Cloning and characterization of special HMW glutenin subunit genes from *Aegilops longissima* L. and their potential for wheat quality improvement

Jinxin Hu¹ · Jian Wang¹ · Xiong Deng¹ · Yueming Yan^{1,2}

Received: 26 April 2019 / Accepted: 7 June 2019 / Published online: 14 June 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

Identification and cloning of new glutenin genes from wheat-related species can provide candidate gene resources for quality improvement of wheat. In this study, ten special high-molecular-weight glutenin subunits (HMW-GS), including five x-type (1S¹2x, 1S¹16x, 1S¹17x, 1S¹23x, and 1S¹25x) and five y-type (1S¹2y, 1S¹6y₁, 1S¹16y, 1S¹17y, and 1S¹23y) from *Aegilops longissima* L. (S¹S¹, 2n = 2x = 14) were identified, and their complete encoding genes were cloned by allelic-specific polymerase chain reaction (AS-PCR). The deduced amino acid (aa) residues of the x-type subunit genes ranged from 821 aa (2469 bp) to 941 aa (2829 bp), while those of the y-type subunit genes varied from 749 aa (2250 bp) to 771 aa (2361 bp). These special HMW-GS had a longer central repetitive domain with more glutamine repeats and glutamine residues compared to the previously characterized HMW-GS in common wheat, which provided a structural basis for superior gluten quality formation. The authenticity of the four cloned genes were verified by matrix-assisted laser desorption ionization time-of-flight/time-of-flight mass spectrometry (MALDI-TOF/TOF-MS). Abundant single-nucleotide polymorphism (SNP) and insertion/deletion (InDel) variations among these genes were identified, which would benefit for developing specific molecular markers used for wheat gluten quality improvement. Phylogenetic analysis revealed that the 1S¹-encoded HMW-GS had close relationships with those from bread wheat, which were divergent from *Triticum* species at 2.10–10.00 million years ago. Our results indicate that the 1S¹ genome contains superior candidate glutenin genes that have potential application values for the improvement of wheat bread making quality.

Keywords *Aegilops longissima* L. · HMW-GS · AS-PCR · MALDI-TOF/TOF-MS · Phylogenetics · Gluten quality

Jinxin Hu, Jian Wang, and Xiong Deng have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1803-8>) contains supplementary material, which is available to authorized users.

✉ Yueming Yan
yanym@cnu.edu.cn

¹ College of Life Science, Capital Normal University, Beijing, Xisanhuan Beilu 105, Beijing 100048, China

² Hubei Collaborative Innovation Center for Grain Industry (HCICGI), Yangtze University, Jingzhou 434025, China

Introduction

Wheat (*Triticum aestivum* L.) is one of the three major grain crops and about 35% of the population in the world uses wheat as a staple food. Meanwhile, wheat is also an important source of plant protein for humans. Wheat flour can be widely used to make bread, noodles, biscuits, cakes, steamed bread, and other foods. Two kinds of major storage proteins gliadins and glutenins are present in the wheat endosperm, which differ in functional property and solubility in different solvents. Glutenins consist of high- and low-molecular-weight glutenin subunits (HMW-GS and LMW-GS), dissolve in dilute acid or dilute alkali, and endow dough viscoelasticity, while gliadins dissolve in 70–90% alcohol and confer wheat dough extensibility (Payne 1987; Wrigley 1996).

HMW-GS are encoded by the *Glu-1* loci located on the long arm of chromosomes 1A, 1B and 1D of the first

partial homologous group of wheat (Lawrence and Shepherd 1981). Each locus consists of two closely linked genes, encoding a larger x-type subunit with molecular weight 80–88 kDa and a smaller y-type subunit with molecular weight 67–73 kDa, depending on the composition of the motifs in the repeating region and the cysteine content (Shewry et al. 2003). Theoretically, hexaploid wheat can express 6 HMW-GS (1Ax, 1Ay, 1Bx, 1By, 1Dx, and 1Dy), but usually 3 (1Bx, 1Dx and 1Dy) to 5 (1Ax, 1Bx, 1By, 1Dx, and 1Dy) genes express due to the gene silencing (Yang et al. 2006). In general, the 1Bx, 1Dx, and 1Dy encoded genes all have expression activity, while the 1Ax and 1By encoded genes are sometimes silent in hexaploid wheat, especially the 1Ay subunit (Jiang et al. 2009). The allelic variations at *Glu-1* loci are closely associated with gluten quality. For example, 1Dx5 + 1Dy10 (*Glu-D1d*), 1Bx7 + 1By8 (*Glu-B1b*), and 1Bx17 + 1By18 (*Glu-B1i*) subunits have higher quality score and confer superior dough strength and breadmaking quality, while 1Dx2 + 1Dy12 (*Glu-D1a*), 1Bx20 (*Glu-B1e*), and 1Bx7 + 1By9 (*Glu-B1c*) subunits are related to poor quality performance (Payne 1987; Shewry et al. 1992; Yan et al. 2009; Jiang et al. 2019; Guo et al. 2019). At present, more than 30 alleles at *Glu-1* loci are identified, and most of these genes have been cloned and well characterized (Payne and Lawrence 1983). However, the allelic variation of HMW-GS genes in common wheat is limited and the identified alleles related to superior gluten quality still lack (Li et al. 2007). Thus, it is highly important to discover and clone new candidate *Glu-1* genes in the related species for further wheat quality improvement (Liang et al. 2015).

