

Novel Insights into the Composition and Characterization of the Glutenin Genes in Common Wheat Xinmai 26

Dale Zhang¹⁺, Jiajia Ren¹⁺, Wenjie Li¹, Xusen Kang¹, Shenglong Bai¹, Lijuan Yang²,
Zhikai Jiang^{2*}, Suoping Li^{1*}

1. Key Laboratory of Plant Stress Biology, State Key Laboratory of Cotton Biology, School of Life Sciences, Henan University, Kaifeng 475001, China.
2. Xinxiang Academy of Agricultural Sciences, Xinxiang 453003, China.

+ Equal contributors

* Corresponding authors: wheat_jiang@126.com (Z. Jiang), lisuoping@henu.edu.cn (S. Li)

Abstract: High molecular weight glutenin subunits (HMW-GS) and Low molecular weight glutenin subunits (LMW-GS) in mature grains play important roles in the formation of glutenin macropolymer and gluten quality. To characterize the expressed glutenin genes of the bread wheat variety Xinmai 26 during seed development, in this study, we measured the dough rheological properties of mature grains through farinograph and gluten testing system, and revealed its strong gluten quality. The compositions of HMW-GS and LMW-GS were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Further, a total of 18 full-length transcripts were obtained by third-generation RNA sequencing, including 5 transcripts of HMW-GS genes and 13 transcripts of LMW-GS genes (8 intact genes and 5 pseudogenes). The deduced protein structures of transcript sequences exhibit the typical HMW-GS and LMW-GS structural characteristics. Moreover, a specific functional marker was developed to make better use of the extra cysteine residue of 1Dx5 subunit. This study provides an efficient method to accurately identify glutenin genes in bread wheat through matching full-length transcripts to their spectrum of glutenin, which will be helpful to understand the contributions of glutenin genes to

wheat quality.

Keywords: HMW-GS; LMW-GS; MALDI-TOF-MS; Third-generation RNA sequencing; Rheological parameters

1. Introduction

As one of the three main cereal crops worldwide, common wheat (*Triticum aestivum* L., $2n = 6x = 42$, AABBDD) provides basic staple food and important protein source for a large proportion of the world population. The dough generated by mixing wheat flour with water can be used to make various food products such as bread, cake, noodle and pasta. The unique physical properties of wheat flour are mainly determined by seed storage proteins that composed of glutenins and gliadins [1]. Gliadins are monomeric proteins divided into α/β , γ , and ω gliadins depending on their mobility in acid-polyacrylamide gel electrophoresis [2]. By contrast, glutenins are polymeric proteins separated based on their molecular weight into high-molecular-weight glutenin subunit (HMW-GS, 60-90 KDa) and low-molecular-weight glutenin subunit (LMW-GS, 20-45 KDa) [3]. It is generally thought that glutenins play a major role in conferring elasticity through the formation of glutenin macropolymers (GMPs) via intra- and intermolecular disulfide bonding [4]. While gliadins are responsible for dough extensibility by non-covalent interactions with GMPs [5]. However, an increasing number of studies have indicated that no adverse effects of dough properties are found by the reduction of gliadins in gluten, but even improved end-use quality parameters [6-11]. Recently Li et al [12] reported that the mutant line lacking *Gli-D2* (DL*GliD2*) expectedly revealed better flour milling quality and dough properties comparing with its progenitor Xy81, without penalties in other agronomic traits. It seems that the quality of wheat food products is mainly influenced by glutenins.

The HMW-GS represents only 10% of gluten proteins, and determines 50-70% of the end-use quality of the wheat grain [13]. The subunits 1 or 2* at *Glu-A1*, 7 + 8, 17 + 18, 13 + 16, and 14 + 15 at *Glu-B1*, and 5 + 10 at *Glu-D1* are known as desirable subunits for bread quality [14]. The contribution of the *Glu-D1* locus to processing

quality is the highest in comparison with *Glu-A1* locus and *Glu-B1* locus [15]. HMW-GS are widely selected and exploited in wheat quality breeding programs since they can be easily identified by SDS-PAGE and molecular markers [16]. The genes encoding HMW-GS are located at the *Glu-A1*, *Glu-B1* and *Glu-D1* loci on the long arms of the group 1 chromosomes, each locus containing two tightly linked genes encoding one x- and one y-type subunit [17]. Only three to five HMW-GSs could be observed in the bread wheat cultivars as a consequence of allelic variations and gene silence [18]. The structure of HMW-GS is generally composed of a large central repetitive domain flanked by short non-repetitive N and C-terminal domains. The quantity and distribution of cysteine residues in the three domains are of great significance for the wheat flour milling quality, since disulfide bond between two cysteine residues can alter the polymer structure and protein conformation [19]. Especially, the serine (118th amino acid) residue in the repetition region of 1Dx5 subunit is replaced by cysteine [20], which is considered superior for wheat quality. The extra cysteine residue is thought to be associated with higher elasticity, better baking and noodle processing quality.

LMW-GS accounts for about 50% of gluten proteins and contributes to 30% of the technological quality. Compared with few HMW-GSs, LMW-GSs are encoded by a multigene family located at the *Glu-A3*, *Glu-B3* and *Glu-D3* loci on the short arms of chromosomes 1A, 1B and 1D, respectively [21]. Gupta and Shepherd [22] identified a total of 20 protein alleles in 222 wheat varieties by SDS-PAGE, including six alleles (*a*, *b*, *c*, *d*, *e*, and *f*) at *Glu-A3* locus, nine alleles (*a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, and *i*) at *Glu-B3* locus, and five alleles (*a*, *b*, *c*, *d*, and *e*) at *Glu-D3* locus. Moreover, at least 15 LMW-GS genes could be determined from individual MCC (micro-core collections) accessions utilizing the LMW-GS gene molecular marker system, of which 4-6 were located at the *Glu-A3* locus, 3-5 at the *Glu-B3* locus, and eight at the *Glu-D3* locus [23]. The mature structure of a typical LMW-GS generally consists of a relatively short N-terminal domain, a central repetitive domain, and a C-terminal domain. Based on the first amino acid residue of N-terminal domain, LMW-GSs could be grouped into three types: LMW-

m, LMW-s and LMW-i (m, s, and i represent methionine, serine and isoleucine, respectively.). Typically, the N-terminal amino acid sequence of the LMW-s type subunits is only SHIPGL-, while LMW-m type subunits contain various N-terminal sequences such as METSHIGPL-, METSRIPGL-, and METSCIPGL- [24]. The LMW-i type subunit starts directly with the repetitive region of ISQQQQ- after the signal peptide due to lacking the N-terminal domain [25]. Most LMW-GSs contain eight cysteine residues, although their positions vary in the three types of subunits, and these cysteine residues play crucial role in the formation of intra- and inter-molecular disulfide bonds in the gluten macropolymer [26].

It is undeniable that the allelic variation of the LMW-GS is very difficult to be distinguished by SDS-PAGE due to their complexity, heterogeneity and co-migration with gliadins [27]. Therefore, a variety of new techniques have been developed to identify the allelic variation of LMW-GS, including two-dimensional electrophoresis (2-DE), ultraperformance liquid chromatography (UPLC), reversed-phase high-performance liquid chromatography (RP-HPLC), and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) [28,29]. MALDI-TOF-MS is currently considered to be the most efficient method to characterize wheat storage proteins due to its accuracy and sensitivity, and requires relatively small samples and only 4-5 minutes per sample for analysis [30]. In regard to cloning of glutenin genes, the sequences could be acquired through PCR amplification from genomic DNA with Sanger sequencing and BAC sequencing as previous studies [31,32]. With the improvement of genome sequencing technology, the Pacific Bioscience (PacBio) third-generation sequencing platform can offer high throughput of 50,000-70,000 reads per reaction and a read length over 3 kb to ensure assembly of genes with repetitive sequences [3,33]. Although the inherent sequencing errors are inevitable in this approach, the errors are random and can be corrected by redundancy. Moreover, the SMRT (single molecule real-time) cell template structure enables the polymerase to loop multiple times on a single molecule. Therefore, the obtained CCS (circular consensus sequences) could achieve improved accuracy [34]. This technology

is efficient and economical in generating sequences from gene families and has been applied to wheat gluten studies [35].

The bread wheat variety Xinmai 26, with high protein content and strong gluten strength, is widely cultivated in the Huanghuai region. Meanwhile, several new varieties with similar high quality have been bred utilizing the accession as one of the parents. However, the characterization of expressed glutenin genes of this variety, especially LMW-GS genes, are still unknown during seed development. In this study, the compositions of HMW-GS and LMW-GS in the bread wheat Xinmai 26 are identified by SDS-PAGE and MALDI-TOF-MS, and its dough rheological character are measured. Further, the full-length transcript sequences acquired by PacBio Sequel were analyzed and characterized to efficiently identify the mRNAs of transcribed HMW-GS and LMW-GS genes from Xinmai 26. Moreover, a specific functional marker was developed to detect the extra cysteine residue of 1Dx5 subunit. The expressed glutenin genes could be accurately determined through matching full-length transcripts to their spectrum of glutenin, which can provide valuable gene resources for wheat quality breeding.

2. Results

2.1 Dough rheological properties of Xinmai26

Dough rheological properties of Xinmai26 were measured through farinograph with 50 g of flour (Fig. 2, Table 1A) and gluten testing system with 10 g of flour (Table 1B). Water absorption (WA), development time (DT), stability time (ST) and degree of softening (DS) based on the farinograph were determined to be 68.4 %, 23.2 min, 23.4 min, and 6 FU, respectively. As one of the most important indexes in evaluating product price calculations, WA is positively related to the yield of finished bakery product and maintenance of food character. Apparently, the WA value of 68.4 % herein exceeds the upper limit of most bread wheat [48]. Moreover, higher ST and lower DS in the Xinmai26 indicate that the corresponding dough would be able to sustain long mechanical processing treatments, implying that the Xinmai26 has strong gluten elasticity. Additionally, wet gluten content (WGC), dry gluten content (DGC), gluten

index (GI) and gluten moisture absorption (GMA) in Xinmai26 are 35.64 g, 12.59 g, 96.67 % and 64.67%, respectively. GI are directly linked to physicochemical characteristics of flour, correlating with protein strength [49]. The larger GI of Xinmai26 reflects its stronger gluten strength.

