Gene 516 (2013) 1–7

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

# Molecular characterization of two y-type high molecular weight glutenin subunit alleles *1Ay12** and *1Ay8** from cultivated einkorn wheat (*Triticum monococcum* ssp. *monococcum*)

Xiao-Hui Guo [a,b], Bi-Hua Wu [a,b,*], Xi-Gui Hu [a,b], Zhe-Guang Bi [a,b], Zhen-Zhen Wang [a,b], Deng-Cai Liu [a,b,c], You-Liang Zheng [a,b]

[a] Triticeae Research Institute, Sichuan Agricultural University, Wenjiang 611130, PR China
[b] Key Laboratory of Crop Genetic Resources and Improvement, Ministry of Education, Sichuan Agricultural University, Ya'an 625014, PR China
[c] Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Science, Xining 810001, PR China

ARTICLE INFO

*Article history:*
Accepted 4 December 2012
Available online 20 December 2012

*Keywords:*
*Triticum monococcum* ssp. *monococcum*
HMW-GS
*1Ay* gene
Heterogeneous expression
Differentiation

ABSTRACT

Two y-type high molecular weight glutenin subunits (HMW-GSs) 1Ay12* and 1Ay8* from the two accessions PI560720 and PI345186 of cultivated einkorn wheat (*Triticum monococcum* ssp. *monococcum*, AA, 2n = 2x = 14), were identified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The mobility of 1Ay12* and 1Ay8* was similar to that of 1Dy12 and 1By8 from common wheat Chinese Spring, respectively. Their ORFs respectively consisted of 1812 bp and 1935 bp, encoding 602 and 643 amino acid residues with the four typical structural domains of HMW-GS including signal peptide, conserved N-, and C-terminal and central repetitive domains. Compared with the most similar active *1Ay* alleles previous published, there were a total of 15 SNPs and 2 InDels in them. Their encoding functions were confirmed by successful heterogeneous expression. The two novel 1Ay alleles were named as *1Ay12** and *1Ay8** with the accession No. JQ318694 and JQ318695 in GenBank, respectively. The two alleles were classed into the two distinct groups, Phe-type and Cys-type, which might be relevant to the differentiation of *Glu-A1-2* alleles. Of which, *1Ay8** belonged to Cys-type group, and its protein possessed an additional conserved cysteine residue in central repetitive region besides the six common ones in N- and C-terminal regions of Phe-type group, and was the second longest in all the known active *1Ay* alleles. These results suggested that the subunit 1Ay8* of cultivated einkorn wheat accession PI345186 might have a potential ability to strengthen the gluten polymer interactions and be a valuable genetic resource for wheat quality improvement.



## 1. Introduction

High molecular weight glutenin subunits are important storage proteins in wheat and its related species (Shewry et al., 1995, 2003b). They are critical components for gluten viscoelastic properties and the end-use quality of wheat flour (Payne, 1987; Shewry et al., 1992). It is known that HMW-GSs in hexaploid common wheat (*Triticum aestivum* ssp. *aestivum*, AABBDD, 2n = 6x = 42) are encoded by three *Glu-1* loci (*Glu-A1*, *Glu-B1* and *Glu-D1*), located on the long arms of homoeologous group 1 chromosomes (Payne, 1987; Shewry et al., 1992). Each locus consists of two tightly linked paralogous genes *Glu-1-1* and *Glu-1-2*, which encode an x-type subunit with higher molecular weight and a y-type subunit with lower molecular weight, respectively (Mackie et al., 1996; Payne, 1987). Theoretically, there should be six HMW-GSs in hexaploid common wheat, but only three to five subunits are observed due to gene silencing (Shewry et al., 2001, 2003b). Of which, 1Ay subunit is always silent (D'Ovidio et al., 1996; Forde et al., 1985; Hu et al., 2012; Xu et al., 2009).

It has been demonstrated that y-type subunits contain more cysteine residues (usually for 7) than x-type ones (usually for 4) (Shewry et al., 2002), being more beneficial to ameliorating flour quality because of the greater polymer resulted from more inter- and intra-molecular disulfide bonds (Shewry et al., 1992, 2002; Wan et al., 2002). It was proposed that the active *1Ay* alleles from other *Triticum* species could be used for improving wheat processing quality and investigating molecular information on expression of *1Ay* gene (Jiang et al., 2009; Waines and Payne, 1987). And, the research by Ciaffi et al. (1991) has proved that the 1Ay subunit from wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*, AABB, 2n = 4x = 28) enhanced gluten properties of durum wheat (*T. turgidum* ssp. *durum*, AABB, 2n = 4x = 28)

*Abbreviations:* aa, amino acid(s); bp, base pair(s); CTAB, cetyltrimethyl ammonium bromide; dNTP, deoxyribonucleoside triphosphate; HMW-GS, high molecular weight glutenin subunit; InDel, insertion and deletion; IPTG, isopropyl β-Δ-thiogalactopyranoside; ORF, open reading frame; PCR, polymerase chain reaction; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SNP, single nucleotide polymorphism.

* Corresponding author at: Triticeae Research Institute, Sichuan Agricultural University, Wenjiang 611130, PR China. Tel./fax: +86 2882650350.
*E-mail address:* wubihua2001@yahoo.com.cn (B.-H. Wu).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.12.037

cultivars. Apparently, it is of significance to search for active *1Ay* gene from the relative species of common wheat. So far, several active *1Ay* alleles had been isolated from *Triticum timopheevii* ssp. *timopheevii* (AAGG, 2n = 4x = 28) (Wan et al., 2002), *Triticum urartu* (AA, 2n = 2x = 14) (Bai et al., 2004; Hu et al., 2010; Jiang et al., 2009), wild einkorn wheat (*Triticum monococcum* ssp. *aegilopoides* AA, 2n = 2x = 14) (Jiang et al., 2009), cultivated emmer wheat (*T. turgidum* ssp. *dicoccum*, AABB, 2n = 4x = 28) (Jiang et al., 2009). However, the information on active *1Ay* alleles in cultivated einkorn wheat, which is one of the earliest cultivated crop domesticated in the Fertile Crescent 10,000 years ago (Zohary and Hopf, 1988) and was considered as a valuable resources for wheat breadmaking quality improvement (Tranquilli et al., 2002), has not been in lack.

Several studies indicated the existence of abundant allelic variations of HMW-GS in cultivated einkorn wheat (Ciaffi et al., 1998; Ma et al., 2007; Waines and Payne, 1987). Our previous study discovered that there is 1Ay subunit for 96.67% in 60 accessions of cultivated einkorn wheat (Hu et al., 2012).

In the present study, two active *1Ay* alleles from cultivated einkorn wheat were isolated to investigate their molecular structural characteristics, and to probe the potential value in quality of common wheat and to analyze the evolution of *Glu-A1-2* alleles.