Aegilops, the closest related genus of wheat, is an important genus of gramineous plants, mainly distributed in Europe and Asia, and contains a large number of valuable genetic resources (Ma et al. 2013). *Aegilops* consists of 22 species, including 7 genomes of S, M, C, U, N, D, and T, of which the S genome contains S (*Ae. speltoides*, which is considered to be the donor of the common wheat B genome), S^b (*Ae. bicornis*), S^l (*Ae. longissima*), S^s (*Ae. searsii*), and S^{sh} (*Ae. sharonensis*) genomes. As a diploid species of *Aegilops*, *Ae. longissima* (S^lS^l, 2n = 2x = 14) has rich genetic resources for wheat improvement such as powdery mildew resistance gene *Pm13* (Alberto et al. 2003), eyespot resistance genes (Sheng and Murray 2013), and drought resistance (Zhou et al. 2016). In addition, *Ae. longissima* contains rich trace elements such as iron, zinc, copper, manganese, calcium, magnesium, and potassium, and the content of trace elements in the hybrid progeny was promoted through crossing with common wheat (Neelam et al. 2013). *Ae. longissima* also has abundant glutenin allelic variations and special HMW-GS that have shown potential values for wheat quality improvement (Wang et al. 2013; Garg et al. 2014; Zhu et al. 2015).

In this study, we aim to investigate the allelic variations at *Glu-1* in *Ae. longissima* accessions and identify new 1S^l-encoded HMW-GS by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Then, allelic-specific polymerase chain reaction (AS-PCR) was used to clone their encoded genes. Ultimately, the complete sequences of 5 x-type and 5 y-type HMW glutenin subunits were obtained from *Ae. longissima*. The molecular characterization of these subunit genes showed special structural features and have potential application values for wheat quality improvement.

Materials and methods

Plant materials

Six *Ae. longissima* accessions kindly provided by Dr. Hsam from Plant Breeding Institute, Technical University of Munich, Germany, were used as materials, including PI542196, PI604108, PI604129-1, PI604130, PI604136-1, and PI60414. Chinese Spring (CS) with HMW-GS (null, 1Bx7 + 1By8, and 1Dx2 + 1Dy12), and CS-*Ae. longissima* 1S^l/1B substitution line CS-1S^l(1B) (null, 1S^lx2.3* + 1S^ly16*, 1Dx2 + 1Dy12) (Wang et al. 2013) were used as the control for HMW-GS identification.

HMW-GS extraction and SDS-PAGE

Seed glutenin extraction was according to the previously described method (Mackie et al. 1996). The albumins, globulins, and gliadins were first extracted and excluded from 15 mg seeds using 75 µL distilled water, 75 µL dilute salt solutions, and 120 µL 30% ethanol, respectively. Then, glutenins were extracted using commonly used glutenin extraction buffer (50% isopropanol, 80 µL Tris-HCl, pH 8.0) with 64.83 µL dithiothreitol (DTT) and 1.4% 4-vinylpyridine (v/v). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed with Bio-Rad PROTEAN II XL equipment in 12% gel and electrophoresed at 15 mA for 2 h based on the previously described method (Yan et al. 2003a). The new HMW-GS identified in *Ae. longissima* were named according to the accession number and subunit types (x- and y-types).

DNA extraction and PCR amplification

Genomic DNA was isolated from seedling leaves at three-leaf stage with CTAB protocol following the procedure of Yan et al. (2004) with slight modifications. The primers for allelic-specific polymerase chain reaction (AS-PCR) were designed based on the previously cloned sequences (Wan et al. 2002), and AS-PCR amplification included an initial

denaturation step at 95 °C for 3 min, followed by denaturation at 95 °C for 15 s, anneal at 58 °C for 15 s, extension at 72 °C for 2 min 30 s in total of 32 cycles, and a final extension step at 72 °C for 10 min.

Molecular cloning and sequencing

The PCR fragments of the expected size were purified from the gel using the Gel Extraction Kit (Omega). Purified products were ligated into CE-Vector (Vazyme, Nanjing, China) and transformed into Trans-2-blue receptor cells (TransGen, Beijing, China). DNA sequencing from three random positive clones of each PCR fragment was carried out by TaKaRa Biotechnology, Dalian, China.

Sequence alignment, SNP, and InDel identification

Multiple sequence comparison of HMW-GS nucleotide and protein sequences were completed by Bioedit 7.0.1.1. Based on sequence alignment, single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) were identified with software Bioedit 7.0.

MALDI-TOF/TOF-MS

To verify the authenticity of the cloned HMW-GS genes, the corresponding subunits were collected from SDS-PAGE gels and digested by trypsin. The digested protein subunits were further identified using matrix-assisted laser desorption ionization time-of-flight/time-of-flight mass spectrometry (MALDI-TOF/TOF-MS) based on the previously described method (Ge et al. 2012).

Construction of phylogenetic tree and estimation of divergence time

Clustal W program and MEGA 7.0 were used to construct phylogenetic tree and the estimation of the divergence times among HMW-GS genes was based on Gaut et al. (1996). The

Clustal W program was used to make a multiple alignment with homologous nucleotide sequences; the alignment file was converted to software MEGA 7.0 by the complete coding regions of HMW-GS genes with bootstrap values 1000. The divergence times of HMW-GS genes were estimated using MEGA 7.0 with the evolution rate as 6.5×10^{-9} substitution/site year according to Allaby et al. (1999).

Results

Identification of novel HMW-GS in *Ae. longissima* L.