2.2 Separation and identification of HMW-GS and LMW-GS

The HMW-GS and LMW-GS of Xinmai26 were separated and identified based on the standard wheat cv. ‘Chinese spring’ by SDS-PAGE (Fig. 3). The HMW-GS encoded by the *Glu-1* loci are determined as 1Ax1, 1Bx7 and 1By8, 1Dx5 and 1Dy10, respectively. A total of 8 protein bands for *Glu-3* loci were identified using the method described by Ibba et al [50], including 1 of the *Glu-A3* locus, 1 of the *Glu-B3* locus, and 6 of the *Glu-D3* locus. The compositions of HMW-GS and LMW-GS were further analyzed by MALDI-TOF-MS. The HMW-GS exhibits molecular masses of 65-90 kDa in MALDI-TOF-MS spectra. The molecular weight of subunit 1Ax1 was close to that of subunit 1Dx5 in the pattern of MALDI-TOF-MS, which resulted in overlap of their mass spectra peaks. Therefore, the peak value of 67,212, 74,813, 82,109, 87,634 Da respectively represent 1Dy10, 1By8, 1Bx7 and 1Dx5/1Ax1 subunits with ascending molecular weights (Fig. 4A). The molecular masses of 25-45 kDa are attributed to LMW-GS in MALDI-TOF-MS spectra. As shown in Fig. 4B, the spectra of LMW subunits consist of complex sets of peaks, which is consistent with the presence of a multigene family located at *Glu-3* loci. Typically, the single peak value of 41,312 Da exhibits the specific peak of *Glu-A3a/c* allele. The *Glu-B3* allele generally contains about 2-4 characteristic peaks [29]. However, only one specific spectrum peak (38,841 Da) was detected in the Xinmai26. The *Glu-D3* alleles were found to be the most complicated in the *Glu-3* loci. Their characteristic peak number for each allele ranges from 1 to 6. A total of 6 peaks (30,866 + 31,576 + 32,172 + 36,902 + 37,212 + 40,353 Da) detected in Xinmai26 were similar with the characteristic peaks of *Glu-D3a/b* allele. In summary, the characteristic peak numbers of the glutenin genes in the MALDI-TOF-MS were consistent with those of protein bands in SDS-PAGE.

2.3 Identification of full-length transcripts of glutenin genes and their corresponding genes.

In order to obtain all the glutenin genes expressed in Xinmai 26, the third-generation transcriptome sequencing was performed through PacBio Sequel II platform on seeds 15 days after flowering, which generated 64,953,298 subreads with mean length 1,566 bp. A total of 1,217,179 reads of circular consensus sequence (CCS) were extracted from the subreads, and 1,150,675 full-length non-chimeric sequences (FLNCs) were acquired after refining by IsoSeq3 pipeline. Due to the highly complex sequence structure of HMW-GS and LMW-GS, the redundant sequences were further removed utilizing a custom pipeline in FLNCs. Finally, 5 transcripts of HMW-GS genes and 13 transcripts of LMW-GS genes were obtained respectively in Xinmai26. To verify the accuracy of transcript sequences of glutenin genes acquired by PacBio Sequel II, the comparison was conducted between the gene sequence of *transcript624* acquired through PCR amplification from genomic DNA with Sanger sequencing and the transcript sequence of that obtained by third-generation sequencing (Fig. S1). In total 11 gene sequences of *transcript624* with PCR amplification were aligned from seven independent clones, amongst two clones were sequenced for three times. Obviously, the sequences of the same clone were completely identical through Sanger sequencing. A few SNPs appeared in gene sequences from different clones, which may be caused by base mismatch in the process of PCR amplification. The final sequence obtained by correcting each other with seven independent clones was identical with the transcript sequence acquired by PacBio Sequel II, indicating that the results with third-generation sequencing are more reliable due to more cycles and stronger error correction ability.

For transcripts of the HMW-GS genes, the sequences of *transcript1854*, *transcript1782*, *transcript272*, *transcript1528* and *transcript624* were demonstrated to be 2496, 2373, 2550, 2166, and 1950 bp in length (Table S2), and encoded 830, 789, 848, 720, and 648 amino acid residues, respectively. The molecular weight calculated from the deduced amino acid sequences of all five transcripts were highly consistent with the peak values of those detected by MALDI-TOF-MS, error ranging from 0.05 %

to 0.54 % (Table 2). High identities were found among the five transcript sequences and those published HMW-GS genes from Triticeae, ranging from 99.96 % to 100 %. Meanwhile, in order to further confirm the phylogenetic relationship between the five transcript sequences and HMW-GS gene family from Triticeae, the phylogenetic tree was constructed based on the deduced amino acid sequences of 93 genes, including *transcript1854*, *transcript1782*, *transcript272*, *transcript1528* and *transcript624* genes obtained in this study and the other 88 HMW-GS genes registered in GenBank (Fig. S2). As clearly shown in this figure, two clades can be classified for the 93 HMW-GS genes: x-type and y-type. The three subgroups from the x-type group and the two subgroups from the y-type group were clearly separated by the Glu-1Ax, Glu-1Bx, Glu-1Dx, Glu-1By and Glu-1Dy. Since the HMW-GS alleles controlled by the same locus were closely clustered in each individual lineage, the 1Ax1, 1Bx7, 1Dx5, 1By8 and 1Dy10 from Xinmai 26 were determined to be encoded by *transcript1854*, *transcript1782*, *transcript272*, *transcript1528* and *transcript624* genes, respectively.

For transcripts of the LMW-GS genes, altogether 13 complete coding sequences were acquired (Table S1), in which eight genes (*transcript4321*, *transcript6479*, *transcript206*, *transcript2357*, *transcript4445*, *transcript1264*, *transcript3598*, and *transcript907*) were found to contain the full-ORF (Open Reading Frame). The other 5 genes (*transcript3234*, *transcript3226*, *transcript3547*, *transcript2435*, *transcript308*) were pseudogenes with at least one stop codon. The sequence alignments indicated that the 13 transcript sequences were completely identical with those published LMW-GS gene sequences from NCBI. Based on the sequence similarity, four genes (*transcript3234*, *transcript3226*, *transcript3547*, *transcript907*) could be attributed to the *Glu-A3* alleles. The transcript sequences of two genes (*transcript3598*, *transcript2435*) were consistent with those of the *Glu-B3* alleles. The other seven genes (*transcript4321*, *transcript6479*, *transcript206*, *transcript2357*, *transcript4445*, *transcript1264*, *transcript308*) were assigned to the *Glu-D3* alleles. Particularly, the nucleotide sequences of *transcript3234* and *transcript308* were completely consistent with those of specific A3-400 (*Glu-A3c*) and D3-589 (*Glu-D3a*) from MCC (micro-

core collections) obtained by the LMW-GS gene molecular marker system [23]. Therefore, the characteristic peaks of *Glu-A3* and *Glu-D3* allele in the MALDI-TOF-MS were respectively determined to be those of *Glu-A3c* and *Glu-D3a*. The specific spectrum peak (38,841 Da) detected in the MALDI-TOF-MS was still uncertain since only two genes (*transcript3598*, *transcript2435*) were identified due to the low expression of a few *Glu-B3* alleles in the mature grain. Further, the eight transcript sequences (*transcript4321*, *transcript6479*, *transcript206*, *transcript2357*, *transcript4445*, *transcript1264*, *transcript3598*, and *transcript907*) with full-ORF were determined to be 1068, 918, 900, 915, 1101, 1059, 1056, and 1134 bp in length, and encoded 354, 304, 298, 303, 365, 351, 350, and 376 amino acid residues, respectively. The molecular weight of the deduced amino acid sequences from the eight genes were highly consistent with the peak values of those detected by MALDI-TOF-MS, with error ranging from 0.18 % to 2.85 % (Table 2). Further, the phylogenetic tree was constructed based on the deduced amino acid sequences of 124 genes, including the eight transcript sequences obtained in this study and the other 116 LMW-GS genes registered in GenBank (Fig. S3). As clearly shown in this figure, the phylogenetic tree was composed of two clear branches, in which all of LMW-i types containing *transcript907* and other LMW-GS genes from *Glu-A3* were clustered in one branch. The LMW-m and LMW-s type subunits composed of other genes from *Glu-B3* and *Glu-D3* were grouped into the other branch since the sequences of them generally display higher consistency. The phylogenetic tree indicated that LMW-i type genes had undergone greater divergence during evolution compared to LMW-s and LMW-m genes [51].

2.4 Molecular characterization of the glutenin genes in the Xinmai26

The deduced protein structures of *transcript1854*, *transcript1782*, *transcript272*, *transcript1528* and *transcript624* exhibited the typical HMW-GS structural characteristics, including a signal peptide of 21 amino acids, a non-repetitive N-terminal domain (A) containing about 80-105 residues, a long repetitive domain (B) of about 480-700 residues, and a non-repetitive domain (C) of 42 amino acids at the C-

terminus. The domain B contained repetitive hexapeptides (unit QQPGQG) as a backbone with inserted hexapeptides (e.g. YYPTSP) and tripeptides (e.g. QQP or QPG). The domains A and C were characterized by the presence of most of the cysteines (Fig. 5). Based on the typical position of cysteines residues [52], the first two residues in the x-type HMW-GS (transcript1854, transcript1782, transcript272) were linked by an intramolecular bond, and thus the others were available for intermolecular linkages. The y-type HMW-GSs (transcript1528 and transcript624) have five cysteines in domain A and one in each of domains B and C. Particularly, compared to the other x-type HMW-GS with four conservative cysteine residues, an extra cysteine residue was found in the N-terminal 118 of the transcript272 (1Dx5) from Xinmai26. The extra cysteine residue is thought to play an important role in the dough elasticity since it might form another intermolecular bond. Based on the first amino acid of the deduced protein sequences of LMW-GS obtained in this study, the eight genes with the full-ORF in Xinmai26 were classified into three types (m-, s- and i-). The transcript6479, transcript206, transcript2357, transcript4445, transcript1264, and transcript3598 were grouped into the m-type. The transcript4321 and transcript907 was the s-type and i-type, respectively. The deduced mature proteins contained three typical conserved domains (N-terminal domain, repetitive domain and C-terminal domain), except for the i-type subunit which lacked the N-terminal domain. All the LMW-GS identified in the Xinmai26 contained eight cysteine residues which could form inter- and intra-chain disulfide bonds (Fig. 5). Among them, the six cysteine residues were proposed to be linked by intrachain disulfide bonds. The other two cysteine residues unique to LMW-GS were located in the N- and C-terminal domain, which could form the interchain disulfide bond. Moreover, the i-type LMW-GS was also different from other genes in molecular weight. The estimated molecular weight of transcript907 was 41.239 kDa, substantially higher than the average molecular weight of other genes because of longer repetitive regions in the i-type subunits.

