RESEARCH NOTE

Characterization of low-molecular-weight glutenin subunit genes of *Aegilops* section *Sitopsis* and comparative analysis with those of wheat (*Triticum aestivum* L.) and some *Aegilops* species

ZHUO HUANG[1,3], HAI LONG[2], XIAO-FANG YU[1], YU-MING WEI[3,4], ZE-HONG YAN[3,4] and YOU-LIANG ZHENG[3,4]*

[1]*College of Landscape Architecture, Chengdu Campus, Sichuan Agricultural University, Huiming Road 211, Wenjiang, Sichuan 611130, People's Republic of China*
[2]*Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan 610041, People's Republic of China*
[3]*Triticeae Research Institute, Sichuan Agricultural University, Wenjiang, Chengdu, Sichuan 611130, People's Republic of China*
[4]*Ministry of Education, Key Laboratory for Crop Genetic Resources and Improvement in Southwest China, Sichuan Agricultural University, Yaan, Sichuan 625014, People's Republic of China*

[Huang Z., Long H., Yu X.-F., Wei Y.-M., Yan Z.-H. and Zheng Y.-L. 2015 Characterization of low-molecular-weight glutenin subunit genes of *Aegilops* section *Sitopsis* and comparative analysis with those of wheat (*Triticum aestivum* L.) and some *Aegilops* species. *J. Genet.* **94**, 497–501]

Introduction

The low-molecular-weight glutenin subunits (LMW-GS) of wheat are associated with dough resistance and extensibility, and have been extensively studied in wheat and its relatives. However, little is known about the LMW-GS genes in *Sitopsis* section of *Aegilops*, the proposed B genome origin of common wheat. In this study, 50 LMW-GS genes were obtained from 13 accessions of five species of *Sitopsis*, among which two novel *Sitopsis* specific types of LMW-GS genes were identified. The comprehensive comparative and phylogenetic analyses of the conserved and nonconserved regions of 109 LMW-GS genes indicated that there might be at least three ancestors in the LMW-GS gene families in *Triticum* and *Aegilops*. The genes obtained from *Sitopsis* species are unambiguously closer to those of B genome of wheat than any other *Aegilops* species analysed, supporting that *Sitopsis* is the donor of wheat B genome. Our results provide better understanding of LMW-GS gene families, and the clues of relationship among wheat and its relatives.

Wheat glutenins are polymeric proteins which are formed by glutenin subunits held together with intermolecular disulphide bonds. They mainly consist of high-molecular-weight glutenin subunits (HMW-GS) and low-molecular-weight glutenin subunits (LMW-GS). The typical LMW-GS are encoded by *Glu-3* loci (Singh and Shepherd 1988; D'Ovidio and Masci 2004).

Genus *Aegilops* plays essential role in origin and evolution of hexaploid wheat. It not only contributes the entire D genome of wheat, but also has been proposed as B genome ancestor. Consequently, extensive studies on novel LMW-GS gene mining have also been conducted in *Aegilops* genus, such as species containing D, C, M, N, U genomes (Ciaffi *et al*. 1999; Gianibelli *et al.* 2002; Johal *et al.* 2004; Huang *et al.* 2005; Wang *et al.* 2011; Li *et al.* 2008, 2010). However, little information is available about the LMW-GS genes in *Sitopsis* section, the putative B genome donor of wheat. In this study, we report the isolation and characterization of the novel LMW-GS genes from the five species of *Sitopsis* section of the genus *Aegilops*, *Ae. longissima* (2n = 2*x* = 14, S^lS^l), *Ae. sharonensis* (2n = 2*x* = 14, $S^{sh}S^{sh}$), *Ae. searsii* (2n = 2*x* = 14, S^sS^s), *Ae. bicornis* (2n = 2*x* = 14, S^bS^b) and *Ae. speltoides* (2n = 2*x* = 14, SS). The comprehensive sequence comparison and phylogenetic analyses extend our knowledge about LMW-GS genes of *Sitopsis* section and their relationship with those of wheat and other *Aegilops* species.

Materials and methods

Plant materials

A total of 13 accessions from five species of *Sitopsis* were used for gene cloning (table 1 in electronic supplementary

*For correspondence. E-mail: ylzheng@sicau.edu.cn.
Zhuo Huang and Hai Long contributed equally to this work.

Keywords. wheat; *Aegilops*; low-molecular-weight glutenin subunits; gene cloning; B genome; *Triticum aestivum*.

material at http://www.ias.ac.in/jgenet/). These materials were obtained from National Small Grains Collection, United States Department of Agriculture–Agricultural Research Service (USDA–ARS), Aberdeen, USA.