Six *Ae. longissima* accessions were consecutively reproduced and saved in our laboratory, which showed higher purity and genetic stability. SDS-PAGE analysis showed extensive allelic variations in HMW-GS compositions among 6 accessions, and 13 new HMW-GS were identified and named: including PI542196 ($1S^2x + 1S^2y$), PI604108 ($1S^6x + 1S^6y_1 + 1S^6y_2$), PI604129 ($1S^16x + 1S^16y$), PI604130 ($1S^17x + 1S^17y$), PI604136 ($1S^23x + 1S^23y$), and PI604142 ($1S^25x + 1S^25y$) as indicated in Fig. 1. Compared to the HMW-GS compositions in CS and CS-1S¹(1B), the new subunits identified in *Ae. longissima* accessions had significantly slower electrophoretic mobilities and higher molecular weight than 1Bx7 + 1By9 subunits of CS. Particularly, six x-type and six y-type subunits moved clearly slower than the greater x-type and y-type subunits $1S^1x2.3^* + 1S^1y16^*$ from CS-1S¹(1B) substitution line, respectively. These special HMW-GS could serve as new genetic resources for wheat quality improvement.

Cloning and characterization of ten novel HMW-GS genes from *Ae. longissima* L.

The specific primers SxF/SxR and SyF/SyR were designed (Table 1) and used for amplify 1S¹-encoded x-type and y-type new subunit genes, respectively. AS-PCR results showed that the specific amplified fragments with about

Fig. 1 Identification of new HMW-GS in *Ae. longissima* L. by SDS-PAGE. Lines 1–8: China Spring (CS), PI542196, PI604108, PI604129, PI604130, PI604136, PI604142, and CS-1S¹(1B)

Table 1 The sequences of primers used to amplify HMW-GS genes

Primers	The sequence of primers
SxF	5'-GGATCTTCCAGAGATCCTTCACTATCTCATCATCACCCAC-3'
SxR	5'-CTGCCGTTTCGACGATTAGGAGTCTGTTCGCATTCACTGGGC-3'
SyF	5'-GGATCTTCCAGAGATAATTTTCATCATCACCCATAACACCG-3'
SyR	5'-CTGCCGTTTCGACGATATTGGGTTTACTCTAGTTACACG-3'

Fig. 2 AS-PCR amplification of x-type and y-type HMW-GS genes from 1S¹ genome of *Ae. longissima* L. with primers SxF+SxR and SyF+SyR. **a** PCR amplification of x-type HMW-GS from 1S genome of *Ae. longissima* L., 1, 2: 1S¹2x; 3, 4: 1S¹16x; 5, 6: 1S¹17x; 7, 8: 1S¹23x; 9, 10: 1S¹25x; M: DNA makers. **b** PCR amplification of y-type HMW-GS from 1S genome of *Ae. longissima* L., 1, 2: 1S¹2y; 3, 4: 1S¹6y₁; 5, 6: 1S¹16y; 7, 8: 1S¹17y; 9, 10: 1S¹23y; M: DNA markers

2400–2800 bp and 2200–2300 bp were obtained (Fig. 2), which were corresponding to the sizes of x-type and y-type subunit genes, respectively. After collection, cloning, and sequencing of the specific amplification fragments, the complete open reading frames (ORFs) of five x-type (1S¹2x, 1S¹16x, 1S¹17x, 1S¹23x, and 1S¹25x) and five y-type (1S¹2y, 1S¹6y₁, 1S¹16y, 1S¹17y, and 1S¹23y) HMW-GS genes were obtained. The full length of five x-type subunit genes ranged from 2469 to 2829 bp, encoding 821–941 amino acid (aa) residues, while five y-type subunit genes had 2250–2316 bp that encoded 749–771 aa residues. Website predictions (https://web.expasy.org/compute_pi) showed that the deduced molecular weight of these encoding sequences were 87,475–100,010 Da for x-type subunits and 80,280–82,761 Da for y-type subunits, much higher than 1Bx7 (82,865 Da) and 1By8 (77,343 Da) subunits from common wheat (Table 2). Among them, the deduced molecular weight of the x-type (1S¹2x, 1S¹16x, and 1S¹25x) and all y-type subunits were also remarkably higher than

those of previously characterized 1S¹x2.3* (97,797 Da) and 1S¹y16* (78,165 Da) subunits (Wang et al. 2013), respectively. All cloned ten novel HMW-GS genes were deposited in GenBank with the accession numbers from MH699857 to MH699866.

Multiple sequence alignment of the deduced amino acid sequences using BioEdit7.0 showed that all ten new subunits had typical structural features of HMW-GS, including the signal peptide (21 aa), conserved N-terminal (81–87 aa), central repeat region (582–792 aa), and C-terminus (42 aa) (Figs. 3, 4). In addition, each x-type subunit contained four conserved cysteines (three at the N-terminus and one at the C-terminus), while the y-type subunits had seven conserved cysteines (five at the N-terminus, one at the repetitive domain, and one at the C-terminus) as other HMW-GS from bread wheat and related species (Figs. 3, 4; Table 2). The amino acid sequence repeat regions showed that five x-type new subunits contained a longer repetitive domain, with 8–11 nonapeptides (consensus GYYPTSPQQ and GYYPTSLQQ), 23–31 hexapeptides (consensus PGQGQQ and SGQGQQ), and 33–41 tripeptides (consensus GQQ). Obviously, the number of tripeptides in five x-type new subunits was much higher than 1S¹x2.3* and 1Bx7 (Table 2). The amino acid sequence alignment indicated high similarity of these x-type subunits with 1S¹x2.3* and 1Bx13 (Fig. 3), and the major differences occurred at the central repeat region. In particular, 1S¹2x, 1S¹16x, 1S¹25x, and 1S¹x2.3* had 84 and 20 aa insertion at the position 669–752 and 778–797, respectively, which resulted in the length increase of the central repeat region. The five y-type new subunits also had a longer central repeat region compared to those from bread wheat (Fig. 4). Their repeat region length ranged from 582 aa to 604 aa, and contained 20–22 hexapeptides (consensus PGQGQQ) and 11–12 nonapeptides (consensus GYYPTSP/LQQ or GHYPASQQQ). The nonapeptide repeats were much higher than 1S¹y16* (8) and 1By8 (4). Both 1S¹6y₁ and 1S¹23y subunits had one 15 aa insertion at the position 454–468. Five x-type and five y-type subunits had 280–332 (34.1–35.4%) and 240–250 (32.0–32.6%) glutamine residues, respectively. Although they had a similar glutamine content with 1S¹x2.3* (35.0%) + 1S¹y16* (31.9%) and 1Bx7 (33.1%) + 1By8 (33.2%), two x-type subunits 1S¹2x (332) and 1S¹25x (331) as well as all y-type subunits contained more glutamine residues than 1S¹x2.3* (329) + 1S¹y16* (239) and 1Bx7 (261) + 1By8 (239) (Table 2).