The secondary structures of the HMW-GS and LMW-GS obtained in this study were predicted by the PSIPRED server, as shown in Table 3. A total of 26 α -helices and

6 β -strands from the HMW-GS only appeared in the N- and/or C-terminal domains, in which 23 of them were found in the N-terminal domains. The other 9 secondary structures were distributed in the C-terminal domains. Compared with the Glu-A1, the subunit combinations of transcript1782 (1Bx7) + transcript1528 (1By8) from Glu-B1 and transcript272 (1Dx5) + transcript624 (1Dy10) from Glu-D1 contained more α -helixes and β -strands, up to 15 and 11, respectively. As for subunit transcript1854 (1Ax1), a total of 59 amino acid residues were involved in the formation of α -helixes (four in N-terminal domain and two in C-terminal domain). In addition, in terms of the number of amino acid residues involved in the formation of secondary structure, the largest number of amino acids (73) came from the transcript1528 subunits (1By8). The secondary structures of the LMW-GS were mainly composed of α -helixes except for a few β -strands. A total of 81 α -helixes were predicted in the LMW-GS, and only 3 β -strands (6 amino acid residues) appeared in the C-terminal domain. In addition, the four α -helixes were also found in the repetitive domain of LMW-GS, and the proportion of amino acid residues participating in secondary structure of LMW-GS was much higher than that of HMW-GS, up to 32.17% - 44.25%. Among the LMW-GS genes, the LMW-i type subunit transcript907 (*Glu-A3c*), located at the 1A chromosome, was established to contain nine α -helixes. It was also found that the number of amino acid residues involved in the formation of its secondary structure were the largest, reaching 137. The *Glu-A3c* of the Aroona NILs was confirmed to play an important role in dough resistance and extensibility [53].

2.5 Specific functional marker for the extra cysteine residue in the 1Dx5 subunit

A specific functional marker was developed to make better use of the extra cysteine residue of transcript272 (1Dx5) subunit from Xinmai26. Though the 1Dx5 subunit could be detected in both Xinmai 26 and Lumai 13 by SDS-PAGE, only the 1Dx5 subunit of Xinmai 26 contained an extra cysteine residue (codon: TGT) by the gene sequencing, while the corresponding site of Lumai 13 was a tyrosine residue (codon: TCT). Thus, the derived cleaved amplified polymorphic sequences (dCAPS) marker was designed based on the single nucleotide polymorphisms (G/C). The PCR

amplification was performed in the two wheat varieties utilizing this specific functional marker, and the PCR products of 95 bp were obtained. After digestion with the enzyme *Sal* I, a band of 76 bp was generated in Xinmai 26 while the product of Lumai 13 was still 95 bp (Fig. S4). To ensure the accuracy and reliability of the dCAPS marker, the amplified products of 95 bp from two wheat varieties were sequenced (Fig. S5). The results showed that only Xinmai 26 contains the recognition site of *Sal* I enzyme. Subsequently, the subunit composition and the extra cysteine residue of *Glu-D1* in 143 wheat varieties was determined by SDS-PAGE and the dCAPS marker (Table S1). The detection of subunit composition and the corresponding extra cysteine residue of *Glu-D1* in the partial wheat varieties were shown in Fig. 6. In the 55 wheat varieties without 1Dx5 subunit, the only 95 bp PCR products were obtained by the dCAPS marker. It is worth noting that the 76 bp products could only be detected in 47 of the 88 wheat varieties containing 1Dx5 subunits using the dCAPS marker, indicating that not all 1Dx5 subunits of bread wheat contain this extra cysteine residue.

3. Discussion

3.1 Advantages of full-length glutenin genes obtained by three-generation transcriptome sequencing

Glutenin genes play a major determinant for the bread-making quality among bread wheat varieties. A large number of HMW-GS genes have been isolated and characterized based on PCR amplification in previous reports [54-58]. For example, the x-type HMW-GS of *Aegilops umbellulata* was determined to carry much longer repetitive domains, which may have potential value in improving the processing properties of hexaploid wheat varieties (Liu et al., 2003). Since HMW-GS contains long repetitive sequences, especially the x-type subunit with high molecular weight, effective primer walking cannot be performed in some regions. Therefore, the full length of many HMW-GS genes was obtained by nested deletion [59]. In addition, LMW-GS genes have also been isolated and characterized by using cDNA or genomic DNA library screening [60,61]. A study based on BAC library screening and proteomics analysis showed that *Glu-A3*, *Glu-B3*, and *Glu-D3* in the Chinese bread wheat cultivar

Xiaoyan 54 contain 4, 3, and 7 genes, respectively [62]. Particularly, the 18, 17 and 17 LMW-GS gene sequences were successfully isolated from Norin 61, Glenlea, and Xiaoyan 54, respectively through the LMW-GS gene marker system [63]. Thus, the LMW-GS gene marker system is considered to be useful and efficient in identifying and characterizing LMW-GS genes in bread wheat [23,50,63,64]. Moreover, this method has advantages over gene-specific PCR and library screening in isolating LMW-GS genes. Recently, with the improved genome sequencing of Chinese Spring with PacBio long reads and utilization of BioNano genome maps to improve and validate the sequence assemblies [33], the complexities of the wheat gluten genomic regions could be better resolved [65]. Through PacBio RSII third-generation RNA sequencing, Wang et al [66] identified 52 full-length gliadin gene transcripts in the bread wheat cultivar Xiaoyan 81. Huo et al [67] constructed high-quality sequences spanning the three wheat homologous α -gliadin loci by aligning PacBio-based sequence contigs with BioNano genome maps. A total of 47 α -gliadin genes were identified, in which 26 genes could encode intact full-length protein products. Zhang et al [35] assembled 424 wheat gliadin and glutenin transcripts from ten wheat cultivars by using just one single-molecule-real-time cell, including 345 proteins and 79 truncated proteins due to early stop codons. In this study, the third-generation transcriptome sequencing was performed through PacBio Sequel II platform on seeds 15 days after flowering, and the 5 transcripts of HMW-GS genes and 13 transcripts of LMW-GS genes were obtained in Xinmai26. The number and molecular weight of HMW-GS and LMW-GS genes obtained by third-generation transcriptome sequencing was consistent with the peak values of those detected by MALDI-TOF-MS. Unexpectedly, only the transcript sequences of two genes (*transcript3598*, *transcript2435*) were acquired at the *Glu-B3* alleles since the expression of a few *Glu-B3* alleles might be low or not expressed in the mature grain of Xinmai26. From this point of view, the *Glu-B3* allele has little influence on flour processing quality in Xinmai26.

Due to the high GC content in glutenin genes, the accuracy of high-fidelity enzyme

remains a challenge during PCR amplification. In this study, the comparison was conducted between the gene sequence of *transcript624* acquired through PCR amplification with Sanger sequencing and the transcript sequence obtained by third-generation sequencing. The result indicates that the sequences of the same clone are completely identical by Sanger sequencing. A few SNPs was found in gene sequences from different clones of the *transcript624*. The final sequence obtained by correcting each other with seven independent clones is identical with the transcript sequence acquired by PacBio Sequel II, showing that the third-generation transcriptome sequencing could improve the sequence accuracy of glutenin genes since the errors are random and can be overcome by redundancy. Therefore, the results obtained by third-generation sequencing are more reliable due to more cycles and stronger error correction ability.

3.2 Composition of glutenin genes in the Xinmai26 and their effects on gluten quality

The strong gluten wheat variety Xinmai 26 has been cultivated in Huanghuai wheat region of China for many years, and the average stability time of the dough can reach 23 minutes (2019-2020 season). New wheat varieties with strong gluten such as Xinmai 38 and Xinmai 45 have been bred using Xinmai 26 as one of parents. It is well known that HMW-GS and LMW-GS in mature seeds play important roles in the formation of glutenin macropolymer and gluten quality, especially for dough extensibility and strength [68]. The effects of HMW-GS on dough properties and pan bread quality have been well studied. According to the contributions of individual HMW-GS to dough strength quality, different subunit combinations at each of the three loci were ranked as: $1Ax1 > 1Ax2^* > 1AxN$ at *Glu-A1*, $1Bx7 + 1By8 \geq 1Bx13 + 1By16 > 1Bx17 + 1By18 = 1Bx7 + 1By9$ at *Glu-B1*, and $1Dx5 + 1Dy10 > 1Dx2 + 1Dy12 > 1Dx4 + 1Dy12$ at *Glu-D1* [69]. The Xinmai26 displays the optimal subunit combinations in this study, which is determined to be 1Ax1, 1Bx7 + 1By8, and 1Dx5 + 1Dy10 based on SDS-PAGE and MALDI-TOF-MS. Moreover, it is generally agreed that the contribution of the *Glu-D1* locus to processing quality is the highest among *Glu-I* loci [15,70].

Especially, 1Dx5 subunit is considered to be a superior subunit for wheat quality due to an extra cysteine residue [20]. The extra cysteine residue facilitates the formation of more intermolecular disulfide bonds, resulting in flexible glutenin macropolymers. Therefore, the subunit combination of 1Dx5 + 1Dy10 is widely used in international wheat quality breeding programs. However, genetic diversity in 1Dx5 is very limited in modern wheat cultivars [20], and not all 1Dx5 contain the extra cysteine residue. In this study, the dCAPS marker based on the single nucleotide polymorphisms (G/C) could produce two types of products after digestion with the enzyme *Sal* I. The only 76 bp band was generated in the 1Dx5 subunit with the extra cysteine residue. A total of 143 wheat varieties from Huanghuai region of China were detected by SDS-PAGE, of which 88 wheat varieties contained 1Dx5 subunit. Unexpectedly, only 47 of the 88 wheat varieties could produce the 76 bp band based on the dCAPS marker. It may be one of the reasons that the flour processing quality of some wheat varieties containing 1Dx5 subunits is not very good due to the absence of this extra cysteine residue. Thus, it is important in wheat quality breeding to select proper 1Dx5 subunit. The 1Ax subunit of Xinmai26 also has positive effect on flour quality. The studies have suggested that the total amount of HMW-GS in grains of wheat varieties containing 1Ax subunit (*1Ax1* or *1Ax2**) increased by 1.5-2.0 % compared with that of wheat varieties containing Null gene [71]. The processing quality of wheat flour can be improved by increasing the ratio of HMW/LMW subunits.