## 2. Materials and methods

### 2.1. Plant materials and SDS-PAGE

Two accessions of cultivated einkorn wheat, PI560720 and PI345186, were kindly provided by National Plant Germplasm System (NPGS, http://www.ars-grin.gov/npgs/index.html). HMW-GSs from hexaploid common wheat Chinese Spring (null, 1Bx7 + 1By8, 1Dx2 + 1Dy12) were used as the control for HMW-GS identification. HMW-GSs were extracted from a half kernel and isolated by SDS-PAGE based on the republished procedure (Wan et al., 2002). To ensure the experimental accuracy, at least three seeds were analyzed for each accession.

### 2.2. DNA preparation and PCR amplification

Genomic DNA was extracted from young leaves using a modified CTAB method (Doyle and Doyle, 1987). A pair of degenerate primers, P1 (5′-ATGGCTAAGCGGC/TTA/GGTCCTCTTTG-3′) and P2 (5′-CTATCA CTGGCTA/GGCCGACAATGG-3′), were used to amplify the ORFs of *1Ay* alleles. PCR amplifications were performed in a reaction volume of 50 μL containing 5 μL of 10× ExPCR buffer, 0.2 mmol $L^{-1}$ of dNTPs, 1 μmol $L^{-1}$ of each primer, 2.5 U of ExTaq polymerase and $ddH_2O$ to 50 μL. PCR was carried out using a GeneAmp 9700 PCR System thermocycler (Applied Biosystems, USA) with the condition of 95 °C for 5 min, followed by 28 cycles of 94 °C for 40 s and 68 °C for 6 min, and then a final incubation at 72 °C for 10 min. PCR products were separated on 1.0% agarose gel.

### 2.3. Cloning, sequencing and prokaryotic expression of the ORFs of 1Ay alleles

The PCR products of expected size were recovered, purified and further ligated into pMD19-T vector (TaKaRa, Dalian, China). Then the ligation mixtures were transformed into *Escherichia coli* DH10B competent cells. The strategy of primer walking and the nest deletion method were used to obtain the full length sequence of *1Ay* ORFs. The DNA sequencing was performed by Sangon Biotechnology Company (Shanghai, China). The final nucleotide sequence for each *1Ay* ORF was determined from the sequencing results of 3 independent clones. The both mature *1Ay* ORFs without the sequence encoding the signal peptide region, were expressed in *E. coli* according to the procedure described by Hu et al. (2010).

### 2.4. Sequence alignments and phylogenetic analysis

Multiple alignments of DNA and the deduced amino acid sequences were completed by Clustal X 2.0 software (Larkin et al., 2007). For maximizing the similarities among the sequences of central repetitive regions, the alignments were further improved by visual examination and manual adjustment. Based on the result of multiple alignments, the sequence identities between the two cloned and other previously published *1Ay* sequences were calculated by BioEdit 7 software (http://www.mbio.ncsu.edu/bioedit/bioedit.html). The nucleotide sequences consisted of the signal peptide and N-terminal and C-terminal domains were selected for phylogenetic analysis because they had been demonstrated to be phylogenetically informative (Chen et al., 2012; Li et al., 2004). The phylogenetic tree was constructed by the Molecular Evolutionary Genetics Analysis software MEGA 5.0 (Tamura et al., 2011), and the boot-strap analysis was conducted with 1000 replicates.

## 3. Results

### 3.1. SDS-PAGE analysis of HMW-GS

SDS-PAGE analysis (Fig. 1A) showed that both the accessions PI560720 and PI345186 of cultivated einkorn wheat possessed a pair of HMW-GSs, an x-type subunit of lower mobility and a y-type subunit of higher mobility. The electrophoretic mobilities of x-type subunits in the two accessions were identical, similar to 1Dx2 of common wheat, while those of y-type ones were obviously different. The 1Ay subunits from PI560720 and PI345186 migrated similarly to 1Dy12 and 1By8 of common wheat, respectively. Based on their electrophoretic mobility, they were designated as 1Ay12* and 1Ay8*, respectively.

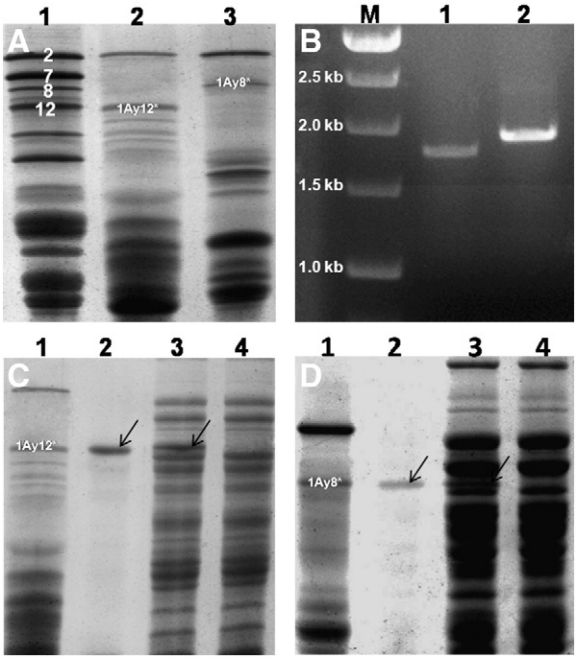


**Fig. 1.** 1Ay subunits, and their encoding gene DNA amplification and heterologous expression of *T. monococcum* ssp. *monococcum*. A: SDS-PAGE profile of the two accessions PI560720 (lane 2) and PI345186 (lane 3) and the reference Chinese Spring (line 1), 1Ay12* and 1Ay8* indicate their 1Ay subunits, respectively; B: the amplified complete ORF of *1Ay12** (lane 1) and *1Ay8** (lane 2) by PCR; C and D: bacterial expression of the modified ORFs of *1Ay12** and *1Ay8** in *E. coli*, respectively, with lane 1 being the HMW-GSs from PI560720 (C) and PI345186 (D), lanes 2 (the selective extract) and 3 (the general extract) being the expressed 1Ay proteins in IPTG-induced bacterial cells, lane 4 being the total proteins in the bacteria cells without IPTG induction for control, and the expressed proteins of *1Ay12** and *1Ay8** were respectively marked by arrow.

### 3.2. Molecular characterization of *1Ay12** and *1Ay8**

There is a single amplified fragment in each of cultivated einkorn wheat (Fig. 1B). The amplified fragments from PI560720 and PI345186 were approximately 1800 bp and 1900 bp in length, close to the most *1Ay* genes in previously published active *Glu-A1-2* genes (Table 1), named as *1Ay12** and *1Ay8**, respectively. Their complete coding sequences were 1812 bp and 1935 bp in length, and encoded 602 and 643 amino acid residues, respectively. The authenticity of *1Ay12** (Fig. 1C) and *1Ay8** (Fig. 1D) alleles was confirmed by successful expression in *E. coli*. The sequences of the two novel *Glu-A1-2* alleles were deposited in GenBank with the accession number JQ318694 for *1Ay12** and JQ318695 for *1Ay8**, respectively.