Table 2 Comparison of the mature protein sequences of new novel HMW-GS

Subunit types	GenBank accession no.	Molecular mass (Da)	ORFs (bp)	Amino acid number	Number of cysteine residues			Number of tripeptides	Number of hexapeptides	Number of nonapeptides	Number of glutamine	Glutamine content (%)
					N-ter.	Rep.	C-ter.					
1S ¹ 2x	MH699858	99,964	2829	941	3	0	1	4	31	11	332	35.3
1S ¹ 16x	MH699857	98,414	2787	927	3	0	1	4	28	11	328	35.4
1S ¹ 17x	MH699859	87,475	2469	821	3	0	1	4	23	8	280	34.1
1S ¹ 23x	MH699860	87,666	2478	824	3	0	1	4	23	9	283	34.3
1S ¹ 25x	MH699861	100,010	2817	938	3	0	1	4	31	11	331	35.3
1S ¹ x2.3*	MF805726	97,797	2829	941	3	0	0	4	31	11	329	35.0
1Bx7	X13927	82,865	2369	789	3	0	1	4	31	8	261	33.1
1S ¹ 2y	MH699862	80,280	2250	749	5	1	1	7	22	11	240	32.0
1S ¹ 6y ₁	MH699863	82,761	2316	771	5	1	1	7	20	11	249	32.3
1S ¹ 16y	MH699864	80,952	2268	755	5	1	1	7	22	11	243	32.2
1S ¹ 17y	MH699865	82,265	2307	768	5	1	1	7	21	12	250	32.6
1S ¹ 23y	MH699866	82,752	2316	771	5	1	1	7	22	11	248	32.2
1S ¹ y16*	JN408502	78,165	2250	749	5	1	1	7	25	8	239	31.9
1By8	AY245797	77,343	2163	720	5	1	1	7	25	4	239	33.2

Analysis of SNPs and InDels present in the cloned new HMW-GS genes

The coding sequences of five x-type and five y-type subunit genes were aligned with other 22 x-type and 18 y-type HMW-GS genes from common wheat and related species, respectively. The SNP and InDel variations present in these HMW-GS genes were identified and the results are listed in Tables 3 and 4. In general, all subunit genes showed extensive allelic variations. Five x-type genes had 64 SNPs at different positions, and the number of SNP per gene was between 12 and 15, of which 31 (48.4%) were resulted from the base transitions of G-A or C-T (Table 3). Five y-type genes contained 79 SNPs at different positions, and the number of SNP per gene was between 12 and 28, and 25 (31.64%) SNPs were resulted from the base transversions of C-A (Table 4). Moreover, more than half of the detected SNPs in five x-type genes and five y-type genes were nonsynonymous. In addition, two deletions at 96–105 and 3082–3089 in 1S¹16x, a deletion at 1355–1364 in both 1S¹6y₁ and 1S¹23y, and one deletion at 2579–2591 in 1S¹2y were detected.

Tandem MS analysis and verification of the cloned HMW-GS genes from *Ae. longissima* L.

To verify the authenticity of the cloned HMW-GS new genes, we selected four subunits (1S¹17x, 1S¹25x, 1S¹6y₁, and 1S¹17y) for further tandem mass spectrometry analysis. The corresponding subunits were collected from SDS-PAGE gel (Fig. 1) and digested by trypsin, and then used for MALDI-TOF/TOF-MS identification. The results showed that 2–6 protein peptides with 11–91 amino acid residues were well matched to their corresponding sequences specifically present in four subunits with high protein score CI % (Table 5). This further confirmed the reliability of the cloned genes.