The effects of *Glu-3* alleles and those of *Glu-1* alleles are largely additive, and the interactions between these loci also have significant effects [25]. It is generally accepted that *Glu-A3* and *Glu-B3* alleles play a major role in determining the flour processing qualities among the three *Glu-3* loci, while *Glu-D3* alleles play minor roles in determining quality variation in bread wheat [53]. Particularly, the *Glu-A3* locus was considered to have the biggest contribution to quality among all *Glu-3* loci. In detail, the ranking of alleles for dough strength is *Glu-A3d* > *Glu-A3b* > *Glu-A3c* > *Glu-A3f* > *Glu-A3a* > *Glu-A3e*, whereas ranking for dough extensibility is slightly different, viz., *Glu-A3c* > *Glu-A3b* ≥ *Glu-A3f* > *Glu-A3e*. However, it is controversial concerning the

role of *Glu-A3* alleles due to different kinds of populations or collections. For example, Zhen et al [72] demonstrated that the LMW-GS allele *Glu-A3a* encodes a specific LMW-i type B-subunit that significantly affects wheat dough strength. Cane et al [73] believed that *Glu-A3e* was correlated with inferior dough resistance and extensibility, whereas Zheng et al [74] found that *Glu-A3e* was a favorable allele for dough-mixing properties. In this study, the LMW-i type subunit transcript907 was determined to be *Glu-A3c* based on the molecular weight and MALDI-TOF-MS. The *Glu-A3c* of the Aroona NILs was confirmed to play an important role in dough resistance and extensibility [53].

The secondary structure is the foundation for a highly complex spatial conformation and is composed of α -helices and β -strands in the wheat gluten. Masci et al [75] believed that helix-helix interactions were involved in guiding the formation of the intramolecular disulfide bonds. The higher α -helix content may contribute to better quality of the dough. The β -strands are generally considered to endow the protein with high elasticity and to improve the capability to resist distortion [51]. In this study, the secondary structures occurred mainly in the N- and/or C-terminal domains of HMW-GS, in which 17 α -helices and 6 β -strands were found in the N-terminal domains (Table 3). The only 9 α -helices appeared in the C-terminal domains. This result indicates that the N-terminal domains contribute more to the elasticity and the capability to resist distortion of the dough than the C-terminal domains in the HMW-GS. Interestingly, the secondary structure mainly exists in the form of α -helices in the LMW-GS, which is mainly concentrated in the C-terminal domains. A total of 73 α -helices and 3 β -strands were found in the C-terminal domains and only 4 α -helices appear in the N-terminal domains. Considering the distribution of the cysteine residues in gluten (Fig. 5), the C-terminal domains of HMW-GS may be more favorable for bonding with the N-terminal domains of LMW-GS to form the elastic backbone of the dough.

4. Materials and Methods

4.1 Plant materials

Wheat variety Xinmai 26 was provided by Xinxiang Academy of Agricultural Sciences and preserved in the Plant Germplasm Resources and Genetic Engineering Laboratory, Henan University, which was sown and harvested in the 2019-2020 crop season. Experimental plot was 2 m in length with 3 rows and a row spacing of 30 cm, with 21 seeds per row. Common wheat cultivars ‘Chinese Spring’ (CS) (null, 1Bx7+1By8, 1Dx2+1Dy12) was used as standard wheat in determining HMW-GS and LMW-GS from Xinmai26. The 143 wheat materials detected by specific functional markers were listed in Table S1.

4.2 SDS-PAGE analysis

Glutenin was extracted according to the method described by Wan et al. [36]. Specimens were separated by SDS-PAGE with a discontinuous system of 4 % (w/v) stacking gel and 10 % (w/v) separating gel. 6 µL samples were separately loaded onto lanes of the gel and electrophoresed at approximately 15 °C and 12 mA for 16 h.

4.3 Dough rheological parameters

Seed sample was tempered to constant moisture content (14.0 %) for 12 h and then milled by ALMB roller mill (Bühler Corp., China) according to the AACC 26-50 method [37]. The test samples were performed in duplicate. Dough farinograph parameters were determined through AACC 54-21 method [38]. Water absorption (WA), development time (DT), stability time (ST), and degree of softening (DS) at 500 BU dough consistency were measured by farinograph with 50 g of flour (Brabender Corp., Germany). Wet gluten content (WGC), dry gluten content (DGC), gluten index (GI) and gluten moisture absorption (GMA) were evaluated by a Glutomatic 2200 apparatus (Perten Instruments AB, Huddinge, Sweden) according to AACC 38-12A method [39].

4.4 Glutenin extraction

Glutenin was prepared by the modified method of Zhang et al. [40]. Flour (15 mg) of mature seeds was violently shocked with 70% ethanol for 1 hour, which followed by centrifugation at 10,000 g for 10 min and removed the supernatant. The obtained

sediment was extracted with 1 mL of 55 % (v/v) propanol at 65 °C for 30 min to remove gliadins. After centrifugation at 10,000 g for 10 min, the supernatant was removed. These steps were repeated two times with the resulting pellet to remove gliadin completely. The resulting pellet was then extracted with 150 µL of extraction buffer (50% (v/v) propanol, 0.08 M Tris-HCl (pH 8.0) containing 1% (w/v) dithiothreitol (DTT)) at 65 °C for 30 min followed by centrifugation at 10,000 g for 5 min. Glutenins in the supernatant fraction were precipitated with 100 µL of cold acetone. Alternately, cysteine residues in the glutenins were alkylated by the addition of 150 µL of extraction buffer containing 1.4% 4-vinylpyridine (v/v) to the supernatant fraction and incubation at 65 °C for 15 min. After centrifuging at 10,000 g for 2 min, 200 µL of supernatant was precipitated with 135 µL of cold acetone and stored at -20 °C. The pellets were washed with cold acetone and dried at RT before MALDI-TOF-MS analysis. The glutenin was extracted in three replicates.

4.5 MALDI-TOF-MS analysis

Add 60 µL of 30% acetonitrile (ACN), 0.4 % trifluoroacetic acid (TFA) and 69.6 % H₂O to dissolve the precipitation for 1 hour at room temperature. Sinapinic acid (SA) was used as matrix, which was dissolved in 50% ACN and 0.05% TFA (10 mg/ml). Sample was spotted onto a MALDI-TOF Voyager DE Pro 100 sample size plate by 0.7 µL SA: 0.7 µL sample: 0.7 µL SA. The sample plate was air dried before analysis. MALDI-TOF-MS was performed in the Analytical and Testing Center, School of Life Science, Henan University, China. Glutenin was measured by New ultrafleXtreme™ MALDI-TOF (Bruker Daltonics, Bremen, Germany) equipped with a 60 Hz nitrogen laser. The parameters of the instrument were used with the following settings; mass range: 60,000 to 90,000 Da (HMW-GS), 30,000 to 45,000 Da (LMW-GS), sample rate: 1.00 GS/s, laser shots: 100, laser power: 85%, laser frequency: 80, and detector gain: 13.3X. Spectra were obtained in positive ion mode. All subunits showed good reproducibility with less than 0.05 % relative standard deviation (RSD) in three replicates. The mean molecular weights of three repeats were adopted in this study.

4.6 PacBio RNA sequencing and data analysis

The timing of anthesis and grain development were inspected and recorded after heading. The grain samples were similarly collected at the 15 DAF (days after flowering). Total RNA was extracted using TRIzol reagent (Invitrogen Corp., Carlsbad, CA), according to the manufacturer's protocol. RNA quality was detected on 1% agarose gels. RNA purity was measured using a nanophotometer (Implen, Inc., Westlake Village, CA, USA). RNA concentration was checked by Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, CA, USA). RNA integrity was assessed through the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

The cDNA libraries (1-10 kb) were constructed and sequenced through PacBio Sequel II platform. Iso-seq reads were processed using the IsoSeq3 pipeline (<https://github.com/PacificBiosciences/IsoSeq>) with parameter "`--min-passes 2`". Due to the highly complex sequence structure of the glutenin gene, a custom pipeline was used to remove redundant sequences to acquire HMW-GS and LMW-GS genes in IsoSeq3, respectively (Fig. 1). Briefly, the nucleotide sequences of the published glutenin genes in NCBI were firstly removed redundant sequences utilizing CD-hit software with parameter "`-c 0.9`". The obtained non-redundant sequences were merged with those of the annotated glutenin genes in Chinese Spring to form query sequences. Then, the Blat software [41] was used to search for full-length non-chimeric (FLNC) sequence generated by IsoSeq3 based on the query sequences, in which the sequences with coverage and identity greater than 70% were retained, error-corrected and removed redundant using IsoCon software [42]. The obtained sequences were aligned by Mafft software [43] and manually removed redundant again according to the support score acquired by IsoCon. Finally, the 5 transcript sequences from HMW-GS and 13 transcript sequences from LMW-GS were obtained, respectively.

4.7 DNA extraction and molecular cloning

Genomic DNA of wheat varieties were extracted according to the method reported by Su et al [44]. The cloning and sequencing of 1Dy10 gene from Xinmai 26 was conducted according to the method described by Wang et al [45]. The DNA sequencing was performed by the Beijing Genomics Institution (BGI, China). The nucleotide sequences for the target genes were analyzed from seven independent clones, in which the two clones were sequenced three times by Sanger sequencing.

4.8 Sequence alignment, phylogenetic analysis and secondary structure prediction

The open reading frame (ORF) of the target gene was translated into amino acid sequence using the NCBI ORF Finder program (<http://www.ncbi.nlm.nih.gov>). The alignment of sequences was carried out using the multiple sequence alignment software Clustal X 2.0 software [46]. Based on the multiple alignment of the full-length amino acid sequences (including signal peptide sequences) of HMW-GS and LMW-GS, the phylogenetic trees were respectively constructed by IQtree software using the ultrafast bootstrap approximation method [47]. The bootstrap values in the phylogenetic tree were estimated based on 1000 replications. Prediction of secondary structure of deduced amino acid sequences was carried out by PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>).