### 3.3. Comparison of *1Ay12** and *1Ay8** with other Glu-1-2 alleles

Both *1Ay12** (JQ318694) and *1Ay8** (JQ318695) were compared with the 32 *Glu-1-2* alleles previously published in *Triticum* species (Table 1). The sequences of both the *1Ay12** and *1Ay8** had the higher identities (from 84.4% to 98.7%) with those of the 24 *1Ay* alleles (except for EU984508) than those of other 8 y-type genes encoded by B, D and G genomes (from 73.2.4% to 80.5%). Compared with the 24 *1Ay* alleles, *1Ay12** was most similar to the active genes EU984503, FJ404595, AY245578 and AM183223 from *T. urartu* with high identities of 98.7%, 98.6% and 98.5%, whereas *1Ay8** was closest to the two pseudogenes GQ184456 and HQ834309 from cultivated einkorn wheat with the high identities of 98.6% and 98.4%, as well as the pseudogene EU984509 and the active gene EU984506 from wild einkorn wheat with a high identity of 97.5%. And, the sequence of *1Ay8** was longer than that of EU984506 with 1899 bp but shorter than that of EU984508 with an unusual length of 2202 bp from wild einkorn wheat, so it was the second longest in all the known active *1Ay* alleles (Table 1).

The coding sequences of *1Ay12** and *1Ay8** were aligned with that of the most similar *1Ay* active allele, respectively. Compared with EU984503, there was one 18 bp deletion at the position from 804 to 805 nucleotide (nt), encoding a hexapeptide repetitive unit (PGQGQQ) in *1Ay12** (Table 2A). And, in comparison with EU984506, one 36 bp insertion at the position from 877 to 912 nt encoding two tandem hexapeptide repetitive units (PGQGQQPGQGQQ) occurred in *1Ay8** (Table 2B). Simultaneously, 4 and 11 SNPs occurred in *1Ay12** and *1Ay8**, respectively (Table 2). The 15 SNPs included 11 transitions (A-G and C-T), and 4 transversions (A-T, C-G and G-T) (Table 2), of which, 9 were of non-synonymous mutation which produced amino acid residue substitutions. In comparison with *1Ay12** (Table 2A) only at 236 for T→A (valine→aspartic acid), *1Ay8** (Table 2B) possessed more amino acid variations including at 142 for A→G (asparagine→aspartic acid), at 868 for T→G (tryptophan→glycine), at 934 for G→A (glycine→arginine), at 1199 for G→A (arginine→glutamine), at 1832 for T→A (leucine→glutamine), at 1863 for C→G (histidine→glutamine), at 1868 for T→C (leucine→proline), and at 1877 for G→A (proline→alanine).

### 3.4. Primary structures of *1Ay12** and *1Ay8**

1Ay12* and 1Ay8* possessed similar primary structure to other previously reported y-type HMW-GSs, containing four domains: a signal peptide with 21 aa, a conserved N-terminal domain with 104 aa, a conserved C-terminal domain with 42 aa, and a central repetitive domain (Table 3). Both the 1Ay12* and 1Ay8* possessed the six conserved cysteine residues, including five at the N-terminal domain and one at the C-terminal domain. In addition, 1Ay8* subunit possessed an additional cysteine residue at the end of central repetitive domain, which is usually absent in most 1Ay subunits with the exception of two 1Ay subunits expressed by EU984506 and EU984508, and four unexpressed 1Ay of EU984509, GQ184456, HQ834309 and AY260548 from 2 diploid and 1

**Table 1**
Identities of sequences between both the *1Ay12** (JQ318694) and *1Ay8** (JQ318695) and 32 other y-type HMW-GS genes from *Triticum* species.

| HMW-GS alleles | GenBank accession | Species | Genome | The size of ORF (bp) | Identity | |
|---|---|---|---|---|---|---|
| | | | | | JQ318694 | JQ318695 |
| 10 *1Ay* active genes | AM183223 | *T. urartu* | AA | 1830 | 98.5% | 88.4% |
| | AY245578 | *T. urartu* | AA | 1830 | 98.6% | 88.5% |
| | FJ404595 | *T. urartu* | AA | 1830 | 98.6% | 88.5% |
| | EU984503 | *T. urartu* | AA | 1830 | 98.7% | 88.5% |
| | EU984504 | *T. urartu* | AA | 1767 | 96.1% | 88.2% |
| | EU984506 | *T. monococcum* ssp. *aegilopoides* | AA | 1899 | 89.2% | 97.5% |
| | EU984507 | *T. monococcum* ssp. *aegilopoides* | AA | 1767 | 96.0% | 88.1% |
| | EU984508 | *T. monococcum* ssp. *aegilopoides* | AA | 2202 | 73.8% | 73.3% |
| | EU984511 | *T. turgidum* ssp. *dicoccum* | AABB | 1767 | 96.0% | 88.1% |
| | AJ306977 | *T. timopheevii* ssp. *timopheevii* | AAGG | 1767 | 95.7% | 88.2% |
| 14 *1Ay* pseudogenes | EU984505 | *T. urartu* | AA | 1920 | 93.7% | 84.5% |
| | AY245579 | *T. urartu* | AA | 1920 | 93.5% | 84.4% |
| | HQ834309 | *T. monococcum* ssp. *monococcum* | AA | 1953 | 86.5% | 98.4% |
| | GQ184456 | *T. monococcum* ssp. *monococcum* | AA | 1953 | 86.7% | 98.6% |
| | EU984509 | *T. monococcum* ssp. *aegilopoides* | AA | 1899 | 89.1% | 97.5% |
| | AY260548 | *T. turgidum* ssp. *dicoccum* | AABB | 1830 | 97.1% | 88.1% |
| | EU984510 | *T. turgidum* ssp. *dicoccum* | AABB | 1830 | 96.9% | 87.9% |
| | AY260549 | *T. turgidum* ssp. *dicoccum* | AABB | 1830 | 96.6% | 87.6% |
| | AY494981 | *T. turgidum* ssp. *turgidum* | AABB | 1830 | 96.8% | 87.8% |
| | AY722710 | *T. turgidum* ssp. *polonicum* | AABB | 1812 | 96.1% | 87.2% |
| | AY303766 | *T. aestivum* ssp. *spelta* | AABBDD | 1830 | 96.9% | 87.9% |
| | DQ537335 | *T. aestivum* ssp. *aestivum* | AABBDD | 1812 | 96.2% | 87.3% |
| | X03042 | *T. aestivum* ssp. *aestivum* | AABBDD | 1812 | 96.2% | 87.3% |
| | HM131807 | *T. timopheevii* ssp. *araraticum* | AAGG | 1866 | 95.1% | 86.2% |
| *1Gy7** | EF151424 | *T. timopheevii* ssp. *timopheevii* | AAGG | 2202 | 73.7% | 73.2% |
| *1Gy7* | HM131806 | *T. timopheevii* ssp. *araraticum* | AAGG | 2202 | 74.1% | 73.5% |
| *1By8* | AY245797 | *T. aestivum* ssp. *durum* | AABB | 2166 | 76.1% | 76.6% |
| *1By9* | X61026 | *T. aestivum* ssp. *aestivum* | AABBDD | 2121 | 77.8% | 76.7% |
| *1By15* | DQ086215 | *T. aestivum* ssp. *aestivum* | AABBDD | 2175 | 75.9% | 76.3% |
| *1By16* | EF540765 | *T. aestivum* ssp. *aestivum* | AABBDD | 2220 | 74.1% | 74.6% |
| *1Dy10* | X12929 | *T. aestivum* ssp. *aestivum* | AABBDD | 1950 | 78.0% | 79.6% |
| *1Dy12* | X03041 | *T. aestivum* ssp. *aestivum* | AABBDD | 1986 | 78.9% | 80.5% |