Phylogenetic analysis

The complete encoding sequences of the cloned 10 new HMW-GS genes from *Ae. longissima*, and 10 x-type and 10 y-type subunit genes from common wheat and related species were used to construct neighbor-joining phylogenetic tree, including MF805726 (1S¹x2.3*), JN001483 (1S¹x2.9), X61009 (1Ax¹), M22208 (1Ax2*), X13927 (1Bx7), KF733216 (1Bx14), AB263219 (1Bx17), AJ437000 (1Bx20), X03346 (1Dx2), X12928 (1Dx5), JN408502 (1S¹y16*), JN001484 (1S¹y2.3), FJ404595 (1Ay), AY245797 (1By8), X61026 (1By9), KF733215 (1By15), EF540765 (1By16), KF430649 (1By18), X12929 (1Dy10), and BK006459 (1Dy12). The results showed that the x-type and y-type HMW-GS genes in the phylogenetic tree were clearly separated into two branches (Fig. 5). Similar to the

Fig. 3 Multiple sequence alignment of the deduced amino acid sequences of 1S1x and 1Bx types HMW-GS and five newly cloned subunit genes in this work in *Ae. longissima*. The red box indicates the cysteine residues

Fig. 4 Multiple sequence alignment of the derived amino acid sequences of IS^{ly} and IBy types HMW-GS and five newly cloned subunit genes in this work in *Ae. longissima*. The red box indicates the cysteine residues

Table 3 The positions of SNPs identified in five new x-type HMW-GS genes

HMW-GS	27	96–105	115	155	188	261	272	276	323	330	341	377	493	498	539	574	864	1010	1683	1699	1713	2055
1S ¹ 2x	G	GCGCGAGCTC	G	G	T	G	C	A	G	C	G	C	A	C	A	C	C	A	A	T	A	C
1S ¹ 16x	G	–	A	G	T	G	C	A	G	C	G	C	G	A	G	C	C	A	A	T	A	C
1S ¹ 17x	A	GCGCGAGCTC	G	G	T	C	T	G	G	T	C	A	G	A	A	C	C	C	T	G	G	T
1S ¹ 23x	A	GCGCGAGCTC	G	A	T	C	T	G	A	T	C	C	A	A	A	T	C	C	T	G	G	T
1S ¹ 25x	G	GCGCGAGCTC	G	G	A	G	C	A	G	C	G	C	G	C	A	C	C	A	A	T	A	C
Other 22	G	GCG/ACGAGCTC	G/T/–	G	T/G	G	C	G	G	C	C	C	G/C	A/G	A	C	G/A	C/T	A	G	G	C
HMW-GS Genes ^a																						
HMW-GS	2061	2497	2650	2950	3071	3082–3089	3179	3267	3348	3475	3514	3583										
1S ¹ 2x	T	A	G	G	C	AGGACAAG	C	A	G	C	T	T										
1S ¹ 16x	T	A	G	G	C	–	C	A	A	C	C	T										
1S ¹ 17x	T	G	T	T	A	AGGACAAG	C	C	G	C	C	T										
1S ¹ 23x	T	G	T	T	A	AGGACAAG	T	A	G	C	C	T										
1S ¹ 25x	T	A	G	G	C	AGGACAAG	C	A	G	T	T	T										
Other 22	A	G	G/A	A/G	C/T	AGGACAAG	C/G	A	G	C	C	C										
HMW-GS Genes ^a																						

^aOther 22 HMW-GS genes were from GenBank, including DQ478571 (*T. aestivum*), X13927 (*T. aestivum*), X61009 (*T. aestivum*), DQ533690 (*T. aestivum*), M22208 (*T. aestivum*), JF938073 (*T. aestivum*), AB263219 (*T. aestivum*), AY553933 (*T. aestivum*), AY367771 (*T. aestivum*), EF540764 (*T. aestivum*), AJ437000 (*T. turgidum* ssp. *durum*), AY455789 (*Ae. comosa*), AY455786 (*Ae. utaristata*), AY611727 (*Ae. bicornis*), HM116373 (*T. timopheevii* ssp.), GQ403043 (*Ae. longissima*), AY611723 (*Ae. searsii*), AF476959 (*Ae. markgrafii*), AF476961 (*Ae. umbellulata*), HQ380225 (*Ae. speltooides*), HQ380224 (*Ae. speltooides*), and AF216868 (*T. Aestivum*)

Table 4 The positions of SNPs identified in five new y-type HMW-GS genes

HMW-GS	36	38	56	156	165	213	274	368	426	478	933	937	1011	1124	1162	1258	1297	1348	1355–1364
1S'2y	A	C	G	C	T	T	G	C	C	G	C	G	C	C	C	C	A	A	GGCACTACCC
1S'6y ₁	A	C	C	C	T	T	A	T	A	G	G	C	A	A	A	A	C	C	–
1S'16y	A	C	C	A	T	C	G	C	C	G	C	G	G	C	C	C	A	A	GGCACTACCC
1S'17y	A	C	C	C	T	T	A	C	C	A	C	G	G	A	C	C	A	A	GGCACTACCC
1S'23y	A	C	C	C	T	T	A	T	A	G	G	C	G	A	A	A	C	C	–
Other 18 HMW-GS Genes ^a	C	T	C	C	G	T	G	C	C	C/G	G	C	G/T	A	C	C	C	C/G/–	G/–G/C/–CATTACAC
HMW-GS	1377	1380	1385	1386	1387	1388	1389	1390	1395	2006	2025	2312	2329	2343	2579–2591	2669	2741		
1S'2y	G	A	G	A	C	A	A	G	A	A	–	C	C	G	–	C	T		
1S'6y ₁	C	C	C	T	T	C	T	C	G	T	C	A	A	A	GCAGCAATCAGG ACAAGA	A	C		
1S'16y	G	A	C	A	C	A	A	G	A	A	–	C	C	A	GCAGCAATCAGG ACAAGA	C	T		
1S'17y	G	A	C	A	C	A	A	G	A	A	T	A	C	A	GCAGCAATCAGG ACAAGA	A	C		
1S'23y	C	C	C	T	T	C	T	C	G	T	T	A	C	A	GCAGCAATCAGG ACAAGA	A	C		
Other 18 HMW-GS Genes ^a	G/A	A/G	G/A/–	A/–	C/–	A/–	A/–	G/A/T/–	A/–	A	T/A/G/–	C/–	C	A/–	GCAGCAATCAGG ACAAGA	C	C		