4.9 Development of specific functional marker

The derived cleaved amplified polymorphic sequences (dCAPS) marker was developed based on the single nucleotide polymorphisms (G/C) at 353 bp of 1Dx5 subunit from NCBI database. The mismatched bases were introduced into the end of forward primer to form a GTCgaC locus (Lowercase letters represent mismatched bases), which can be recognized by restriction enzyme *Sal* I based on the principle of dCAPS marker. The sequences of specific primers (P1: 5'-GCAGCAACTCCAACAACGTA-3' and P2: 5'-GATAGTATGAAACCTGCTGTCGA-3') were designed by Primer Premier 5.0 (Premier Biosoft International, Palo, Alto, CA). The 25 µl PCR reaction volume contained 50 ng of genomic DNA, 1 × *Taq* DNA polymerase buffer, 1.5 mM MgCl₂, 0.5 µM each primer, 200 µM each dNTP and 1U *Taq* DNA polymerase (Takara, Japan).

The program for PCR amplification was as follows: initial denaturation at 94 °C for 3 min, 30 cycles of 94 °C for 25 s, 58 °C for 25 s, 72 °C for 15 s, and a final extension at 72 °C for 10 min. The PCR products were digested with restriction enzyme *Sal* I (Takara, Japan). The 10 µL reaction volume consisted of 5 µL PCR product, 1 U restriction enzyme and 1 × buffer. The reaction was performed at 37°C for 3 hours, and the digested products were detected with 3% agarose gel.

5. Conclusion

This study provides an efficient method for accurately identifying glutenin genes in bread wheat through third-generation RNA sequencing. The Xinmai26 displays the optimal subunit combinations, in which HMW-GS is determined to be 1Ax1, 1Bx7 + 1By8, 1Dx5 + 1Dy10 and LMW-GS is determined to be Glu-A3c and Glu-D3a based on the specific spectrum peaks of MALDI-TOF-MS and their molecular weight. The type of *Glu-B3* allele could not be determined due to low expression level in the mature grain. Moreover, a specific functional marker was developed to make better use of 1Dx5 subunit with the extra cysteine residue, which will play an important role in wheat quality breeding through molecular marker-assisted selection.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 31871615) and Project of Science and Technology Department of Henan Province (Grant No. 202102110031).

Conflicts of interest

The authors declare no conflict of interest.

Reference

- [1] Shewry, P.R.; Tatham, A.S.; Barro, F.; Barcelo, P.; Lazzeri, P. Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. *Nat Biotechnol*, **1995**, 13, 1185-1190.
- [2] Barak, S.; Mudgil, D.; Khatkar, B.S. Biochemical and functional properties of wheat gliadins: a review. *Crit Rev Food Sci*, **2015**, 55, 357–368.
- [3] Huo, N.; Zhu, T.; Zhang, S.; Mohr, T.; Luo, M.C.; Lee, J.Y.; Distelfeld, A.;

- Altenbach, S.; Gu, Y.Q. Rapid evolution of α -gliadin gene family revealed by analyzing *Gli-2* locus regions of wild emmer wheat. *Funct Integr Genomic*, **2019**, 19, 993–1005.
- [4] Luo, C.; Griffin, W.B.; Branlard, G.; McNeil, D.L. Comparison of low and high molecular weight wheat glutenin allele effects on flour quality. *Theor Appl Genet*, **2001**, 102, 1088-1098.
- [5] Shewry, P.R.; Halford, N.G.; Lafiandra, D. Genetics of wheat gluten proteins. *Adv Genet*, **2003**, 49, 111–184.
- [6] Gil-Humanes, J.; Piston, F.; Tollefsen, S.; Sollid, L.M. and Barro, F. Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. *Proc Natl Acad Sci USA*, **2010**, 107, 17023–17028.
- [7] Gil-Humanes, J.; Piston, F.; Gimenez, M.J.; Martin, A. and Barro, F. The introgression of RNAi silencing of c-gliadins into commercial lines of bread wheat changes the mixing and technological properties of the dough. *PLoS One*, **2012**, 7, e45937.
- [8] Gil-Humanes, J.; Piston, F.; Barro, F. and Rosell, C.M. The shutdown of celiac disease-related gliadin epitopes in bread wheat by RNAi provides flours with increased stability and better tolerance to over-mixing. *PLoS One*, **2014**, 9, e91931.
- [9] Gil-Humanes, J.; Piston, F.; Altamirano-Fortoul, R.; Real, A.; Comino, I.; Sousa, C.; Rosell, C.M. and Barro, F. Reduced-gliadin wheat bread: an alternative to the gluten-free diet for consumers suffering gluten-related pathologies. *PLoS One*, **2014**, 9, e90898.
- [10] Becker, D.; Wieser, H.; Koehler, P.; Folck, A.; Mühling, K.H. and Zörb, C. Protein composition and techno-functional properties of transgenic wheat with reduced α -gliadin content obtained by RNA interference. *J Appl Bot Food Qual*, **2012**, 85, 23–33.
- [11] Altenbach, S.B.; Tanaka, C.K. and Seabourn, B.W. Silencing of omega-5 gliadins in transgenic wheat eliminates a major source of environmental variability and improves dough mixing properties of flour. *BMC Plant Biol*, **2014**, 14, 393-405.
- [12] Li, D.; Jin, H.; Zhang, K.; Wang, Z.; Wang, F.; Zhao, Y.; Huo, N.; Liu, X.; Gu, Y.Q.; Wang, D.; Dong, L. Analysis of the *Gli-D2* locus identifies a genetic target for simultaneously improving the breadmaking and health-related traits of common wheat.

The Plant Journal, **2018**, 95, 414-426.

[13] Maria, I.I.; Kiszonas, A.M.; Morris, C.F. Development of haplotype-specific molecular markers for the low-molecular-weight glutenin subunits. *Mol Breeding*, **2018**, 38, 68.

[14] Jin, H.; Zhang, Y.; Li, G.; Mu, P.; Fan, Z.; Xia, X.; He, Z. Effects of allelic variation of HMW-GS and LMW-GS on mixograph properties and Chinese noodle and steamed bread qualities in a set of Aroona near-isogenic wheat lines. *J Cereal Sci*, **2013**, 57, 146–152.

[15] Zhang, Y.; Tang, J.W.; Yan, J.; Zhang, Y.L.; Zhang, Y.; Xia, X.C.; He, Z.H. The gluten protein and interactions between components determine mixograph properties in an F₆ recombinant inbred line population in bread wheat. *J Cereal Sci*, **2009**, 50, 219-226.

[16] Goel, S.; Yadav, M.; Singh, K.; Jaat, R.S.; Singh, N.K. Exploring diverse wheat germplasm for novel alleles in HMW-GS for bread quality improvement. *J Food Sci Technol*, **2018**, 55, 3257–3262.

[17] He, Z.H.; Liu, L.; Xia, X.C.; Liu, J.J.; Pena, R.J. Composition of HMW and LMW glutenin subunits and their effects on dough properties, pan bread, and noodle quality of Chinese bread wheats. *Cereal Chem*, **2005**, 82, 345–350.

[18] Dai, S.; Xu, D.; Yan, Y.; Wen, Z.; Zhang, J.; Chen, H.; Lu, Z.; Li, H.; Cong, H.; Wei, Y.; Zheng, Y.; Yan, Z. Characterization of high- and low-molecular-weight glutenin subunits from Chinese Xinjiang wheat landraces and historical varieties. *J Food Sci Technol*, **2020**, 57, 3823–3835.

[19] Zhang, D.L.; He, T.T.; Liang, H.H.; Huang, L.Y.; SU, Y.Z.; LI, Y.G.; Li, S.P. Flour quality and related molecular characterization of high molecular weight glutenin subunit genes from wild emmer wheat accession TD-256. *J Agr Food Che*, **2016**, 64, 5128-5136.

[20] Dong, Z.; Yang, Y.; Li, Y.; Zhang, K.; Lou, H.; An, X. Haplotype Variation of *Glu-D1* Locus and the Origin of *Glu-D1d* Allele Conferring Superior End-Use Qualities in Common Wheat. *PLoS One*, **2013**, 8, e74859.

- [21] Rasheed, A.; Xia, X.C.; Yan, Y.M.; Appels, R.; Mahmood, T.; He, Z.H. Wheat seed storage proteins: advances in molecular genetics, diversity and breeding applications. *J Cereal Sci*, **2014**, 60, 11–24.
- [22] Gupta, R.B.; Shepherd, K.W. Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutenin. I. Variation and genetic control of the subunits in hexaploid wheats. *Theor Appl Genet*, **1990**, 80, 65–74.
- [23] Zhang, X.F.; Liu, D.C.; Zhang, J.H.; Jiang, W.; Luo, G.B.; Yang, W.L.; Sun, J.Z.; Tong, Y.P.; Cui, D.Q. and Zhang, A.M. Novel insights into the composition, variation, organization, and expression of the low-molecular-weight glutenin subunit gene family in common wheat. *J Exp Bot*, **2013**, 64, 2027–2040.
- [24] Masci, S.; Lew, E.J.L.; Lafiandra, D.; Porceddu, E.; Kasarda, D.D. Characterization of low molecular weight glutenin subunits in durum wheat by reversed-phase high-performance liquid chromatography and N-terminal sequencing. *Cereal Chem*, **1995**, 72, 100–104.
- [25] Zhang, W.; Gianibelli, M.C.; Rampling, L.; Gale, K.R. Characterization and marker development for low molecular weight glutenin genes from *Glu-A3* alleles of bread wheat (*Triticum aestivum* L.). *Theor Appl Genet*, **2004**, 108, 1409–1419.
- [26] Beom, H.R.; Kim, J.S.; Jang, Y.R.; Lim, S.H.; Kim, C.K.; Lee, C.K.; Lee, J.Y. Proteomic analysis of low-molecular-weight glutenin subunits and relationship with their genes in a common wheat variety. *3 Biotech*, **2018**, 8, 56-64.
- [27] Liu, T.M.; Ikeda, T.M.; Branlard, G.; Peña, R.J.; Rogers, W.J.; Lerner, S.E.; Kolman, M.A.; Xia, X.; Wang, L.; Ma, W.; Appels, R.; Yoshida, H.; Wang, A.; Yan, Y.; He, Z. Comparison of low molecular weight glutenin subunits identified by SDS-PAGE, 2-DE, MALDI-TOF-MS and PCR in common wheat. *BMC Plant Biol*, **2010**, 10, 124-147.
- [28] Yu, Z.; Han, C.; Yan, X.; Li, X.; Jiang, G.; Yan, Y. Rapid characterization of wheat low molecular weight glutenin subunits by ultraperformance liquid chromatography (UPLC). *J Agric Food Chem*, **2013**, 61, 4026-4034.
- [29] Wang, A.; Liu, L.; Peng, Y.; Islam, S.; Applebee, M.; Appels, R. Identification of

low molecular weight glutenin alleles by matrix-Assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) in common Wheat (*Triticum aestivum* L.). *PLoS One*, **2015**, 10, e0138981.