PCR primers

Two primer sets were used to obtain the complete open reading frames (ORFs) of LMW-GS genes. The first primer set, pGluB-F (5′-ATCAAAACCAAGCAAC-AGTAT-3′) and pGluB-R (5′-ATATTACTAGAGATATTTTCTTATCA-3′), was proved being B-genome specific in wheat (Long *et al*. 2006), and the second primer set, pGluU-F (5′-GAT CATCACAG/AGCACAAT/GCATCA-3′) and pGluU-R (5′-TTTTCTTATCAGTAGG/C/ACACCAACT-3′) (underlined nucleotides showed that they are degenerate), was developed as degenerate universal primers based on the conserved sequences of known LMW-GS genes of wheat.

DNA extraction and PCR amplification

Genomic DNA was extracted from 3 to 5 day old seedlings by using CTAB procedure (Murray and Thompson 1980). The PCR amplification was conducted in a volume of 50 μL containing 200 ng of genomic DNA, 200 μM of each dNTP, 5 μM of each primer, 3 U *ExTaq* DNA polymerase with high fidelity (TaKaRa, Dalian, China), and 1× PCR buffer (supplied with the *Taq* polymerase). The cycling parameters were 94°C for 5 min, followed by 35 cycles of 94°C for 45 s, 58°C for 1 min, and 72°C for 1 min 30 s, and a final extension at 72°C for 5 min. The electrophoretic separation of amplified products was conducted using 1% agarose gels in 1× TAE buffer.

Cloning, sequencing and sequence analysis

The amplified products were recovered from the gels and ligated onto the pMD18-T vector (TaKaRa), and then transformed into competent cells of *E. coli* strain DH5α. The positive colonies were confirmed by PCR using M13 universal primers. Several positive colonies were randomly selected for bidirectional Sanger sequencing by commercial company (Invitrogen, Shanghai, China).

The complete coding sequence was assembled by DNAMAN 6.0 (Lynnon Corporation, San Ramon, USA) with minimum overlap of 300 bp and identity more than 95%. The homology analysis with known LMW-GS genes was conducted by Blastn search against the non-redundant (nr) database on NCBI network (http://www.ncbi.nlm.nih/gov/). Sequence alignment was performed by ClusterW (Larkin *et al*. 2007). Phylogentic analysis was conducted by MEGA5.2 using Neighbor-joining (NJ) method with 1000-bootstrap replications (Tamura *et al.* 2011). For nucleotide tree, the Kimura 2-parameter model was used with the following parameters: transitions and transversions substitution, homogeneous pattern among lineages, uniform rates among sites and pair-wise deletion for gaps/missing data. For amino acid tree, the Poisson model was used with following parameters: homogeneous pattern among lineages, uniform rates among sites and pair-wise deletion for gaps/missing data. Motif logo was drawn by WebLogo (Crooks *et al.* 2004).

Results

Using primer sets pGluB and pGluU, about 900 to 1100 bp bands were amplified, which were then recovered and purified from gels and ligated to T-vector. The positive clones were screened by PCR using M13 universal primers. In each accession, a maximum 10 positive clones were chosen for bidirectional sequencing. Finally, the DNA sequences of totally 50 positive clones were harvested.

Combined with the six sequences of *Ae. longissima* obtained in our previous study (Huang *et al.* 2010), a total of 56 clones were isolated from *Sitopsis*, including six sequences from *Ae. bicornis*, 24 from *Ae. longissima*, nine from *Ae. searsii*, 12 from *Ae. sharonesis* and five from *Ae. speltoides* (table 2 in electronic supplementary material). The lengths of obtained sequences range from 813 to 1101 bp. Twenty clones were putative pseudogenes due to premature translation termination. Homology analysis indicated that they share high degree similarities (86–99%) with known LMW-GS genes from *T. aestivum* and *T. durum* (table 2 in electronic supplementary material).