^aOther 18 HMW-GS genes were from GenBank, including DQ000162 (*T. aestivum*), X61026 (*T. aestivum*), EF540765 (*T. aestivum*), AY245797 (*T. turgidum* ssp. *durum*), AF513640 (*Ae. speltooides*), HQ380223 (*Ae. speltooides*), HQ380222 (*Ae. speltooides*), GQ403044 (*Ae. longissima*), HQ380229 (*Ae. kotschy*), HQ380227 (*Ae. kotschy*), HQ380228 (*Ae. kotschy*), AF476960 (*Ae. Markgrafii*), AF476962 (*Ae. umbellulata*), AY611728 (*Ae. Bicornis*), AY611724 (*Ae. searsii*), AY455788 (*Ae. comosa*), and AY455787 (*Ae. unitaristata*)

Table 5 HMW-GS from *Ae. longissima* L. identified by MALDI-TOF/TOF-MS

Types	HMW-GS	Tryptic fragments identified by MALDI-TOF/TOF-MS	Positions	Matched peptides	Protein score CI %
x-type	1S ¹ 17x	ACQQVMDQQLR GGSSYPGDTTPLQQLQSQIFWGIPALLKR YYPTSPQQPGQWQQPEQGQPR	42–52 80–108 150–170	3	241
x-type	1S ¹ 25x	ACQQVMDQQLR DGSFYPGETTPPQQLQSQIFWGIPALLKR YYPSVSSPQQVSYYPGQASPQR- PGQGQQPGQGQQSR HRQQGYPTSPQQPGQWQQPEQGQPR AQLLAAQLPAMCR	42–52 80–170 915–927	3	150
y-type	1S ¹ 6y ₁	QVVDQQLAGRLPWSTGLQMR CRPVAVSQVAR	45–64 76–86	2	98
y-type	1S ¹ 17y	QVVDQQLAGRLPWSTGLQMR CRPVAVSQVARQYEQTMLVLPK QGSYYPGQASPQQLGQGQQPGK WQKPGQGQQGYPTSLQPGQGQQI GK GQQGHYPASLQPGQGQPGQR QGYDSPYHVSAEQQAAR AQQPTTQLTTVCR	45–64 76–96 136–184 563–583 723–739 745–757	6	340

previous report (Wang et al. 2013), the x-type and y-type subunit genes from S¹ genome showed closer relationships with 1Dx2 and 1Dx5 and 1Dy10 and 1Dy12 subunit genes from bread wheat, respectively.

As shown in Tables S1 and S2, the calculated divergence times between five x-type and y-type HMW-GS genes isolated from *Ae. longissima* and those from A, B and D genomes of *Triticum* species ranged from 2.10 to 10.00 million years ago (MYA), from *Aegilops* species ranged from 1.50 to 7.48 MYA, and from *Ae. longissima* ranged from 0.99 to 6.51 MYA.

Discussion

It is well known that HMW glutenin subunits play an important role in the structure and properties of gluten, as well as the wheat processing quality. Considerable work demonstrated that the variations at *Glu-1* loci in common hexaploid wheat are limited (Li et al. 2007), and only a few of HMW-GS genes have positive effects on gluten quality, although numerous subunits are identified (Payne et al. 1981). However, higher diversity in HMW-GS compositions in the wheat-related species is widely present (Ciaffi et al. 1993; Yan et al. 2003b). The *Aegilops* genus has rich storage protein gene resources (Liu et al. 2003; Yan et al. 2008), and particularly, some potential superior HMW-GS from *Ae. longissima* were identified (Wang et al. 2013; Garg et al. 2014; Zhu et al. 2015). In this work, we identified and cloned ten new HMW-GS genes in *Ae. longissima*, and they showed

special molecular structural features with higher molecular weight and more glutamine residues that may confer superior gluten quality. Thus, the S¹ genome has extensive allelic variations at *Glu-1* loci and contains potential gene resources for wheat quality improvement.

In theory, high glutamine content can stabilize the polymeric structure of glutenin by forming more hydrogen bonds (Bekes et al. 1995), and therefore, the newly identified HMW-GS rich in glutamines may have a positive effect on dough strength. Meanwhile, these subunits contained more repetitive domains, leading to a higher molecular mass. In general, longer and more regular repeating units have a positive effect on the elasticity and viscosity of the dough through intermolecular interactions (Masci et al. 2000). The insertion of the intermediate repeat region fragments directly affects the function of glutenins (Hassani et al. 2005). Compared with other HMW-GS, 1S¹2x, 1S¹16x, and 1S¹25x have 84 and 20 aa insertions at positions 669–752 and 778–797, respectively (Table 3), and both 1S¹6y₁ and 1S¹23y have one 15 aa insertion at the position 454–468 (Table 4), significantly increasing the length of the repeat region. In addition, the consensus of tripeptide and hexapeptide in repetitive domains has effects on wheat quality (Masci et al. 2000). Three new x-type subunits (1S¹2x, 1S¹16x, and 1S¹25x) have more tripeptide motifs in repetitive domain than 1Bx7 and 1S¹x2.3*, while five y-type subunits had much more nonapeptide repeats than 1S¹y16* and 1By8 (Table 2). D'Ovidio and Anderson (1994) reported that y-type HMW-GS plays the most important role in various major

Fig. 5 The phylogenetic tree of 30 HMW-GS based on complete amino acid sequences. 1S¹2x* and 1S¹16y* is cloned from the S genome of Chinese Spring chromosome substitution line CS-1S1/1B, in which the 1B chromosome was substituted by 1S¹ from *Ae. longissima*

factors affecting flour technology. Thus, the molecular characterization of these special HMW-GS may improve the formation of gluten macropolymers and breadmaking quality, and have potential application values during wheat quality breeding.