[30] Jang, Y.R.; Cho, K.; Kim, S.W.; Susan, B.; Altenbach; Lim, S.H.; Sim, J.R.; Lee, J.Y. Development of an optimized MALDI-TOF-MS method for high-throughput identification of high-molecular-weight glutenin subunits in wheat. *Molecules*, **2020**, 25, 4347-4364.

[31] Dong, L.; Huo, N.; Wang, Y.; Deal, K.; Wang, D.; Hu, T.; Dvorak, J.; Anderson, O.D.; Luo, M.C.; Gu, Y.Q. Rapid evolutionary dynamics in a 2.8-Mb chromosomal region containing multiple prolamin and resistance gene families in *Aegilops tauschii*. *Plant J*, **2016**, 87, 495–506.

[32] Li, Y.G.; Liang, H.H.; Bai, S.L.; Zhou, Y.; Sun, G.L.; Su, Y.R.; Gao, A.L.; Zhang, D.L.; Li, S.P. Molecular characterization and variation of the celiac disease epitope domains among α -gliadin genes in *Aegilops tauschii*. *J Agr Food Chem*, **2017**, 65, 3422-3429.

[33] Zimin, A.V.; Puiu, D.; Hall, R.; Kingan, S.; Clavijo, B.J.; Salzberg, S.L. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *GigaScience*, **2017**, 6, 1–7.

[34] Travers, K.J.; Chin, C.S.; Rank, D.R.; Eid, J.S.; Turner, S.W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids*, **2010**, 38, e159.

[35] Zhang, W.; Ciclitira, P.; Messing, J. PacBio sequencing of gene families — A case study with wheat gluten genes. *Gene*, **2014**, 533, 541–546.

[36] Wan, Y.; Wang, D.; Shewry, P.R.; Halford, N.G. Isolation and characterization of five novel high molecular weight subunit of glutenin genes from *Triticum timopheevi* and *Aegilops cylindrica*. *Theor Appl Genet*, **2002**, 104, 828-839.

[37] AACC 26–50 (2000a) Approved methods of the American Association of cereal chemists, 10th edn. American Association of Cereal Chemists, St. Paul.

[38] AACC 38–12A (2000b) Approved methods of the American Association of cereal

chemists, 10th edn. American Association of Cereal Chemists, St. Paul.

[39] AACC 54–21 (2000c) Approved methods of the American Association of cereal chemists, 10th edn. American Association of Cereal Chemists, St. Paul.

[40] Zhang, Q.; Dong, Y.; An, X.; Wang, A.; Zhang, Y.; Lia, X.; Gao, L.; Xia, X.; He, Z.; Yan, Y. Characterization of HMW glutenin subunits in common wheat and related species by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). *J Cereal Sci*, **2008**, 47, 252–261.

[41] Kent, W.J. BLAT - The BLAST-like alignment tool. *Genome Res*, **2002**, 12, 656–664.

[42] Sahlin, K.; Tomaszewicz, M.; Makova, K.D.; Medvedev, P. Deciphering highly similar multigene family transcripts from iso-seq data with isocon. *Nat Commun*, **2018**, 9, 4601-4612.

[43] Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*, **2013**, 30, 772–780.

[44] Su, Y.Z.; Zou, M.W.; Zhu, Y.M.; Han, X.; Li, Y.G.; Zhang, D.L.; Li, S.P. Analysis of population structure and origin in *Aegilops tauschii* Coss. from China through SNP markers. *Genet Resour Crop Ev*, **2020**, 67, 923-934.

[45] Wang, K.; An, X.L.; Pan, L.P.; Dong, K.; Gao, L.Y.; Wang, S.L.; Xie, Z.Z.; Zhang, Z.; Appels, R.; Ma, W.; Yan, Y.M. Molecular characterization of HMW-GS *1Dx3'* and *1Dx4'* genes from *Aegilops tauschii* and their potential value for wheat quality improvement. *Hereditas*, **2012**, 149, 41-49.

[46] Larkin, M.A. Blackshields, G. Brown, N.P. Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Clustal W and Clustal X version 2.0. *Bioinformatics*, **2007**, 23, 2947-2948.

[47] Li, G.; Wang, L.; Yang, J.; He, H.; Jin, H.; Li, X.; Ren, T.; Ren, Z.; Li, F.; Han, X.; Zhao, X.; Dong, L.; Li, Y.; Song, Z.; Yan, Z.; Zheng, N.; Shi, C.; Wang, Z.; Yang, S.; Xiong, Z.; Zhang, M.; Sun, G.; Zheng, X.; Gou, M.; Ji, C.; Du, J.; Zheng, H.; Doležal, J.; Deng, X.W.; Stein, N.; Yang, Q.; Zhang, K.; Wang, D. A high-quality genome

assembly highlights rye genomic characteristics and agronomically important genes. *Nat. Genet.* **2021**, 53, 574-584.

[48] Tsilo, T.J.; Nygard, G.; Khan, K.; Simsek, S.; Hareland, G.A.; Chao, S.; Anderson, J.A. Molecular genetic mapping of QTL associated with flour water absorption and farinograph related traits in bread wheat. *Euphytica*, **2013**, 194, 293-302.

[49] Edwards, N.M. Gianibelli, M.C.; McCaig, T.N.; Clarke, J.M.; Ames, N.P.; Larroque, O.R.; Dexter, J.E. Relationships between dough strength, polymeric protein quantity and composition for diverse durum wheat genotypes. *J Cereal Sci.* **2007**, 45, 140-149.

[50] Ibba, M.I.; Kiszonas, A.M.; Guzman, C.; Morris, C.F. Definition of the low molecular weight glutenin subunit gene family members in a set of standard bread wheat (*Triticum aestivum* L.) varieties. *J Cereal Sci.* **2017**, 74, 263-271.

[51] Jiang, C.; Pei, Y.; Zhang, Y.; Li, X.; Yao, D.; Yan, Y.; Ma, W.; Hsam, S.; Zeller, F. Molecular cloning and characterization of four novel LMW glutenin subunit genes from *Aegilops longissima*, *Triticum dicoccoides* and *T. zhukovskyi*. *Hereditas*, **2008**, 145, 92-98.

[52] Wieser, H. Chemistry of gluten proteins. *Food Microbiol.* **2007**, 24, 115–119.

[53] Zhang, X.; Jin, H.; Zhang, Y.; Liu, D.; Li, G.; Xia, X.; He, Z.; Zhang, A. Composition and functional analysis of low-molecular-weight glutenin alleles with Aroona near-isogenic lines of bread wheat. *BMC Plant Biol.* **2012**, 12, 243-258.

[54] Liu, Z.; Yan, Z.; Wan, Y.; Liu, K.; Zheng, Y.; Wang, D. Analysis of HMW glutenin subunits and their coding sequences in two diploid *Aegilops* species. *Theor Appl Genet.* **2003**, 106, 1368–1378.

[55] Wang, J.R.; Yan, Z.H.; Wei, Y.M.; Zheng, Y.L. Characterization of high-molecular-weight glutenin subunit genes from *Elytrigia elongate*. *Plant Breed.* **2006**, 125, 89–95.

[56] Sun, X.; Hu, S.; Liu, X.; Qian, W.; Hao, S.; Zhang, A.; Wang, D. Characterization of the HMW glutenin subunits from *Aegilops searsii* and identification of a novel variant HMW glutenin subunit. *Theor Appl Genet.* **2006**, 113, 631–641.

- [57] Liu, S.; Gao, X.; Xia, G. Characterizing HMW-GS alleles of decaploid *Agropyron elongatum* in relation to evolution and wheat breeding. *Theor Appl Genet*, **2008**, 116, 325–334.
- [58] Li, X.; Zhang, Y.; Gao, L.; Wang, A.; Ji, K.; He, Z.; Appels, R.; Ma, W.; Yan, Y.; Molecular cloning, heterologous expression, and phylogenetic analysis of a novel y-type HMW glutenin subunit gene from the G genome of *Triticum timopheevi*. *Genome*, **2007**, 50, 1130–1140.
- [59] Liu, S.W.; Zhao, S.Y.; Chen, F.G. and Xia, G.M. Generation of novel high quality HMW-GS genes in two introgression lines of *Triticum aestivum*/*Agropyron elongatum*. *BMC Evol Biol*, **2007**, 7, 76-83.
- [60] Ikeda, T.M.; Nagamine, T.; Fukuoka, H.; Yano, H. Identification of new low-molecular-weight glutenin subunit genes in wheat. *Theor Appl Genet*, **2002**, 104, 680–687.
- [61] Huang, Z.; Long, H.; Jiang, Q.T.; Wei, Y.M.; Yan, Z.H.; Zheng, Y.L. Molecular characterization of novel low-molecular-weight glutenin genes in *Aegilops longissima*. *J Appl Genet*, **2010**, 51, 9–18.
- [62] Dong, L.L.; Zhang, X.F.; Liu, D.C.; Fan, H.J.; Sun, J.Z.; Zhang, Z.J.; Qin, H.J.; Li, B.; Hao, S.T.; Li, Z.S.; Wang, D.W.; Zhang, A.M.; Ling, H.Q. New insights into the organization, recombination, expression and functional mechanism of low molecular weight glutenin subunit genes in bread wheat. *PLoS ONE*, **2010**, 5, e13548.
- [63] Zhang, X.F.; Liu, D.C.; Yang, W.L.; Liu, K.F.; Sun, J.Z.; Guo, X.L.; Li, Y.W.; Wang, D.W.; Ling, H.Q.; Zhang, A.M. Development of a new marker system for identifying the complex members of the low-molecular-weight glutenin subunit gene family in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet*, **2011**, 122, 1503–1516.
- [64] Luo, G.B.; Zhang, X.F.; Zhang, Y.L.; Yang, W.L.; Li, Y.W.; Sun, J.Z.; Zhan, K.H.; Zhang, A.M. and Liu, D.C. Composition, variation, expression and evolution of low-molecular-weight glutenin subunit genes in *Triticum Urartu*. *BMC Plant Biol*, **2015**, 15, 68-81.
- [65] Huo, N.X.; Zhu, T.T.; Susan, Altenbach, Dong, L.L.; Wang, Y.; Toni, Mohr.; Liu,