**Table 2**
SNP and InDel analysis of *1Ay12** (A) and *1Ay8** (B) compared with their most similar active *1Ay* alleles previously published.

A

| Sequence | Position | | | | |
|---|---|---|---|---|---|
| | 236 | 804–805 | 1491 | 1572 | 1800 |
| *1Ay12** | T (D) | ----------(------) | A | T | C |
| EU984503 | A (V) | CCAGGACAAGGGCAACAA (PGQGQQ) | G | C | T |

B

| Sequence | Position | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 142 | 753 | 868 | 877–912 | 934 | 1002 | 1199 | 1719 | 1832 | 1863 | 1868 | 1877 |
| *1Ay8** | G (D) | G | G (G) | CCAGGACAAGGGCAACAACCAGGACAAGGGCAACAA (PGQGQQPGQGQQ) | A (R) | A | A (R) | A | A (Q) | G (Q) | C (P) | A (Q) |
| EU984506 | A (N) | A | T (W) | ----------(----------) | G (G) | G | G (Q) | G | T (L) | C (H) | T (L) | G (R) |

Note: EU984503 and EU984506 are of the active *1Ay* alleles from *T. urartu* and *T. monococcum* ssp. *aegilopoides*, respectively. The InDels were marked by the dashes "-". The letters and "-" in parentheses are the substitutions and deletions of amino acids resulted from the corresponding SNPs and InDels, respectively.

tetraploid species, but present in the y-type subunits encoded by B, D and G genomes of *Triticum* species (Table 3).

In order to further investigate the distribution of the additional cysteine residue in central repetitive region of *1Ay* alleles, alignment of amino acid sequences of 34 y-type HMW-GSs was conducted (Fig. 2). As the additional cysteine residue of HQ834309, GQ184456, EU984506, EU984508 and EU984509 from wild and cultivated einkorn wheat, that of 1Ay8* (JQ318695) was strictly conserved and located in central repetitive region at the position −73 towards C-terminal same as those of the y-type HMW-GSs encoded by B, D and G genomes, different from the additional cysteine residue located in the position −40 of the pseudogenes AY260548 from wild emmer wheat (Fig. 2). But, in 1Ay12*, this conserved cysteine residue was always replaced by a phenylalanine residue, as in AY260548 and the other 18 *1Ay* active or pseudogenes (Fig. 2). Based on this, the 26 *Glu-A1-2* alleles (*1Ay* active or pseudogenes) were divided into two groups, i.e. Cys-type group, which possessed the additional conserved cysteine residue in the central repetitive region, and Phe-type group, in which the additional conserved cysteine residue was replaced by phenylalanine residue (Fig. 2). Comparing the coding sequences of all these y-type HMW-GSs, the additional conserved cysteine residues from both all the 6 *1Ay* alleles of Cys-type group and the y-type HMW-GS alleles from B, D and G genomes were always encoded by the triplet codon TGC, while the corresponding phenylalanine residues of all the 20 *1Ay* alleles of Phe-type group always by the triplet codon TTC (Fig. 3).

**Table 3**
Comparison of primary structures of 1Ay12* and 1Ay8* with 32 other y-type subunits.

| HMW-GS subunits | GenBank accession | Number of amino acid residues | | | | | Number of cysteine residues | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | SP | NT | CR | CT | Total | NT | CR | CT | Total |
| 1Ay12* | JQ318694 | 21 | 104 | 435 | 42 | 602 | 5 | 0 | 1 | 6 |
| 1Ay8* | JQ318695 | 21 | 104 | 476 | 42 | 643 | 5 | 1 | 1 | 7 |
| 10 1Ay subunits | AM183223 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | AY245578 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | FJ404595 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | AJ306977 | 21 | 104 | 420 | 42 | 587 | 5 | 0 | 1 | 6 |
| | EU984503 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | EU984504 | 21 | 104 | 420 | 42 | 587 | 5 | 0 | 1 | 6 |
| | EU984506 | 21 | 104 | 464 | 42 | 631 | 5 | 1 | 1 | 7 |
| | EU984507 | 21 | 104 | 420 | 42 | 587 | 5 | 0 | 1 | 6 |
| | EU984508 | 21 | 104 | 565 | 42 | 732 | 5 | 1 | 1 | 7 |
| | EU984511 | 21 | 104 | 420 | 42 | 587 | 5 | 0 | 1 | 6 |
| 14 1Ay amino acid sequences[a] | EU984505 | 21 | 104 | 471 | 42 | 638 | 5 | 0 | 1 | 6 |
| | AY245579 | 21 | 104 | 471 | 42 | 638 | 5 | 0 | 1 | 6 |
| | HQ834309 | 21 | 104 | 482 | 42 | 649 | 5 | 1 | 1 | 7 |
| | GQ184456 | 21 | 104 | 482 | 42 | 649 | 5 | 1 | 1 | 7 |
| | EU984509 | 21 | 104 | 464 | 42 | 631 | 5 | 1 | 1 | 7 |
| | AY260548 | 21 | 104 | 441 | 42 | 608 | 5 | 1 | 1 | 7 |
| | EU984510 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | AY260549 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | AY494981 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | AY722710 | 21 | 104 | 435 | 42 | 602 | 5 | 0 | 1 | 6 |
| | AY303766 | 21 | 104 | 441 | 42 | 608 | 5 | 0 | 1 | 6 |
| | DQ537335 | 21 | 104 | 435 | 42 | 602 | 5 | 0 | 1 | 6 |
| | X03042 | 21 | 104 | 435 | 42 | 602 | 5 | 0 | 1 | 6 |
| | HM131807 | 21 | 104 | 453 | 42 | 620 | 5 | 0 | 1 | 6 |
| 1Gy7* | EF151424 | 21 | 104 | 565 | 42 | 732 | 5 | 1 | 1 | 7 |
| 1Gy7 | HM131806 | 21 | 104 | 565 | 42 | 732 | 5 | 1 | 1 | 7 |
| 1By8 | AY245797 | 21 | 104 | 553 | 42 | 720 | 5 | 1 | 1 | 7 |
| 1By9 | X61026 | 21 | 104 | 538 | 42 | 705 | 5 | 1 | 1 | 7 |
| 1By15 | DQ086215 | 21 | 104 | 556 | 42 | 723 | 5 | 1 | 1 | 7 |
| 1By16 | EF540765 | 21 | 104 | 571 | 42 | 738 | 5 | 1 | 1 | 7 |
| 1Dy10 | X12929 | 21 | 104 | 481 | 42 | 648 | 5 | 1 | 1 | 7 |
| 1Dy12 | X03041 | 21 | 104 | 493 | 42 | 660 | 5 | 1 | 1 | 7 |

