). It could conclude that the LMW-s type gene was a variant form, which diverged from the LMW-m type gene about 4–6 MYA (Table 4).

It is clear that the traditional classification needs to be refined. A new LMW-GS gene classification system (Fig. 7) is proposed in the current study according to the N-terminal sequences, the positions of the cysteine residues and sequence homologous analysis (Table 1; Fig. 3). LMW-GS genes were divided into two major types: LMW-M and LMW-I based on the first amino acid of N-terminal sequences and each type was further separated into two subclasses based on the positions of cysteine residues (Table 1), namely LMW-m and LMW-s in the LMW-M type and LMW-i and LMW-l in the LMW-I type. In addition to the difference of the first cysteine residues in N-terminal, the LMW-M type genes had a special cysteine residue in the N-terminal or repetitive domain while the LMW-i type genes possessed an extra cysteine in the C-terminal I domain. The LMW-l genes generally had a special cysteine in the signal peptide and repetitive domain, and an insertion with 10–22 residues (consensus $LGQQPQ_{5-17}$) in the end of the C-terminal II (Fig. 1). The homology tree analysis of 34 LMW-GS genes (Fig. 3) also supported this new classification. The LMW-M and LMW-I type genes had 75% sequence homology while 79 and 77% homology were presented between LMW-m and LMW-s, and LMW-l and LMW-i type genes, respectively. Genes in each subclass had higher homology, up to 87–96%.

According to the recent reports by Ikeda et al. (2006) and Huang and Cloutier (2008), some special LMW-GS genes with the N-terminal sequences of IENSHIPGL were confirmed to present in common wheat. In fact, this kind of subunit belongs to LMW-s type based on the positions of cysteine residues and homology analysis although their first amino acid residue of N-terminal was isoleucine. This residue resulted is most likely resulted from a G–T dot mutation (ATG → ATT), consequently changing methionine to isoleucine. For the same reason, the AcLMW-I1 subunit should belong to LMW-l type.

Origin of LMW-l type genes and evolution of the *Glu-3* loci

According to our results and previous reports so far, origin of the LMW-l type genes and their evolutionary relationships among LMW-GS genes at the *Glu-3* loci could be drawn. During the evolutionary history of LMW-GS genes, some major variation events, including the N-terminal deletion, two special Cys substitutions in the signal peptide and in the end of repetitive domain II and a long insertion of 10–22 residues (consensus $LGQQPQ_{5-17}$) in the end of the C-terminal II, could occur in the LMW-m type genes that might be the progenitor of other LMW-GS genes. These events, possibly resulting from unequal crossing over and/or slippage during homologous recombination as well as dot mutation, leaded to the generation of LMW-l type genes. In addition, intra-strand illegitimate recombination might be another important molecular mechanism for the origin and evolution of the *Glu-3* genes as reported by Li et al. (2008b). Since each *Glu-3* locus could have at least 2 LMW-GS genes that locate 81–150 kb apart (Wicker et al. 2003; Huang and Cloutier 2008), intra-strand unequal crossing over could occur between two genes, leading to a fragment insertion of 30–69 bp (10–23 residues) in the LMW-l type genes as shown in Fig. 6. The same mechanism would also result in N-terminal deletion. These variation events were estimated to occur in 12–14 MYA. Illegitimate recombination required only a few base pairs of sequence identity (Wicker et al. 2003) or even no sequence homology (Arguello et al. 2006), and therefore, higher identity among LMW-GS genes could facilitate intra-strand illegitimate recombination at the *Glu-3* locus.