- Z.Y.; Jan, Dvorak.; Luo, M.C.; Gu, Y.Q. Dynamic evolution of α -gliadin prolamin gene family in homeologous genomes of hexaploid wheat. *Sci Rep*, **2018**, 8, 5181-5193.
- [66] Wang, D.W.; Li, D.; Wang, J.J.; Zhao, Y.; Wang, Z.J.; Yue, G.D.; Liu, X.; Qin, H.J.; Zhang, K.P.; Dong, L.L.; Wang, D.W. Genome-wide analysis of complex wheat gliadins, the dominant carriers of celiac disease epitopes. *Sci Rep*, **2017**, 7, 44609-44622.
- [67] Huo, N.; Zhang, S.; Zhu, T.; Dong, L.; Wang, Y.; Mohr, T.; Hu, T.; Liu, Z.; Dvorak, J.; Luo, M.C.; Wang, D.; Lee, J.Y.; Altenbach, S.; Gu, Y.Q. Gene duplication and evolution dynamics in the homeologous regions harboring multiple prolamin and resistance gene families in hexaploid wheat. *Front Plant Sci*, **2018**, 9, 673-685.
- [68] Shewry, P.R.; Halford, N.G. Cereal seed storage proteins: structures, properties and role in grain utilization. *J Exp Bot*, **2002**, 53, 947-958.
- [69] Rasheed, A.; Xia, X.C.; Yan, Y.M.; Appels, R.; Mahmoodb, T.; He, Z.H. Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications. *J Cereal Sci*, **2014**, 60, 11-14.
- [70] Payne, P.I.; Corfield, K.G.; Holt, L.M.; Blackman, J.A. Correlations between the inheritance of certain high-molecular-weight subunits of glutenin and breadmaking quality in progenies of six crosses of bread wheat. *J Sci Food Agr*, **1981**, 32, 51-60.
- [71] Rakszegia, M.; Pastoric, G.; Jonesc, H.D.; Békés, F.; B, s.b.; Butowb; Láng, L.; Bedo, Z.; Shewry, P.R. Technological quality of field grown transgenic lines of commercial wheat cultivars expressing the 1Ax1 HMW glutenin subunit gene. *J Cereal Sci*, **2008**, 47, 310–321.
- [72] Zhen, S.; Han, C.; Ma, C.; Gu, A.; Zhang, M.; Shen, X.; Li, X.; Yan, Y. Deletion of the low-molecular-weight glutenin subunit allele *Glu-A3a* of wheat (*Triticum aestivum* L.) significantly reduces dough strength and breadmaking quality. *BMC Plant Biol*, **2014**, 14, 367-383.
- [73] Cane, K.; Sharp, P.J.; Eagles, H.A.; Eastwood, R.F.; Hollamby, G.J.; Kuchel, H.; Lu, M.Q.; Martin, P.J. The effects on grain quality traits of a grain serpin protein and the VPM1 segment in southern Australian wheat breeding. *Aust J Agric Res*, **2008**, 59, 883-

890.

[74] Zheng, S.S.; Byrne, P.F.; Bai, G.H.; Shan, X.Y.; Reid, S.D.; Haley, S.D.; Seabourn, B.W. Association analysis reveals effects of wheat glutenin alleles and rye translocations on dough-mixing properties. *J Cereal Sci*, **2009**, 50, 283-290.

[75] Masci, S.; D'Ovidio, R.; Lafiandra, D.; Kasarda, D.D. Characterization of a low-molecular-weight glutenin subunit gene from bread wheat and the corresponding protein that represents a major subunit of the glutenin polymer. *Plant Physiol*, **1998**, 118, 1147-1158.

Table 1 Dough rheological parameters of Xinmai26 obtained from micro-doughlab (A) and gluten testing system (B).

A	Variety	Water absorption (WA, %)	Development time (DT, min)	Stability time (ST, min)	Degree of softening (DS, FU)
	Xinmai26	64.76	8.88	8.70	64.97

B	Variety	Wet gluten content (WGC, g)	Dry gluten content (DGC, g)	Gluten index (GI, %)	Gluten moisture absorption (GMA, %)
	Xinmai26	35.64	12.59	96.67	64.67

Table 2 Comparison of HMW-GS and LMW-GS molecular weights deduced from transcript gene sequences with those determined by MALDI-TOF-MS in Xinmai26

Glutenin	Deduced <i>Mr</i> from transcript genes (Da)	<i>Mr</i> by MALDI- TOF-MS (Da)	Difference (Da)	Error (%)
HMW-GS				
Transcript1854	87708	87634	74	0.08
Transcript1782	82524	82109	415	0.50
Transcript272	88109	87634	475	0.54
Transcript1528	75128	74813	315	0.42
Transcript624	67472	67212	260	0.39
LMW-GS				
Transcript907	41239	41312	-73	0.18
Transcript3598	37805	38841	-1036	2.74
Transcript206	31771	30866	905	2.85
Transcript2357	32045	31576	469	1.46
Transcript6479	32565	32172	393	1.21
Transcript1264	37862	36902	960	2.54
Transcript4321	37967	37212	755	1.99
Transcript4445	39632	40353	-721	1.82

Table 3 The secondary structure prediction of the HMW-GS and LMW-GS obtained in this study.

Subunits	Sequence ID	Structure motifs	Amino acid number	Ratio	Total	Dispersal in every region		
						N-terminal domain	Repetitive domain	C-terminal domain
1Ax1	transcript1854	α -helix	59	7.29	6	4	-	2
1Bx7	transcript1782	α -helix	51	6.64	4	3	-	1
		β -strand	8	1.04	3	3	-	-
1By8	transcript1528	α -helix	71	10.16	7	4	-	3
		β -strand	2	0.29	1	1	-	-
1Dx5	transcript272	α -helix	62	7.50	3	2	-	1
		β -strand	9	1.09	2	2	-	-
1Dy10	transcript624	α -helix	61	9.73	6	4	-	2
A3c	transcript907	α -helix	137	38.48	9	-	1	8
B3	transcript3598	α -helix	118	35.76	11	2	-	9
		β -strand	2	0.61	1	-	-	1
D3a	transcript206	α -helix	121	43.53	10	-	1	9
		β -strand	2	0.72	1	-	-	1
D3a	transcript1264	α -helix	125	37.76	9	-	-	9
D3a	transcript2357	α -helix	112	39.58	15	-	2	13
D3a	transcript4321	α -helix	118	35.33	9	1	-	8
D3a	transcript6479	α -helix	108	38.03	8	-	-	8
D3a	transcript4445	α -helix	109	31.59	10	1	-	9
		β -strand	2	0.58	1	-	-	1

Figure Captions

Fig. 1 Custom pipeline of full-length transcripts of glutenin genes acquired by third-generation transcriptome sequencing

Fig.2 Dough rheological qualities of Xinmai26 measured through farinograph with 50 g of flour. (DT: development time; ST: stability time).

Fig. 3 Separation of the glutenin subunits by SDS-PAGE. (1: Chinese spring, 2: Xinmai26, Bold arrowheads, white circles, and white arrowheads mark unique *Glu-A3*, *Glu-B3*, and *Glu-D3* LMW-GS protein bands, respectively.)

Fig.4 Spectrum pattern of HMW-GS and LMW-GS alleles from Xinmai26 by MALDI-TOF-MS. (A: HMW-GS, B: LMW-GS, arrows indicate identification pattern of *Glu-1* and *Glu-3* loci.)

Fig.5 Distribution of cysteine residues and disulphide bonds in the HMW-GS and LMW-GS. (Triangles: The additional cysteine residues involved in interchain *S-S* bonds; Square: The cysteine residues usually involved in interchain *S-S* bonds in common wheat; Stars: The cysteine residues involved in intrachain *S-S* bonds.)

Fig.6 Analysis of *Glu-D1* subunits composition and corresponding extra cysteine residue detected in partial wheat varieties. (A: *Glu-D1* subunits detected by SDS-PAGE; B: Extra cysteine residue identified by specific functional marker; 1: Jimai 229; 2: Zhengmai 336; 3: Shiluan 02-1; 4: Gaoyou 503; 5: Wenmai 6; 6: Yemao; 7: Linyou 145; 8: Henong 58-3; 9: Yumai 34; 10: Neixiang 188; 11: Gaoyou 9409; 12: Shannong 16; 13: Pingan 8; 14: Pingan 9; 15: Luomai 26)

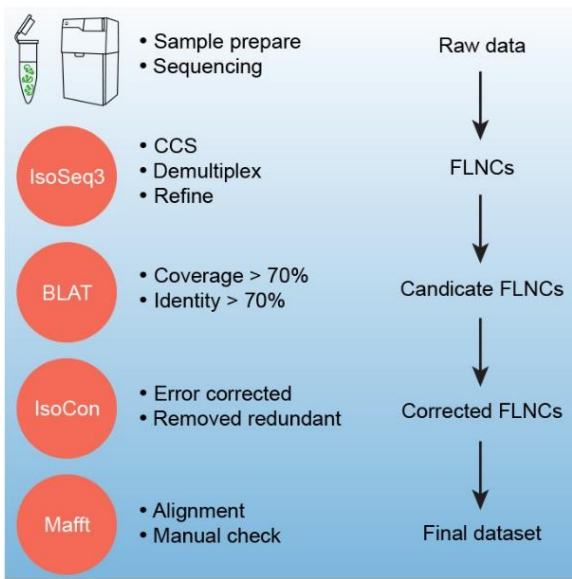


Fig. 1 Custom pipeline of full-length transcripts of glutenin genes acquired by third-generation transcriptome sequencing.