SP: Signal peptide, NT: N-terminal domain, CR: central repetitive region, CT: C-terminal domain.
[a] 14 amino acid sequences deduced from the corresponding *1Ay* pseudogenes in which the in-frame stop codons and frame-shift mutations were ignored.

### 3.5. Phylogenetic analysis

To further analyze the evolutionary relationship between the two groups of *Glu-A1-2* alleles, a neighbor-joining tree was constructed by using the 26 *Glu-A1-2* alleles and 9 other HMW-GS genes reported (Fig. 4). The result showed all the *Glu-A1-2* alleles were clustered together and separated from the clades of *Glu-D1-2*, *Glu-B1-2*, *Glu-G1-2* alleles, except the *1Ay* gene EU984508 of *T. monococcum* ssp. *aegilopoides*, which was clustered with the *Glu-G1-2* clade of *1Gy7** from *Triticum timopheevii* ssp. *timopheevii* and *1Gy7* from *T. timopheevii* ssp. *araraticum* but not with any group of the *1Ay* alleles (Fig. 4). In the clade of *Glu-A1-2* alleles, there were clearly two subclades with a very high bootstrap value of 100, corresponding well to Phe-type and Cys-type groups of *Glu-A1-2* alleles, respectively (Fig. 4).

## 4. Discussion

SDS-PAGE, is a kind of analytical methods to separate components of a protein mixture based on molecular weight size (Shapiro et al., 1967; Weber and Osborn, 1969), which has been widely used to identify allelic variation of the HMW-GSs from common wheat and the relatives (Branlard et al., 1989; Hu et al., 2012; Waines and Payne, 1987; Wan et al., 2002). However, the mobility is sometimes disaccordant with the molecular weight size. For example, the subunits 1Bx13 with 774 aa, 1By16 with 717 aa, 1Dx5 with 827 aa, and 1Dyl2 with 639 aa are larger, than the allelic subunits 1Bx7 with 768 aa, 1By15 with 702 aa, 1Dx2 with 817 aa, 1Dyl0 with 627 aa, respectively, but migrated faster on SDS-PAGE (Pang and Zhang, 2008; Shewry et al., 1992). Similarly, the subunit 1By16* with 696 aa is much smaller, in comparison with the allelic subunit 1By16 with 717 aa, but possessed a similar electrophoretic mobility (Jin et al., 2012). In this work, the

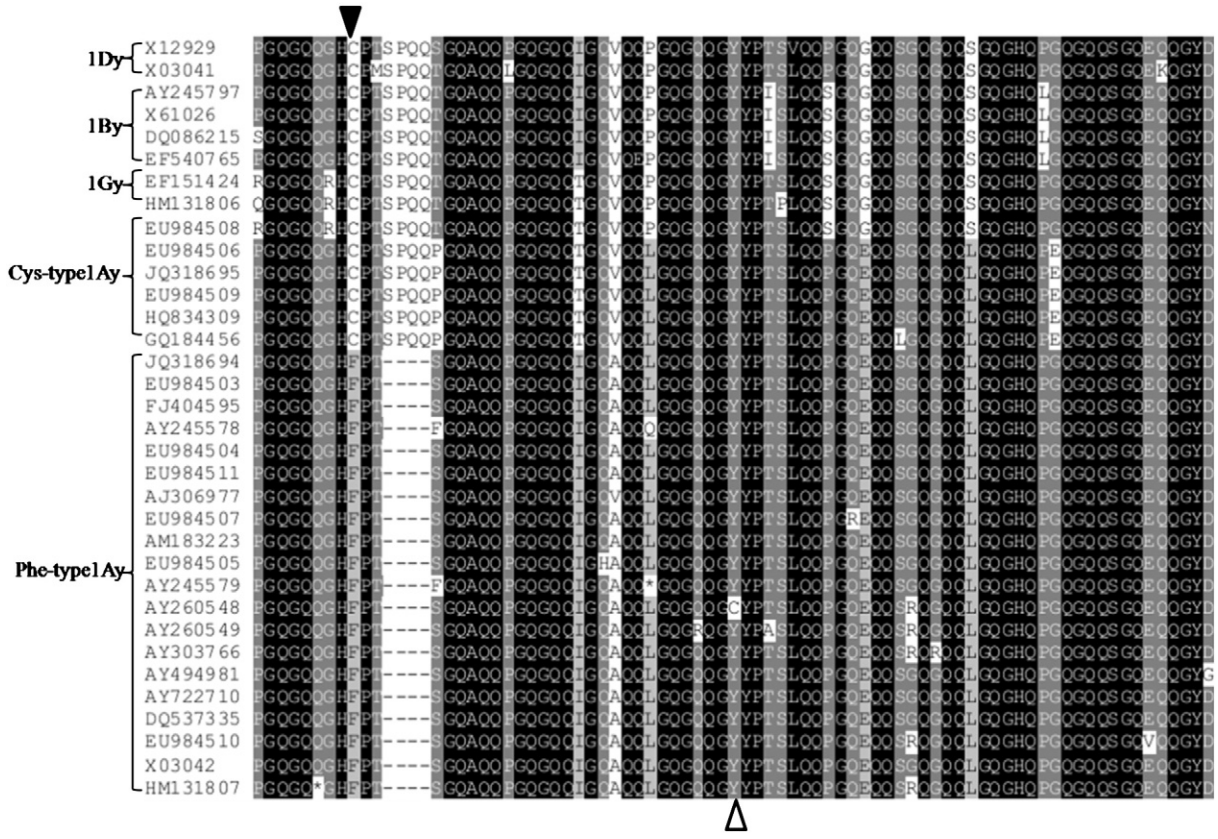


**Fig. 2.** Alignment of amino acid sequences of the partial central repetitive domain next the C-terminal among 34 y-type HMW-GSs. The black triangle marked the position of the additional conserved cysteine residues (C) in position −73 of central repetitive domain toward C-terminal of 6 Cys-type *1Ay* alleles and 8 y-type HMW-GSs from B, D, and G genomes, where the substitutions of cysteine residues with phenylalanine residues (F) occurred in 20 Phe-type *1Ay* alleles. The hollow triangle marked the position of the additional cysteine residue in the central repetitive domain of the Phe-type allele AY260548. The in-frame stop codons were represented by asterisks.