Molecular markers based on SNPs have shown to be a fast and effectively for identifying HMW-GS genes in wheat breeding (Rafalski 2002). To date, various molecular markers for HMW-GS genes have been developed, which provide a powerful tool for wheat quality improvement (Ma et al. 2003; Liang et al. 2015). The HMW-GS genes identified and cloned from *Ae. longissima* showed abundant SNP and InDel variations (Tables 3, 4), which could facilitate to develop molecular markers used for marker-assisted selection in the early generations during quality improvement programs. HMW-GS genes can be successfully transferred to wheat by transgenic technology (Liu et al. 2011). Recently, wheat genetic transformation has made a greater progress (Wang et al. 2017), which can greatly shorten the breeding period. Therefore, the newly cloned special HMW-GS genes in this

study are expected to be used for wheat quality improvement by combining with effective genetic transformation and marker-assisted selection.

Conclusion

In this study, ten special HMW-GS in *Ae. longissima* were identified and their complete encoding genes were cloned by AS-PCR. Molecular characterization showed that all subunits contained a longer central repetitive domain with more glutamine repeats and glutamine residues than those from common wheat, which could benefit the formation of superior breadmaking quality. The encoding sequences of these new genes had abundant SNP and InDel variations that could facilitate to develop specific molecular markers for gluten quality improvement. The 1S¹-encoded HMW-GS showed closer relationships with those at *Glu-D1* locus of bread wheat as revealed by phylogenetic analysis, and they diverged from *Triticum* species at about 2.10–10.00 MYA.

These novel HMW-GS genes have potential application values for breadmaking quality improvement of wheat through molecular marker development and marker-assisted selection as well as direct genetic transformation.

Acknowledgements This research was financially supported by Grants from National Key R & D Program of China (2016YFD0100502) and the National Natural Science Foundation of China (31771773).