According to our classification and phylogenetic analysis results (Figs. 3, 4; Table 4), the evolutionary origin of different types of LMW-GS genes at the *Glu-3* loci was proposed as shown in Fig. 7. Among the LMW-GS gene

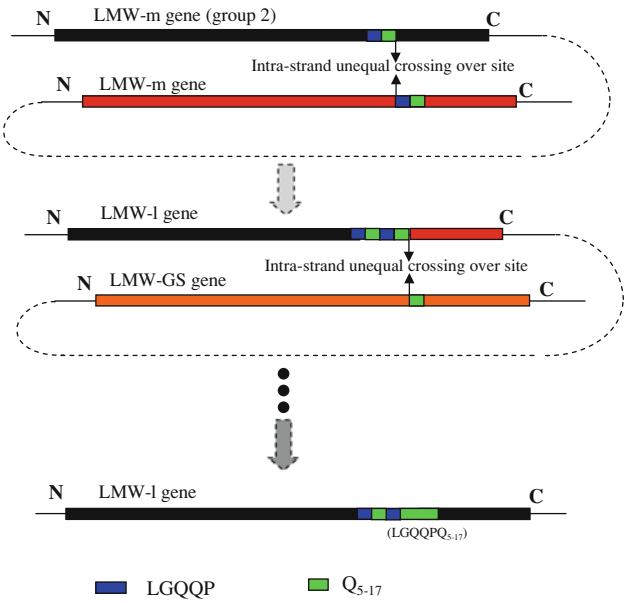

Fig. 6 Intrastrand unequal crossing over (illegitimate recombination) mechanism resulting in a long fragment insertion with 10–22 residues (consensus $LGQQPQ_{5-17}$) in the LMW-l type genes. At least two times of unequal crossing over between two LMW-GS genes at the same locus might occur. The first time leaded to a $LGQQPQ_5$ insertion and the second time resulted in an insertion of polyglutamine chain (Q_{1-12}). These unequal crossing over events might occur 12–14 MYA

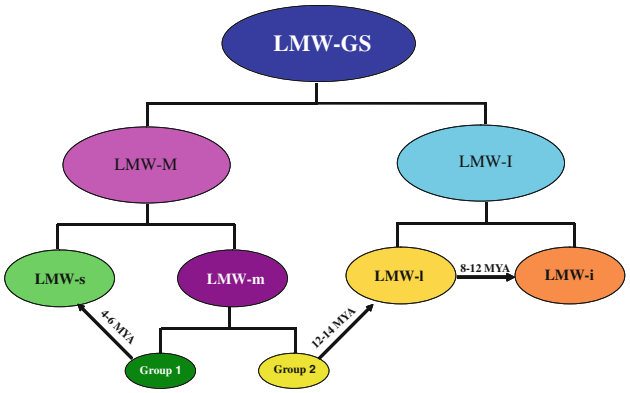

Fig. 7 The new LMW-GS classification system and their evolutionary relationships among LMW-GS gene family

family, the LMW-m type genes could be considered to be progenitor of other LMW-GS genes, and LMW-s type genes were a variant form, which was divergent from the group 1 of LMW-m type genes at 4–6 MYA. The LMW-l genes is a mid-type, which was diverged from the group 2 of LMW-m type genes at about 12–14 MYA. LMW-i type genes evolved from LMW-l genes at about 8–12 MYA through a dipeptide (LP) and a fragment (30–66 bp) deletion due to homologous or non-homologous illegitimate recombination.

To date, the evolutionary relationships between the M genome and A, B and D genomes in common wheats are not clear. Earlier work suggested that either some extinct species were the donors of the modified genomes or they were significantly rearranged during evolution (Kihara 1954). The M and D genomes were shown to have the same origin (Badaeva et al. 1996). Our phylogenetic (Fig. 3) and divergent time analysis (Fig. 5) indicated that the LMW-i and LMW-m type subunits encoded by the M genome were more similar with those encoded by A and D genomes, suggesting that the M genome may have introgressions with A and D genomes. Our results also suggested that the genome divergence among *Triticum* and *Aegilops* species were much later than that of LMW-GS gene types at *Glu-3* loci.

Acknowledgments This research was financially supported by grants from the National Natural Science Foundation of China (30830072, 30771334), the Chinese Ministry of Science and Technology (2009CB118303, 2006AA10Z186) and Key Project of National Plant Transgenic Genes of China (2008ZX08002-004).