subunits 1Ay8* with 622aa and 1Ay12* with 581 aa of cultivated einkorn wheat accessions PI560720 and PI345186 were clearly different from 1By8 with 699 aa and 1Dy12 with 639 aa of common wheat Chinese Spring in size, respectively, but the mobilities between 1Ay8* and 1By8, as well as between 1Ay12* and 1Dy12 on SDS-PAGE are very similar (Fig. 1A), which was similar to the result of the two 1Ay subunits EU984506 and EU984503 reported by Jiang et al. (2009). Shewry et al. (1992) proposed that different stabilities to denaturation by detergent SDS between the subunits (1Dx2 and 1Dx5, 1Dyl0 and 1Dy12), which arose from their conformational difference, could account for the anomalous relative mobilities. Pang and Zhang (2008) also suggested that the different spatial structure and different charge might be responsible for the faster migration of the subunits 1Bx13, 1By16. So, the anomalous migration between 1Ay12* and 1Dy12 as well as between 1Ay8* and 1By8 might be associated with their different spatial structures or charge or the both interaction resulting from their sequence difference.

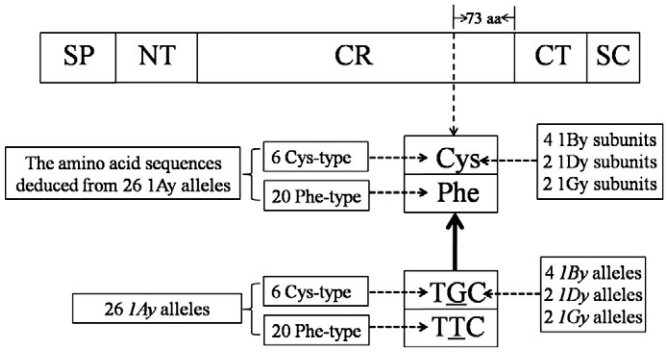


**Fig. 3.** Comparison of the triplet codons encoding the conserved cysteine residues and phenylalanine residues at position −73 of the central repetitive domains toward C-terminal in 34 y-type HMW-GS alleles. SP: Signal peptide domain, NT: N-terminal domain, T: Central repetitive domain, CT: C-terminal domain, SC: Stop codons.

The gene encoding 1Ay type HMW-GS is considered as always silent in common wheat (D'Ovidio et al., 1996; Forde et al., 1985; Hu et al., 2012; Xu et al., 2009). In this research, two functional *1Ay* alleles encoding 1Ay12* and 1Ay8* subunits were successfully isolated (Figs. 1A and B). And, the authenticity of the cloned *1Ay* genes *1Ay12** and *1Ay8** was confirmed by heterogeneous expression in *E. coli* (Figs. 1C and D). Their encoding proteins shared the typical characteristic structure of HMW-GS (Table 3) which contains each of signal peptide, N-terminal, C-terminal and central repetitive regions (Shewry et al., 1995). They were different from the previously published active *1Ay* alleles due to some nucleotide substitutions as well as the length of central repetitive domain resulted from insertion/deletion. Compared with the most similar active *1Ay* alleles previously published, a total of 15 SNPs and 2 InDels were identified in *1Ay12** and *1Ay8** (Table 2). Especially, the polypeptide of *1Ay8** possessed not only the six common cysteine residues in N- and C-terminal regions like in 1Ay12*, but also the additional conserved cysteine residue at position −73 in the central repetitive region towards the C-terminal of y-type subunits encoded by the related B, D and G genomes, while not done by A genome (Li et al., 2007; Shewry et al., 2002). Interestingly, this subunit 1Ay8* is quite similar to the two 1Ay ones of EU984506 and EU984508 containing the additional cysteine residue in wild einkorn wheat (Jiang et al., 2009), which are all with longer central repetitive region than all the other expressed 1Ay subunits without this cysteine residue (Table 3). Of them, 1Ay8* subunit is the second longest (Table 3).