Phylogentic analysis was conducted based on the alignment of nucleotide sequences of 109 LMW-GS genes, including 56 obtained in this study, known genes representing nine typical LMW-GS groups (Long *et al.* 2005) in A, B and D genomes of wheat and known genes from other species of genus *Aegilpos* (table 3 in electronic supplementary material). An orthologue of LMW-GS in *Secale cereale*, encoding the γ-secalin (EF432547) was used as out group. All LMW-GS could be remarkably separated into three main clusters (figure 1; figure 2 in electronic supplementary material). Cluster I contains 14 sequences: one derived from A genome and the others were from *Ae. markgrafii*, *Ae. umbellulata*, *Ae. uniaristata*, *Ae. comosa* and *Ae. speltoides*. All these genes lack the typical 13-aa amino acid N-terminal domain. Cluster II is comprised of 60% sequences analysed. Thirty-five sequences obtained in this study were grouped distinctly from others. Two A-genome derived LMW-GS are clustered solely. Three M-types of LMW-GS from D genome, designated as METSRV-D, METSC-D and METRC-D were clustered with genes mainly from *Aegilops* species with C, U, M, N, CU, CD, MD genomes and formed three sub-clusters, respectively, excepting only one known gene from *Ae. speltoides*. Cluster III is composed of 29 genes, including three typical LMW-GS originating from B and D genomes of wheat, designated as METSHIP-B, METSHIP-D and MEN-B, respectively. The genes clustered closely with the B-genome derived LMW-GS genes are all from *Sitopsis* species, including *Ae. bicornis*, *Ae. longissima*, *Ae. sharonesis* and *Ae. speltoides*.

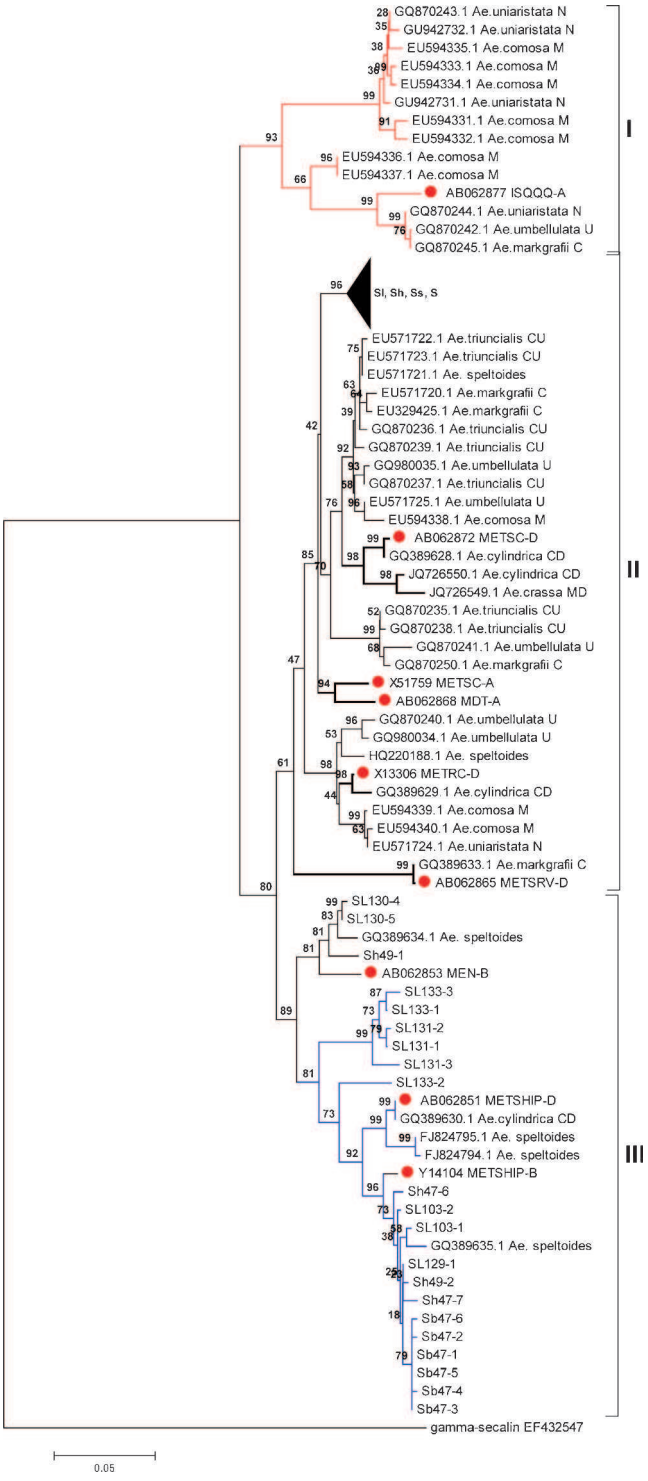

Figure 1. Phylogenetic tree of LMW-GS obtained in species of *Sitopsis* and other known LMW-GS from wheat and species in *Aegilops* genus. The representing typical LMW-GS in specific genomes of wheat are indicated by small rounds. The compressed branch (the black triangle) represents some genes obtained in this study and is shown in figure 2 in electronic supplementary material. The bar under the tree indicates the distance scale.