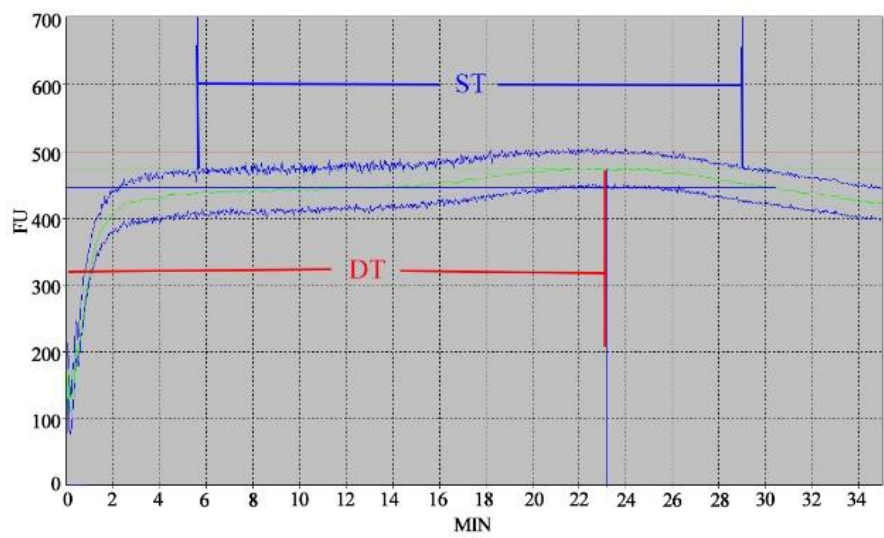


Fig.2 Dough rheological qualities of Xinmai26 measured through farinograph with 50 g of flour.
(DT: development time; ST: stability time).

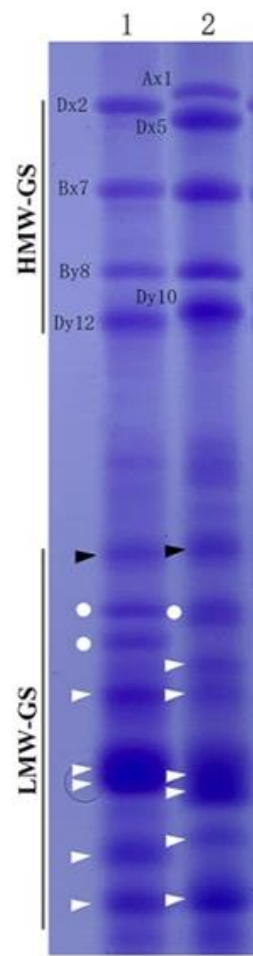


Fig. 3 Separation of the glutenin subunits by SDS-PAGE. (1: Chinese spring, 2: Xinmai26, Bold arrowheads, white circles, and white arrowheads mark unique *Glu-A3*, *Glu-B3*, and *Glu-D3* LMW-GS protein bands, respectively.)

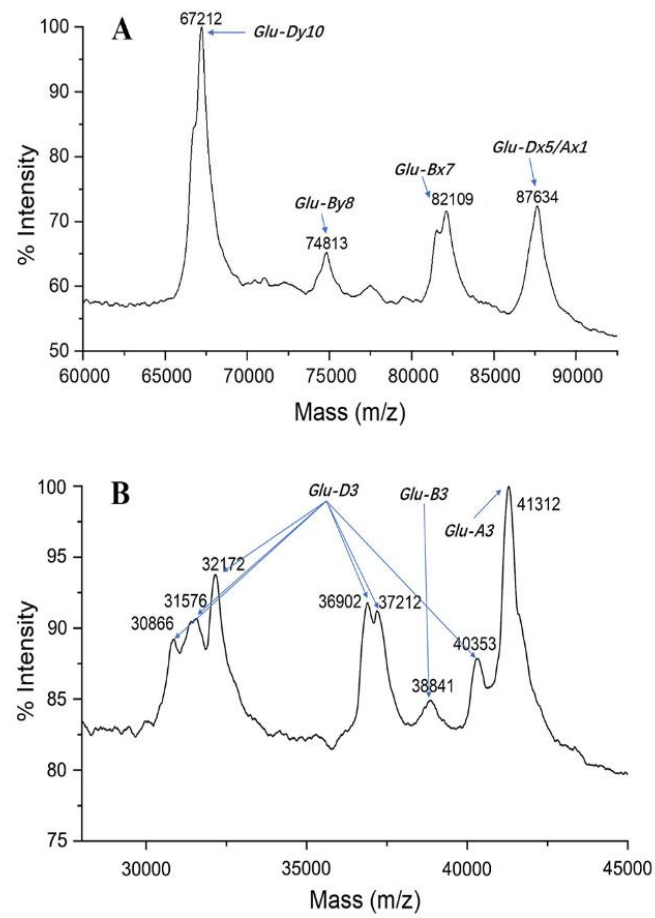


Fig.4 Spectrum pattern of HMW-GS and LMW-GS alleles from Xinmai26 by MALDI-TOF-MS.

(A: HMW-GS, B: LMW-GS, arrows indicate identification pattern of *Glu-1* and *Glu-3* loci.)

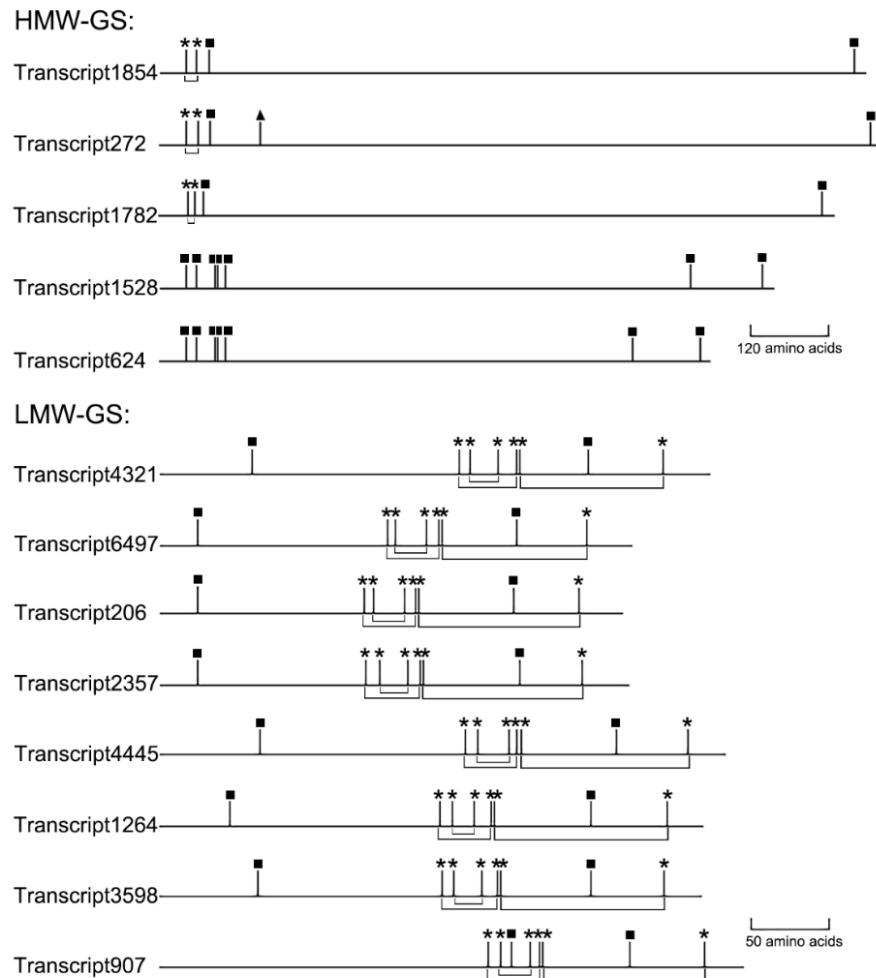


Fig.5 Distribution of cysteine residues and disulphide bonds in the HMW-GS and LMW-GS

Triangles: The additional cysteine residues involved in interchain $S-S$ bonds; Square: The cysteine residues usually involved in interchain $S-S$ bonds in common wheat; Stars: The cysteine residues involved in intrachain $S-S$ bonds.

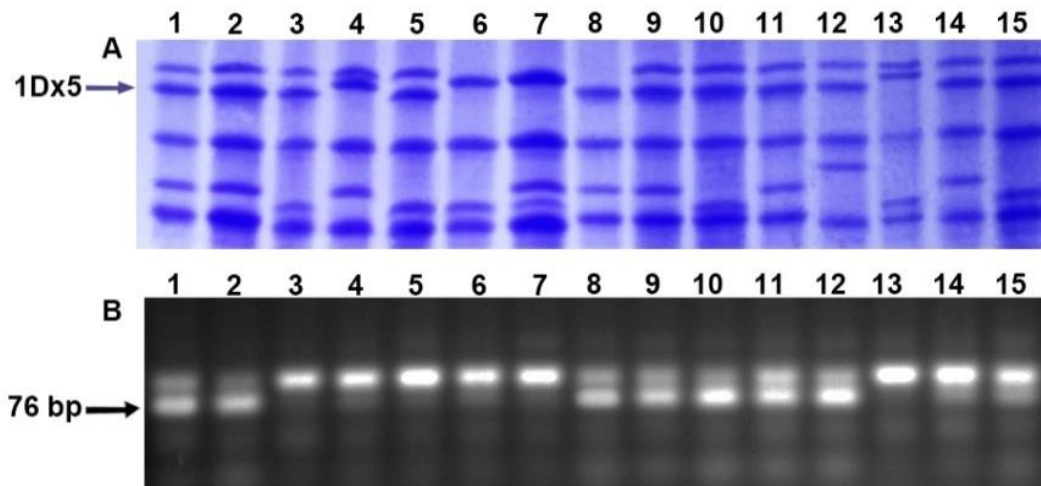


Fig.6 Analysis of *Glu-D1* subunits composition and corresponding extra cysteine residue detected in partial wheat varieties. (A: *Glu-D1* subunits detected by SDS-PAGE; B: Extra cysteine residue identified by specific functional marker; 1: Jimai 229; 2: Zhengmai 336; 3: Shiluan 02-1; 4: Gaoyou 503; 5: Wenmai 6; 6: Yemao; 7: Linyou 145; 8: Henong 58-3; 9: Yumai 34; 10: Neixiang 188; 11: Gaoyou 9409; 12: Shannong 16; 13: Pingan 8; 14: Pingan 9; 15: Luomai 26)