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 (*T. monococcum* L.). Theor Appl Genet 113:383–395

Anderson OD, Greene 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

Arguello JR, Chen Y, Yang S, Wang W, Long M (2006) Origination of an X-linked testes chimeric gene by illegitimate recombination in *Drosophila*. PLoS Genet 2:745–754

Badaeva ED, Friebe B, Gill B (1996) Genome differentiation in Aegilops. 2. Physical mapping of 5S and 18S–26S ribosomal RNA gene families in diploid species. Genome 39:1150–1158

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Protein structure prediction servers at University College London. Nucleic Acids Res 33(Web Server Issue):W36–W38

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

Cornish GB, Bekes F, Allen HM, Martin DJ (2001) Flour proteins linked to quality traits in an Australian doubled haploid wheat population. Aust J Agric Res 52:1339–1348

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

D'Ovidio R, Lafiandra D, Porceddu E (1996) Identification and molecular characterization of a large insertion within the repetitive domain of a high-molecular-weight glutenin subunit gene from hexaploid wheat. Theor Appl Genet 93:1048–1053

Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res 12:1075–1079

Dvorak J (1998) Genome analysis in the *Triticum-Aegilops* alliance. In: Proceedings of the 9th international wheat genetics symposium, Saskatoon, Sask. pp 8–11

Gao SC, Gu YQ, Wu JJ, Coleman-Derr D, Huo N, Crossman C, Jia JZ, Zuo Q, Ren ZL, Anderson OD, Kong XY (2007) Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. Plant Mol Biol 65:189–203

Gaut BS, Morton BR, Mccaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* Parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93:10274–10279

Gupta RB, Shepherd KW (1990) Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutenin. 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

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

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99:8133–8138

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

Ikeda TM, Araki E, Fujita Y, Yano H (2006) Characterization of low-molecular-weight glutenin subunit genes and their protein products in common wheats. Theor Appl Genet 112:327–334

Jackson EA, Holt LM, Payne PI (1983) Characterisation of high molecular weight gliadin and low-molecular-weight glutenin

subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localisation of their controlling genes. Theor Appl Genet 66:29–37

Jiang CX, Pei YH, Zhang YZ, Li XH, Yao DN, Yan YM, 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(5):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

Kihara H (1954) Considerations on the evolution and distribution of *Aegilops* species based on the analyser-method. Cytologia 19:336–357

Kimber G, Feldman M (1987) Wild wheat, an introduction. College of Agriculture University of Missouri, Columbia, p 142

Kumar S, Tamura K, Nei M (2004) *MEGA3*: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Lee Y-K, Ciaffi M, Appels R, Morell MK (1999) The low-molecular-weight glutenin subunit proteins of primitive wheats. II. The genes from A-genome species. Theor Appl Genet 98:126–134

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 XH, Zhang YZ, Gao LY, Wang AL, Ji KM, He ZH, Appels R, Ma WJ, Yan YM (2007) Molecular cloning, heterologous expression and phylogenetic analysis of a novel y-type HMW glutenin subunit gene from G genome of *Triticum timopheevi*. Genome 50:1130–1140

Li XH, Wang AL, Xiao YH, Yan YM, He ZH, Appels R, Ma W, Hsam SLK, Zeller FJ (2008a) Cloning and characterization of a novel low molecular weight glutenin subunit gene at the *Glu-A3* locus from wild emmer wheat (*Triticum turgidum* L. var. dicoccoides). Euphytica 159:181–190

Li XH, Ma W, Gao LY, Zhang YZ, Wang AL, Ji KM, Wang K, Appels R, Yan Y (2008b) A novel chimeric LMW-GS gene from the wild relatives of wheat *Ae. kotschyi* and *Ae. juvenalis*: evolution at the *Glu-3* loci. Genetics 180:93–101

Long H, Wei YM, Yan ZH, Baum B, Nevo E, Zheng YL (2005) Classification of wheat low-molecular-weight glutenin subunit genes and its chromosome assignment by developing LMW-GS group-specific primers. Theor Appl Genet 111:1251–1259

Luo Z, Chen F, Feng D, Xia G (2005) LMW-GS genes in *Agropyron elongatum* and their potential value in wheat breeding. Theor Appl Genet 111:272–280