N-terminal domain

Four types of N-terminal domains, METSHILSLEKPL, METSCIPSLERPW, MENSHIPGLEKPS and MENSHIPGLERPS were identified in *Sitopsis* and each consists of 13, 35, 6 and 1 clones, respectively. Among them, MENSHIPGLERPS obtained in *Ae. sharonesis* (tables 1 & 3 in electronic supplementary material) is identical to one of B genome. METSCIPSLERPW and METSHILSLEKPL are *Sitopsis* specific, and are highly similar to the known LMW-GS in A and B genomes of wheat, respectively (tables 1 & 2 in electronic supplementary material). Seven clones from *Ae. longissima* contain the N-terminal sequences of MENSHIPGLEKPS, which is one residue different from MENSHIPGLERPS.

We also compared the N-terminal domain of other *Aegilops* species with *Sitopsis* and wheat. In all, 15 typical N-terminal sequences were identified (tables 3 & 4; figure 1 in electronic supplementary material). The motif logos were drawn from alignments of both nucleotide and deduced amino acids (figure 2, a&b). Two putative consensus forms of N-terminal domain, METSCIPGLERPW and METSHIPGLEKPW, are then identified. The phytogenetic analysis based on the alignment of nucleotide sequences also indicated that typical N-terminal domains are grouped into two main clusters, whereas the LMW-GS lacking a typical N-terminal sequence are significantly separated from others (figure 2c).

Repetitive domain

Analysis of the 109 LMW-GS showed high degree of length variations from 52 to 184 amino acids. Sequence alignment analysis showed that the composition of repeat motifs seem to be divergent between different types of LMW-GS (according to N-terminal sequences), whereas relatively conserved interiorly. Then, a phylogenetic analysis based on the alignment of deduced amino acid sequences of repetitive domain was conducted and showed that similar to the cluster patterns of N-terminal domains, all repetitive domains could be separated into three major clusters: one composed of LMW-GS with N-terminal sequence of METSCIPGLERPW and related forms of MDTSCIPGLERPW, METRCIPGLERPW, METSCIPSLERPW, METSCISGLERPW, and the singletons of MATSCIPGSESPW and METRCVPGLERPW, in which two types of N-terminal sequences derived from A genome are clustered solely. The LMW-GS with N-terminal sequences of METSHIPGLEKPS, MENSHIPGLE(K/R)PS, METSHIPSLEKPL, METSHILSLEKPL, METSHIPGLENPS and IENSHIPGLEKPS are grouped to another cluster. Besides the genes of B and D genomes, almost all other genes in this cluster are from *Sitopsis* species, except for one LMW-GS of *Ae. cylindrica* with D genome. The last cluster mainly consists of LMW-GS without typical N-terminal sequence. Although two LMW-GS with METSRVPGLEKPW were also found in this cluster, they are clustered distinctly from others.