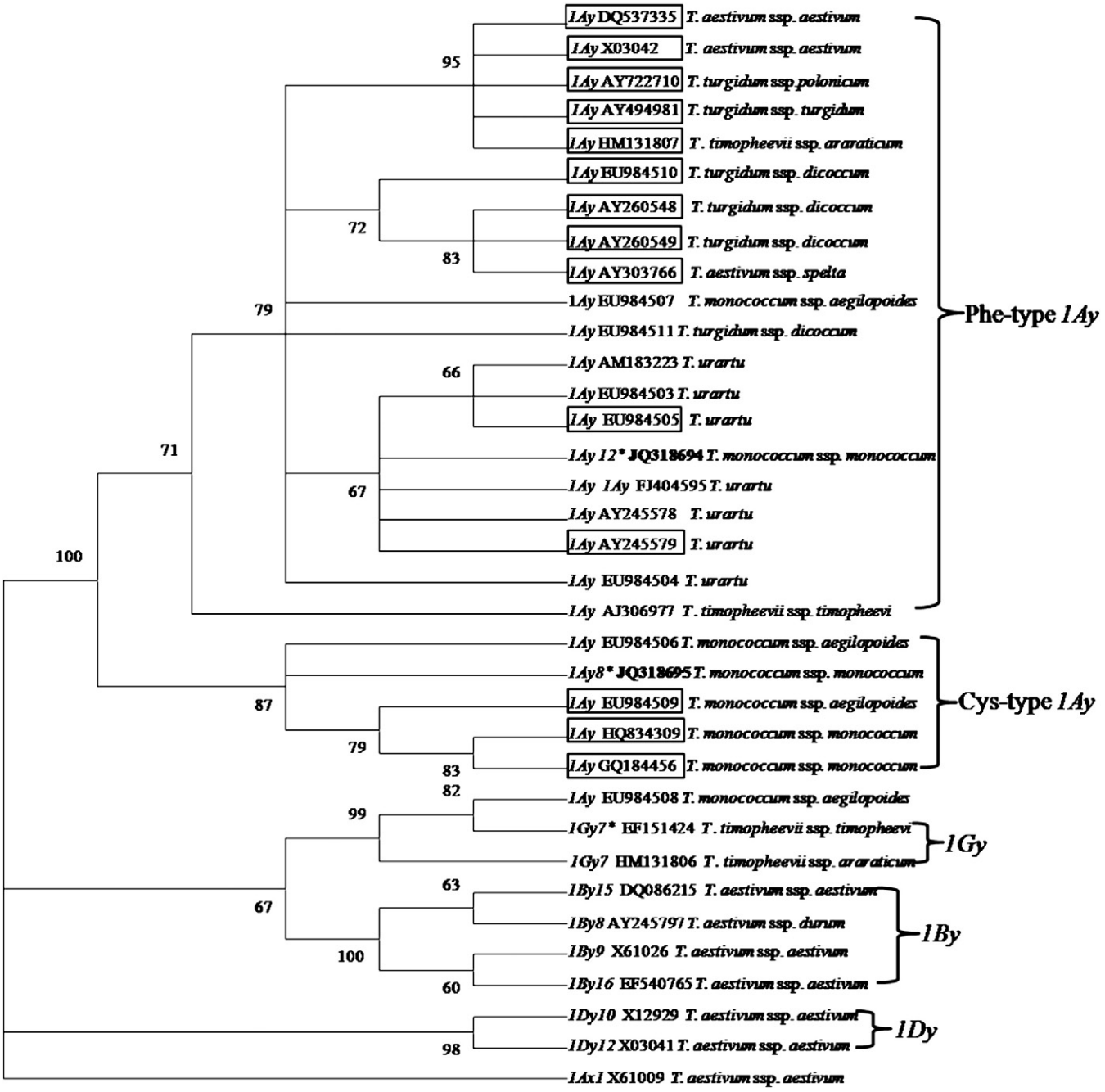


**Fig. 4.** Phylogenetic relationship between twenty Phe-type and six Cys-type *Glu-A1-2* alleles. The neighbor-joining tree was constructed with the non-repetitive regions of the 35 Glu-1 alleles. *1Ax1* (X61009) was used as outgroup. 14 *1Ay* pseudogenes were marked by box.

Cysteine residue can be encoded by both the triplet codons TGC and TGT, and phenylalanine residue by both the TTC and TTT. However, in this study, the additional conserved cysteine residues of all the *1Ay* alleles of Cys-type group and other y-type HMW-GS alleles encoded by genomes B, D and G were always encoded by the codon TGC but never by TGT, and the corresponding phenylalanine residues of all the *1Ay* alleles of Phe-type group always by the codon TTC but never by TTT (Fig. 3). It suggested that the second nucleotide residue transversion for G-T between TGC and TTC codons might be responsible for that the conserved cysteine residue of Cys-type group was always replaced by a phenylalanine residue of the Phe-type group in *1Ay* alleles. Based on that the two conserved cysteine residues in the N-terminal domain existed in other 1Bx subunits were replaced by two tyrosine residues in 1Bx14 and 1Bx20, the two subunits were considered as a novel subclass different from other 1Bx ones, which might be associated with the differentiation of *Glu-B1-1* alleles (Li et al., 2004; Shewry et al., 2003a). In this study, of all the 26 known *Glu-A1-2* alleles (including 12 functional and 14 pseudogenes), 6 (such as *1Ay8**, HQ834309, etc.) contained the additional conserved cysteine residue in position −73 of central repetitive domain toward C-terminal like the y-type subunit genes of other *Glu-1* loci (*Glu-B1*, *Glu-D1* and *Glu-G1*), whereas this cysteine residue in the other 20 alleles was always replaced by phenylalanine residue (such as *1Ay12**, EU984504, etc.). The neighbor-joining tree based on 35 HMW-GS genes showed that all the 25 *Glu-A1-2* alleles (except for EU984508) were clearly clustered into two separate subclades, corresponding well to Phe-type and Cys-type groups of *Glu-A1-2* alleles, respectively (Figs. 2 and 4). Comparison result discovered that all the six *1Ay* alleles with this additional cysteine residue were from cultivated and wild einkorn wheat while nothing from *T. urartu* of diploid with AA

genome, tetraploid with AABB/AAGG genomes, and hexaploid with AABBDD/AAAAGG genomes. It is well known that cultivated einkorn wheat is domesticated from wild einkorn wheat (Zohary and Hopf, 1988). This is agreement with our previous study result, in which cultivated and wild einkorn wheat were with alike higher frequencies and more identical alleles of 1Ay subunit, and both them are relatively distant from the species *T. urartu* (Hu et al., 2012) which is widely considered as the A genome progenitor of tetraploid and hexaploid wheats (Dvořák et al., 1993). These results suggested that the Phe-type and Cys-type groups might involve in the differentiation of *Glu-A1-2* alleles.

Both the conserved cysteine residues in central repetitive region and the size of the repetitive domains contribute to the high order structure of HMW-GSs, the former is involved in the formation of inter- or intra-molecular disulfide bonds, the latter may promote intermolecular interactions through hydrogen bonding (Anjum et al., 2007; Shewry and Tatham, 1997; Shewry et al., 2002; Veraverbeke and Delcour, 2002). The positive relationship of the additional cysteine residue in the central repetitive region or the longer repetitive region of HMW-GSs on dough strength has been reported (Belton, 1999; Feeney et al., 2003; Juhász et al., 2003; Lafiandra et al., 1993; Payne et al., 1981). For example, Lafiandra et al. (1993) and Juhász et al. (2003) considered that the presence of an extra cysteine residue in the central repetitive region of the subunits 1Dx5 and 1Ax2$^{*B}$, facilitating the formation of further disulfide bonds, might lead to an improvement in gluten quality characters. Consequently, the subunit 1Ay8$^{*}$ in cultivated einkorn wheat accession PI345186 might also have a potential ability to strengthen the gluten polymer interactions, and be a valuable genetic resource for wheat quality improvement.

It is an interesting phenomenon that the *1Ay* allele EU984508 of wild einkorn wheat was clustered into the subclade containing two *1Gy* alleles but not into any group of the *1Ay* alleles. In fact, the *1Ay* allele EU984508 possessed much higher sequence identity with *1Gy7*$^{*}$ (99.4%) and *1Gy7* (99.3%) as well as identical length with them, compared with other 25 *1Ay* alleles (from 66.2% to 68.8%) (Supplementary Table 1), which might be responsible for the phenomenon. Of course, it is needed to further investigate the relationship between EU984508 and the two *1Gy* alleles.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2012.12.037.

## Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (No. 30571139 and No. 30671271), as well as by both the Personal Training Foundation and the Education Committee Accented Term in Sichuan Province.