References

- Alberto C, Renato D, Antonio TO, Carla C, Marina P, Enrico P (2003) Genetic analysis of the *Aegilops longissima*, 3S chromosome carrying the *Pm13*, resistance gene. *Euphytica* 130:177–183
- Allaby RG, Banerjee M, Brown TA (1999) Evolution of the high molecular weight glutenin loci of the A, B, D, and G genomes of wheat. *Genome* 42:296–307
- Bekes F, Gras PW, Gupta RB (1995) The effects of purified cereal polypeptides on the mixing properties of dough. In: Williams YA, Wrigley CW (eds) Capturing the benefits of research for consumers. R. Aust. Chem. Inst, Cereal Chem Div, pp 92–98
- Ciaffi M, Lafiandra D, Porceddu E, Benedettelli S (1993) Storageprotein variation in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) from Jordan and Turkey. I. Electrophoretic characterization of genotypes. *Theor Appl Genet* 86:474–480
- D'Ovidio R, Anderson OD (1994) PCR analysis to distinguish between alleles of a member of a multigene family correlated with wheat bread-making quality. *Theor Appl Genet* 88:759–763
- Garg M, Kumar R, Singh RP, Tsujimoto H (2014) Development of an *Aegilops longissima*, substitution line with improved bread-making quality. *J Cereal Sci* 60:389–396
- 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* 93:10274–10279
- Ge P, Ma C, Wang S, Gao L, Li X, Guo G, Ma W, Yan Y (2012) Comparative proteomic analysis of grain development in two spring wheat varieties under drought stress. *Anal Bioanal Chem* 402:1297–1313
- Guo H, Wu J, Lu Y, Yan Y (2019) High molecular weight glutenin 1Bx17 and 1By18 subunits encoded by *Glu-B1i* enhance wheat dough rheological properties and breadmaking quality. *J Food Quality* 2019:8
- Hassani ME, Gianibelli MC, Shariflou MR, Sharp PJ (2005) Molecular structure of a novel y-type HMW glutenin subunit gene present in *Triticum tauschii*. *Euphytica* 141:191–198
- Jiang QT, Wei YM, Wang F, Wang JR, Yan ZH, Zheng YL (2009) Characterization and comparative analysis of HMW glutenin 1Ay alleles with differential expressions. *BMC Plant Biol* 9:16
- Jiang P, Xue J, Duan L, Gu Y, Mu J, Han S, Chen L, Li Y, Ma W, Yan Y, Li X (2019) Effects of high-molecular-weight glutenin subunit combination in common wheat on the quality of crumb structure. *J Sci Food Agr* 99:1501–1508
- Lawrence GJ, Shepherd KW (1981) Chromosomal location of genes controlling seed protein in species related to wheat. *Theor Appl Genet* 59:25–31
- Li XH, Zhang YZ, Gao LY, Dong YM, Ji KM, He ZH, Appels RD, 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
- Liang XN, Zhen SM, Han CX, Wang C, Li XH, Ma WJ, Yan YM (2015) Molecular characterization and marker development for hexaploid wheat specific HMW glutenin subunit 1By18 gene. *Mol Breed* 35:221
- Liu ZJ, Yan ZH, Wan YF, Liu KF, Zhang YF, Wang DW (2003) Analysis of HMW glutenin subunits and their coding sequences in two diploid *Aegilops* species. *Theor Appl Genet* 106:1368–1378
- Liu X, Jin W, Liu J, Zhao H, Guo A (2011) Transformation of wheat with the HMW-GS 1Bx14 gene without markers. *Genetika* 47:206
- Ma W, Zhang W, Gale KR (2003) Multiplex-PCR typing of high molecular weight glutenin subunit alleles in wheat. *Euphytica* 134:51–60
- Ma CY, Yang Y, Li XH, Ge P, Guo GF, Subburaj S, Zeller FJ, Hsam SLK, Yan YM (2013) Molecular cloning and characterization of six novel HMW-GS genes from *Aegilops speltoides* and *Aegilops kotschy*. *Plant Breed* 132:284–289
- Mackie AM, Lagudah ES, Sharp PJ, Lafiandra D (1996) Molecular and biochemical characterization of HMW glutenin subunits from *T. tauschii* and the D genome of hexaploid wheat. *J Cereal Sci* 23:213–225
- Masci S, D'Ovidio R, Lafiandra D, Kasarda DD (2000) A 1B coded low-molecular-weight glutenin subunit associated with quality in durum wheats show strong similarity to subunits present in some bread wheat cultivars. *Theor Appl Genet* 100:396–400
- Neelam K, Rawat N, Tiwari VK, Ghandhi N, Arun PC, Kumar S, Tripathi SK, Randhawa GS, Prasad R, Dhaliwal HS (2013) Development and molecular characterization of wheat-*Aegilops longissima* derivatives with high grain micronutrients. *Aust J Crop Sci* 7:508–514
- 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, Corfield KG, Holt LM, Blackman JA (1981) Correlation between the inheritance of certain high molecular weight subunits of glutenin and bread-making quality in progenies of six crosses of bread wheat. *J Sci Food Agric* 32:51–60
- Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 5:94–100
- Sheng H, Murray TD (2013) Identifying new sources of resistance to eyespot of wheat in *Aegilops longissima*. *Plant Dis* 97:346–353
- Shewry PR, Halford NG, Tatham AS (1992) High molecular weight subunits of wheat glutenin. *J Cereal Sci* 15:105–120
- Shewry PR, Gilbert SM, Savage AW, Tatham AS, Wan YF, Belton PS, Wellner N, D'Ovidio R, Be'ke's F, Halford NG (2003) Sequence and properties of HMW subunit 1Bx20 from pasta wheat (*Triticum durum*) which is associated with poor end use properties. *Theor Appl Genet* 106:744–750
- Wan Y, Wang D, Shewry PR, Halford NG (2002) Isolation and characterization of five novel high-molecular-weight subunit genes from *Triticum timopheevi* and *Aegilops cylindrica*. *Theor Appl Genet* 104:828–839
- Wang SL, Yu ZT, Cao M, Shen XX, Li N, Yan YM (2013) Molecular mechanisms of HMW glutenin subunits from 1S¹ Genome of *Aegilops longissima* positively affecting wheat breadmaking quality. *PLoS ONE* 8:e58947
- Wang K, Liu HY, Du LP, Ye XG (2017) Generation of marker-free transgenic hexaploid wheat via an *Agrobacterium*-mediated co-transformation strategy in commercial Chinese wheat varieties. *Plant Biotechnol J* 15:614–623
- Wrigley CW (1996) Biopolymers giant proteins with flour power. *Nature* 381:738–739
- Yan YM, Hsam SLK, Yu JZ, Jiang Y, Zeller FJ (2003a) 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 YM, Hsam SLK, Yu JZ, Jiang Y, Ohtsuka I, Zeller FJ (2003b) HMW and LMW glutenin alleles among putative tetraploid and hexaploid European spelt wheat (*Triticum spelta* L.) progenitors. *Theor Appl Genet* 107:1321–1330
- Yan YM, Zheng J, Xiao Y, Yu J, Hu Y, Cai M, Hsam SLK, Zeller FJ (2004) Identification and molecular characterization of a novel y-type HMW glutenin gene from *Glu-1D¹* locus of *Aegilops tauschii*. *Theor Appl Genet* 108:1349–1358
- Yan ZH, Guo ZF, Liu DC, Dai SF, Wei YM, Zheng YL (2008) Characterization of HMW-GS genes *Dx5^t* and *Dy12^t* from *Aegilops tauschii* accession with subunit combination *Dx5^t*+*Dy12^t*. *Cereal Res Commun* 36:477–487
- Yan YM, Jiang Y, An XL, Pei YH, Li XH, Zhang YZ, Wang AL, He Z, Xia X, Bekes F, Ma W (2009) Cloning, expression and functional analysis of HMW glutenin subunit *1By8* gene from Italy pasta wheat (*Triticum turgidum* L. ssp. *durum*). *J Cereal Sci* 50:398–406
- Yang ZJ, Guang-Rong LI, Liu C, Feng J, Zhou JP, Ren ZL (2006) Molecular characterization of a HMW glutenin subunit allele providing evidence for silencing of x-type gene on *Glu-B1*. *Acta Genet Sin* 33:929–936
- Zhou J, Ma C, Zhen S, Cao M, Zeller FJ, Hsam SLK, Yan Y (2016) Identification of drought stress related proteins from 1S¹(1B) chromosome substitution line of wheat variety Chinese Spring. *Bot Stud* 57:20
- Zhu GR, Wang SL, Zhen SM, Shen XX, Prodanovic S, Yan YM (2015) Molecular characterization and phylogenetic analysis of unusual x-type HMW glutenin subunits from 1S¹ genome of *Aegilops longissima*. *Genetika* 47:185–203