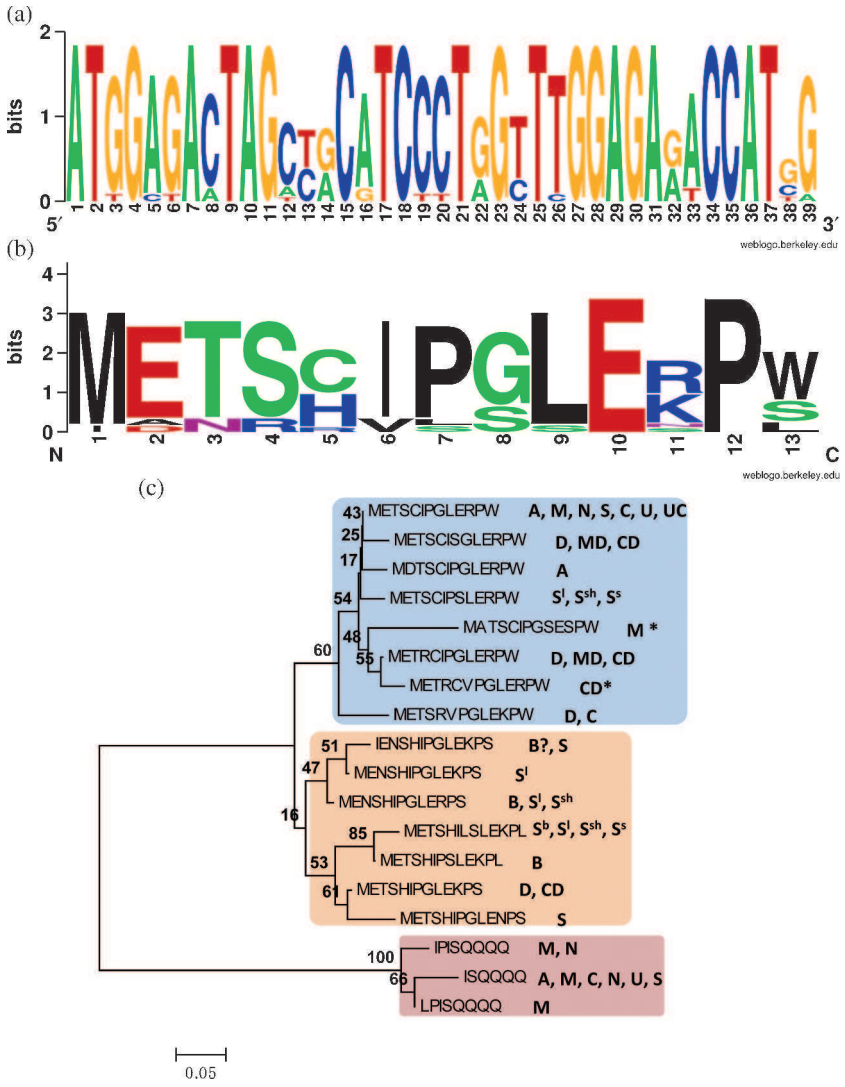

Figure 2. Sequence analysis of deduced amino acids of N-terminal domain. Motif logos of N-terminal domain constructed by alignment of nucleotide (a) and deduced amino acid sequences (b) of all N-terminal domain analysed in this study. (c) The NJ tree of nucleotide sequences of N-terminal domains. The letters on the right side of the tree show the genome symbols, in which a corresponding LMW-GS with a N-terminal sequence is found. The bar under the tree indicates the distance scale.

Discussion

Although several LMW-GS genes were isolated from *Ae. longissima* and *Ae. speltoides* (Wang *et al.* 2011), the knowledge about LMW-GS genes of the *Sitopsis* section of the genus *Aegilops* is very limited. In this study, we isolated 50 LMW-GS genes from 13 accessions of five species of *Sitopsis*. Two novel types with unique N-terminal sequences are identified compared with the known genes. Five obtained LMW-GS were found to contain nine cysteines in the C-terminal domain (data not shown), though they are all pseudogenes (table 5 in electronic supplementary material). These results not only extend the knowledge of the LMW-GS families, but also suggested that there are abundant variations of LMW-GS in *Sitopsis* of *Aegilops* which could be considered as valuable resource for wheat quality improvement.

Fifteen typical N-terminal sequences were concluded from 109 LMW-GS genes of wheat and *Aegilop* species. Two consensus forms of N-terminal sequences, METSCIPGLERPW and METSHIPGLEKPW were deduced. The former is the most frequent type which is not only found in the A genome of wheat but also distributed in species of *Aegilops* with genomes of M, N, S, U, C and UC (figure 2c). The latter has not been found in known genes. Therefore, METSCIPGLERPW might be among the most ancient form.

The repetitive domain is highly variable in number of repeated motif which is responsive for extensive length variations of LMW-GS. In this study, we found that the repeat composition in LMW-GS with same N-terminal sequences is quite similar. A phylogenetic analysis indicated that similar to the N-terminal sequence-based analysis, three main clusters could be built, and the LMW-GS under same or similar N-terminal types are clustered closely, (figure 3 in electronic

supplementary material). A specific motif 'ILWY' is found in three closely-related types, METSHIL-S, D_METSHIP-D, and B_METSHIP, and 'TLSH' appears in LMW-GS whose N-terminal sequences starting with MEN/IEN–implying they might have same origins (data omitted). The repeat composition of METSCIPGLERPW is the most divergent among all types which are separately clustered with genes of A, D genomes and *Sitopsis* species with N-terminal sequences of METRC-D, METSC-A and METSCIPSLERPW, respectively (figure 3 in electronic supplementary material). The combined results from the analyses of conserved and highly variable regions suggested that there might be at least three ancestors during evolution of LMW-GS gene families in wheat and *Aegilops* genus.

