Theor Appl Genet (2008) 116:953–966
DOI 10.1007/s00122-008-0727-1

ORIGINAL PAPER

Molecular characterization and genomic organization of low molecular weight glutenin subunit genes at the *Glu-3* loci in hexaploid wheat (*Triticum aestivum* L.)

Xiu-Qiang Huang · Sylvie Cloutier

Received: 5 October 2007 / Accepted: 1 February 2008 / Published online: 28 February 2008
© Springer-Verlag 2008

Abstract In this study, we report on the molecular characterization and genomic organization of the low molecular weight glutenin subunit (LMW-GS) gene family in hexaploid wheat (*Triticum aestivum* L.). Eighty-two positive BAC clones were identified to contain LMW-GS genes from the hexaploid wheat 'Glenlea' BAC library via filter hybridization and PCR validation. Twelve unique LMW glutenin genes and seven pseudogenes were isolated from these positive BAC clones by primer-template mismatch PCR and subsequent primer walking using hemi-nested touchdown PCR. These genes were sequenced and each consisted of a single-open reading frame (ORF) and untranslated 5′ and 3′ flanking regions. All 12 LMW glutenin subunits contained eight cysteine residues. The LMW-m-type subunits are the most abundant in hexaploid wheat. Of the 12 LMW-GS, 1, 2 and 9 are i-type, s-type and m-type, respectively. The phylogenetic analysis suggested that the LMW-i type gene showed greater differences to LMW-s and LMW-m-type genes, which, in turn, were more closely related to one another. On the basis of their N-terminal sequences, they were classified into nine groups. Fingerprinting of the 82 BAC clones indicated 30 BAC clones assembled into eight contigs, while the remaining clones were singletons. BAC end sequencing of the 82 clones revealed that long terminal repeat (LTR) retrotransposons were abundant in the *Glu-3* regions. The average physical distance between two adjacent LMW-GS genes was estimated to be 81 kb. Most of LMW-GS genes are located in the D-genome, suggesting that the *Glu-D3* locus is much larger than the *Glu-B3* locus and *Glu-A3* locus. Alignments of sequences indicated that the same type (starting with the same N-terminal sequence) LMW-GS genes were highly conserved in the homologous genomes between hexaploid wheat and its donors such as durum wheat and *T. tauschii*.

Keywords BAC clones · BAC fingerprinting · Chromosomal location · Genomic organization · Hexaploid wheat · LMW-GS

Communicated by J. Snape.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0727-1) contains supplementary material, which is available to authorized users.

X.-Q. Huang · S. Cloutier (✉)
Cereal Research Centre, Agriculture and Agri-Food Canada,
195 Dafoe Road, R3T 2M9 Winnipeg, MB, Canada
e-mail: scloutier@agr.gc.ca

Present Address:
X.-Q. Huang
Department of Plant Science, University of Manitoba,
66 Dafoe Road, R3T 2N2 Winnipeg, MB, Canada

Introduction

Seed storage proteins play an important role in determining wheat (*Triticum aestivum* L.) quality (Payne 1987). The major seed storage proteins are called prolamins because of their high proline and glutamine content. Wheat prolamins are classified into two major groups, namely, gliadins and glutenins. Gliadins are monomeric proteins that form intramolecular disulphide bonds, whereas glutenins are polymeric proteins, which are held together by intermolecular disulphide bonds thereby forming the glutenin polymer (Shewry and Thatam 1997). Wheat glutenins consist of high molecular weight glutenin subunits (HMW-GS) and low molecular weight glutenin subunits (LMW-GS). HMW-GS account for about 10% of the total wheat gluten

fraction. They have been extensively studied because of their importance in gluten elasticity and their relatively small number (three to five subunits per hexaploid cultivar) resulting in relative ease of analysis at the protein level. Recently, the complete sequences and genomic organization of HMW-GS gene loci in hexaploid wheat, tetraploid wheat (*T. turgidum*) and diploid wheat (*T. tauschii*) have been reported (Anderson et al. 2003; Gu et al. 2004b, 2006; Kong et al. 2004).

LMW-GS represent approximately 40% of the total wheat gluten fraction. Despite their abundance, they have received less attention than the HMW-GS mainly due to difficulties associated with their characterization by SDS-PAGE and their overlap with gliadins of similar size and mobility (Gianibelli et al. 2001). LMW-GS have molecular weight values ranging from 30 to 50 KDa (D'Ovidio and Masci 2004). LMW-GSs are encoded at the *Glu-A3*, *Glu-B3* and *Glu-D3* loci on the short arms of chromosome 1A, 1B and 1D, respectively (Singh and Shepherd 1988; Gupta and Shepherd 1990). The number of LMW glutenin-like genes in hexaploid wheat was estimated to range from 30 to 40 based on Southern hybridization (Cassidy et al. 1998). This estimate is in agreement with N-terminal sequencing of glutenin fractions where 26 different LMW-GS sequences were identified in bread wheat cultivar 'Yecora Rojo' (Lew et al. 1992). These LMW-GS genes were classified into three types, LMW-i, LMW-m and LMW-s, based on the first amino acid residue of the mature protein, which correspond to isoleucine, methionine and serine, respectively (Lew et al. 1992; Cloutier et al. 2001).