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

Maruyama-Funatsuki W, Takata K, Nishio Z, Tabiki T, Yahata E, Kato A, Saito K, Funatsuki H, Saruyama H, Yamauchi H (2004) Identification of low-molecular weight glutenin subunits of wheat associated with bread-making quality. Plant Breed 123:355–360

Masci S, Lew EJL, Lafiandra D, Porceddu E, Kasarda DD (1995) Characterization of low-molecular-weight glutenin subunits in durum wheat by RP-HPLC and N-terminal sequencing. Cereal Chem 72:100–104

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–405

Nieto-Taladriz MT, Ruiz M, Martinez MC, Vazquez JF, Carrillo JM (1997) Variation and classification of B-low-molecular-weight glutenin subunit alleles in durum wheat. Theor Appl Genet 95:1155–1160

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

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. Nucl Acids Res 16:11376

Riley R, Chapman V, Johnson R (1968) Introduction of yellow rust resistance of *Aegilops comosa* into wheat by genetically induced homoeologous recombination. Nature 217:378–384

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

Shewry PR, Halford NG (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot 53:947–958

Tanaka H, Shimizu R, Tsujimoto H (2005) Genetical analysis of contribution of low-molecular-weight glutenin subunits to dough strength in common wheat (*Triticum aestivum* L.). Euphytica 141:157–162

Tao HP, Kasarda DD (1989) Two-dimensional gel mapping and N-terminal sequencing of LMW-glutenin subunits. J Exp Bot 40:1015–1020

Tatham AS, Miflin BJ, Shewry PR (1985) The beta-turn conformation in wheat gluten proteins: relationship to gluten elasticity. Cereal Chem 62:405–442

Tatham AS, Field JM, Smith SJ, Shewry PR (1987) The conformations of wheat gluten proteins. II. Aggregated gliadins and low molecular weight subunits of glutenin. J Cereal Sci 51:203–214

Tatham AS, Drake AF, Shewry PR (1990) Conformational studies of synthetic peptides corresponding to the repetitive region of the high molecular weight (HMW) glutenin subunits of wheat. J Cereal Sci 11:189–200

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15:1186–1197

Yan Y, Prodanovic S, Mladenov N, Milovanovic M (1999) Studies on genetic control of low molecular weight glutenin subunits in wheat by A-PAGE analysis. Cereal Res Commun 27:251–257

Yan Y, Hsam SLK, Yu JZ, Jiang Y, Zeller FJ (2003) Allelic variation of the HMW glutenin subunits in *Aegilops tauschii* accessions detected by Sodium Dodecyl Sulphate (SDS-PAGE), Acid Polyacrylamide Gel (A-PAGE) and Capillary Electrophoresis. Euphytica 130:377–385

Yan Y, Zheng J, Xiao Y, Yu J, Hu Y, Cai M, Li Y, Hsam SLK, Zeller FJ (2004) Identification and molecular characterization of a novel y-type *Glu-D^t1* glutenin gene of *Aegilops tauschii*. Theor Appl Genet 108:1349–1358

Zhang W, Gianibelli MC, Rampling LR, 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

Zhang Y, Li Q, Yan Y, Zheng J, An X, Xiao Y, Wang A, Pei Y, Wang H, Hsam SLK, Zeller FJ (2006) Molecular characterization and phylogenetic analysis of a novel glutenin gene (*Dy10.1^t*) from *Aegilos tauschii*. Genome 49:735–745

Zhang YZ, Li XH, Wang AL, An XL, Zhang Q, Pei YH, Gao LY, Ma WJ, Appels R, Yan YM (2008) Novel x-type high-molecular-weight glutenin genes from Aegilops tauschii and their

implications on the wheat origin and evolution mechanism of Glu-D1–1 proteins. Genetics 178:23–33

Zhao H, Wang R, Guo A, Hu S, Sun G (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 bread wheat. Hereditas 141:193–198

Zhao XL, Xia XC, He ZH, Gale KR, Lei ZS, Appels R, Ma WJ (2006) Characterization of three low-molecular-weight *Glu-D3* subunit genes in common wheat. Theor Appl Genet 113: 1247–1259