It was also found that two types of B genome derived LMW-GS belong to cluster III and are unambiguously closely related to LMW-GS genes of four species in *Sitopsis* (figures 1 & 3 in electronic supplementary material). This result supports the species of *Sitopsis* as the putative origin of B genome of wheat (Sarkar and Stebbins 1956; Kerby and Kuspira 1988). However, no sufficient evidence supports that *Ae. speltoides* alone as the B genome donor.

Acknowledgements

This work is supported by National Natural Science Foundation of China (no. 31230053). Zhuo Huang is supported by 'Shuangzhi' project found by Sichuan Agricultural University.

References

Ciaffi M., Lee Y. K., Tamas L., Gupta R., Skerritt J. and Appels R. 1999 The low-molecular-weight glutenin subunit proteins of primitive wheats III; the genes from D-genome species. *Theor. Appl. Genet.* **98**, 135–148.

Crooks G. E., Hon G., Chandonia J. M. and Brenner S. E. 2004 WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.

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

Gianibelli M. C., Wrigley C. W. and MacRitchie F. 2002 Polymorphism of low Mr Glutenin subunits in *Triticum tauschii*. *Cereal Chem.* **35**, 277–286.

Huang Z., Long H., Yan Z. H., Wei Y. M. and Zheng Y. L. 2005 Sequence analysis of LMW-GS genes from *Aegilops tauschii* containing HMW glutenin subunits 5+12 with good quality. *High Technol. Lett.* **15**, 67–72.

Huang Z., Long H., Jiang Q. T., Wei Y. M., Yan Z. H. and Zheng Y. L. 2010 Molecular characterization of novel low-molecular-weight glutenin genes in *Ageilops longissima*. *J. Appl. Genet.* **51**, 9–18.

Johal J., Gianibelli M. C., Rahman S., Morell M. K. and Gale K. R. 2004 Characterization of low-molecular-weight glutenin genes in *Aegilops tauschii*. *Theor. Appl. Genet.* **109**, 1028–1040.

Kerby K. and Kuspira J. 1988 Cytological evidence bearing on the origin of the B genome in polyploid wheats. *Genome* **30**, 36–43.

Larkin M. A., Blackshields G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H. *et al.* 2007 Clustal W and Clustal X version 2.0. *Bioinfo* **23**, 2947–2948.

Li X. H., Ma W., Gao L. Y., Zhang Y. Z., Wang A. L. and Ji K. M. 2008 A novel chimeric low molecular weight glutenin subunit gene from the wild relatives of wheat *Aegilops kotschyi* and *Ae juvenalis*: evolution at the *Glu-3* loci. *Genetics* **180**, 93–101.

Li X. H., Wang K., Wang S. L., Gao L. Y., Xie X. X. and Hsam S. L. K. 2010 Molecular characterization and comparative transcriptional analysis of LMW-m type genes from wheat (*Triticum aestivum* L.) and *Aegilops species*. *Theor. Appl. Genet.* **121**, 845–856.

Long H., Wei Y. M., Yan Z. H., Baum B., Nevo E. and Zheng Y. L. 2005 Classification of wheat low-molecular-weight glutenin subunit genes and its chromosome assignment by developing LMW-GS group-specific primers. *Theor. Appl. Genet.* **111**, 1251–1259.

Long H., Wei Y. M., Yan Z. H., Baum B., Nevo E. and Zheng Y. L. 2006 Analysis and validation of genome-specific DNA variations in 5′ flanking conserved sequences of wheat low-molecular-weight glutenin subunit genes. *Sci. China Ser. C-life Sci.* **49**, 322–331.

Murray M. and Thompson W. F. 1980 Rapid isolation of high molecular weight plant. *DNA Nucleic Acids Res.* **8**, 4321–4325.

Sarkar P. and Stebbins G. L. 1956 Morphological evidence concerning the origin of the B genome in wheat. *Am. J. Bot.* **43**, 297–304.

Singh N. K. and Shepherd K. W. 1988 The structure and genetic control of a new class of disulphide-linked proteins in wheat endosperm. *Theor. Appl. Genet.* **71**, 79–92.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum Parsimony method. *Mol. Biol. Evol.* **28**, 2731–2739.

Wang S., Li X., Wang K., Wang X., Li S., Zhang Y. *et al.* 2011 Phylogenetic analysis of C, M, N, and U genomes and their relationships with *Triticum* and other related genomes as revealed by LMW-GS genes at *Glu-3* loci. *Genome* **54**, 273–284.

Received 13 August 2014, in revised form 5 March 2015; accepted 13 March 2015
Unedited version published online: 13 March 2015
Final version published online: 16 September 2015