## References

Anjum, F.M., Khan, M.R., Din, A., Saeed, M., Pasha, I., Arshad, M.U., 2007. Wheat gluten: high molecular weight glutenin subunits — structure, genetics, and relation to dough elasticity. J. Food Sci. 72, R56–R63.
Bai, J.R., Jia, X., Liu, K.F., Wang, D.W., 2004. Cloning and characterization of the coding sequences of the *1Ay* high molecular weight glutenin subunit genes from *Triticum urartu*. Acta Bot. Sin. 46, 463–471.
Belton, P.S., 1999. On the elasticity of wheat gluten. J. Cereal Sci. 29, 103–107.
Branlard, G., Autran, J.C., Monneveux, P., 1989. High molecular weight glutenin subunit in durum wheat (*T. durum*). Theor. Appl. Genet. 78, 353–358.
Chen, W.J., et al., 2012. Novel HMW glutenin genes from *Aegilops tauschii* and their unique structures. Genes Genom. 34, 339–343.
Ciaffi, M., Benedettelli, S., Giorgi, B., Porceddu, E., Lafiandra, D., 1991. Seed storage proteins of *Triticum turgidum* ssp. *dicoccoides* and their effect on the technological quality in durum wheat. Plant Breed. 107, 309–319.
Ciaffi, M., Dominici, L., Lafiandra, D., 1998. High molecular weight glutenin subunit variation in wild and cultivated einkorn wheats (*Triticum* spp., *Poaceae*). Plant Syst. Evol. 209, 123–137.
D'Ovidio, R., Masci, S., Porceddu, E., 1996. Sequence analysis of the 5′ non-coding regions of active and inactive *1Ay* HMW glutenin genes from wild and cultivated wheats. Plant Sci. 114, 61–69.
Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.
Dvořák, J., Terlizzi, P.d., Zhang, H.B., Resta, P., 1993. The evolution of polyploid wheats: identification of the A genome donor species. Genome 36, 21–31.
Feeney, K.A., et al., 2003. Molecular structures and interactions of repetitive peptides based on wheat glutenin subunits depend on chain length. Biopolymers 72, 123–131.
Forde, J., et al., 1985. The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.). Nucleic Acids Res. 13, 6817–6832.
Hu, X.G., Wu, B.H., Yan, Z.H., Liu, D.C., Wei, Y.M., Zheng, Y.L., 2010. Characterization of a novel *1Ay* gene and its expression protein in *Triticum urartu*. Agric. Sci. China 9, 1543–1552.
Hu, X.G., et al., 2012. Allelic variation and distribution of HMW glutenin subunit 1Ay in *Triticum* species. Genet. Resour. Crop Evol. 59, 491–497.
Jiang, Q.T., Wei, Y.M., Wang, F., Wang, J.R., Yan, Z.H., Zheng, Y.L., 2009. Characterization and comparative analysis of HMW glutenin *1Ay* alleles with differential expressions. BMC Plant Biol. 9, 16.
Jin, M., et al., 2012. Identification and molecular characterisation of HMW glutenin subunit *1By16*$^{*}$ in wild emmer. J. Appl. Genet. 53, 249–258.
Juhász, A., et al., 2003. Bankuti 1201-an old Hungarian wheat variety with special storage protein composition. Theor. Appl. Genet. 107, 697–704.
Lafiandra, D., D'Ovidio, R., Porceddu, E., Margiotta, B., Colaprico, G., 1993. New data supporting high Mr glutenin subunit 5 as the determinant of quality differences among the pairs 5 + 10 vs. 2 + 12. J. Cereal Sci 18, 197–205.
Larkin, M.A., et al., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Li, W., et al., 2004. Molecular characterization of HMW glutenin subunit allele *1Bx14*: further insights into the evolution of *Glu-B1-1* alleles in wheat and related species. Theor. Appl. Genet. 109, 1093–1104.
Li, X., et al., 2007. Molecular cloning, heterologous expression, and phylogenetic analysis of a novel y-type HMW glutenin subunit gene from the G genome of *Triticum timopheevi*. Genome 50, 1130–1140.
Ma, Z.C., Wei, Y.M., Yan, Z.H., Zheng, Y.L., 2007. Genetic variations of gliadin and high-molecular-weight glutenin subunits in diploid wheats. Plant Genet. Resour. Newsl. 150, 10–15.
Mackie, A.M., Sharp, P.J., Lagudah, E.S., 1996. The nucleotide and derived amino acid sequence of a HMW glutenin gene from *Triticum tauschii* and comparison with those from the D genome of bread wheat. J. Cereal Sci. 24, 73–78.
Pang, B.S., Zhang, X.Y., 2008. Isolation and molecular characterization of high molecular weight glutenin subunit genes *1Bx13* and *1By16* from hexaploid wheat. J. Integr. Plant Biol. 50, 329–337.
Payne, P.I., 1987. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annu. Rev. Plant Physiol. 38, 141–153.
Payne, P.I., Corfield, K.G., Holt, L.M., Blackman, J.A., 1981. Correlations 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.
Shapiro, A.L., Vinuela, E., Maizel Jr., J.V., 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochem. Biophys. Res. Commun. 28, 815–820.
Shewry, P.R., Tatham, A.S., 1997. Disulphide bonds in wheat gluten proteins. J. Cereal Sci. 25, 207–227.
Shewry, P.R., Halford, N.G., Tatham, A.S., 1992. High molecular weight subunits of wheat glutenin. J. Cereal Sci. 15, 105–120.
Shewry, P.R., Napier, J.A., Tatham, A.S., 1995. Seed storage proteins: structures and biosynthesis. Plant Cell 7, 945–956.
Shewry, P.R., Popineau, Y., Lafiandra, D., Belton, P., 2001. Wheat glutenin subunits and dough elasticity: findings of the EUROWHEAT project. Trends Food Sci. Technol. 11, 433–441.
Shewry, P.R., Halford, N.G., Belton, P.S., Tatham, A.S., 2002. The structure and properties of gluten: an elastic protein from wheat grain. Philos. Trans. R. Soc. Lond. B Biol. Sci. 357, 133–142.
Shewry, P.R., et al., 2003a. 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.
Shewry, P.R., Halford, N.G., Lafiandra, D., 2003b. Genetics of wheat gluten proteins. Adv. Genet. 49, 111–184.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Tranquilli, G., et al., 2002. Effect of *Triticum monococcum* glutenin loci on cookie making quality and on predictive tests for bread making quality. J. Cereal Sci. 36, 9–18.
Veraverbeke, W.S., Delcour, J.A., 2002. Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality. Crit. Rev. Food Sci. Nutr. 42, 179–208.
Waines, J.G., Payne, P.I., 1987. Electrophoretic analysis of the high-molecular-weight glutenin subunits of *Triticum monococcum*, *T. urartu*, and the A genome of bread wheat (*T. aestivum*). Theor. Appl. Genet. 74, 71–76.
Wan, Y., Wang, D., Shewry, P.R., Halford, N.G., 2002. Isolation and characterization of five novel high molecular weight subunit of glutenin genes from *Triticum timopheevi* and *Aegilops cylindrica*. Theor. Appl. Genet. 104, 828–839.
Weber, K., Osborn, M., 1969. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. J. Biol. Chem. 244, 4406–4412.
Xu, L.L., Li, W., Wei, Y.M., Zheng, Y.L., 2009. Genetic diversity of HMW glutenin subunits in diploid, tetraploid and hexaploid *Triticum* species. Genet. Resour. Crop Evol. 56, 377–391.
Zohary, D., Hopf, M., 1988. Domestication of plants in the old world: the origin and spread of cultivated plants in west Asia, Europe, and the Nile Valley. Clarendon Press, Oxford.