Although the role of HMW-GS in bread-making quality is better understood, LMW-GS also play a significant role in the formation of large polymers. Some allelic forms of LMW-GS showed greater effects on these properties than HMW-GS (Gupta et al. 1989, 1994). Different allelic forms of LMW-GS seem to play different roles in determining different quality parameters (Luo et al. 2001). The *Glu-A3* alleles significantly influenced protein content, SDS sedimentation volume and mixograph midline peak value, whereas *Glu-B3* alleles strongly affected all characteristics and the *Glu-D3* alleles significantly affected all characteristics other than SDS sedimentation volume. Tanaka et al. (2005) found that LMW-GS significantly affected dough strength in common wheat. Using recombinant inbred lines, Maruyama-Funatsuki et al. (2004) reported that one LMW-GS, encoded by an LMW-s-type gene, was associated with good bread-making quality, thereby confirming that LMW-GSs are a major contributor to dough strength. Other studies have also shown that allelic variation of HMW-GS and LMW-GS are both associated with difference in the rheological properties of wheat flour (Payne 1987; Gupta et al. 1989; Nieto-Taladriz et al. 1994).

'Glenlea' is a Canadian Western extra strong (CWES) spring wheat cultivar. The LMW-i- and LMW-s-type genes in 'Glenlea' have been recently cloned using genomic DNA and developing seed cDNA libraries (Cloutier et al. 2001; Maruyama-Funatsuki et al. 2005). To date, more than 200 genomic DNA and cDNA clones of LMW-GS genes from hexaploid wheat and its relatives have been published (An et al. 2006; Ikeda et al. 2002; Long et al. 2005; Özdemir and Cloutier 2005; Zhao et al. 2006, 2007). Most studies reported that very few LMW-GS genes were isolated from a given wheat cultivar by PCR due to their complex compositions. Recently, LMW-GS genes were isolated from the Japanese soft wheat cultivar 'Norin 61' and classified into 12 groups based on the deduced amino acid sequences of the conserved N- and C-terminal domains (Ikeda et al. 2002). Wicker et al. (2003) reported on the genomic organization of the LMW-GS genes via comparative analysis of sequences of LMW-GS containing BAC clones from *T. monococcum* and *T. durum*, whereas Johal et al. (2004) characterized LMW-GS genes from BAC clones of *T. tauschii*, the D-genome donor of hexaploid wheat. To understand genomic organization of the LMW-GS gene family, a BAC library was constructed from 'Glenlea' (Nilmalgoda et al. 2003). The present study was undertaken with the goal of characterizing the *Glu-3* loci of hexaploid wheat by (i) physical mapping of LMW-GS containing BAC clones from the 'Glenlea' BAC library via SNaPshot fingerprinting; (ii) genome organization around the *Glu-3* loci using BAC end sequencing; (iii) isolation and molecular characterization of LMW-GS genes at the *Glu-3* loci in hexaploid wheat and (iv) phylogenetic analysis of LMW-GS genes.

Materials and methods

BAC library screening

The Glenlea BAC library contains 656,640 clones with an estimated 3.1× haploid genome coverage and has been gridded onto 24 high-density filters (Nilmalgoda et al. 2003). A composite probe, consisting of the complete coding regions of an LMW-i (AY542896, Cloutier et al. 2001), LMW-m (AY542897) and LMW-s (AY542898) (Özdemir and Cloutier 2005), was labeled with [α-^{32}P]dCTP using Ready-to-Go DNA labeling beads (Amersham Pharmacia Biotech, Inc.) and used to screen the 24 high-density filters. Hybridization was performed as described in Nilmalgoda et al. (2003).

BAC DNA extraction

Positive BAC clones identified by high-density filter hybridization were inoculated in 96-well plates containing

1.5 ml of 2× YT supplemented with 12.5 µg chloramphenicol/ml and were grown in a C25KC incubator shaker (New Brunswick Scientific CO., Edison, NJ, USA) at 37°C and 300 rpm for 20 h. BAC DNA was isolated and purified using an Eppendorf Perfectprep BAC 96 purification kit (Hamburg, Germany) adapted for a liquid handling robot (Qiagen 3000, Mississauga, ON, Canada). BAC DNA was eluted in a final volume of 60 µl.

BAC fingerprinting and contig assembly

Fingerprinting of positive clones was performed according to Luo et al. (2003). Briefly, 42 µl of BAC plasmid DNA was digested simultaneously with five restriction enzymes, namely, *Bam*HI, *Eco*RI, *Xba*I, *Xho*I and *Hae*III. The first four enzymes are all 6-bp cutters that produce a 3′ recessed end, which leaves a different nucleotide on the negative strand for each enzyme. Restriction enzyme *Hae*III is a blunt-end 4-base cutter used to generate fragments of appropriate size and number for analysis on a capillary system. The restricted BAC DNA was then labeled with dideoxynucleotides using the SNaPshot™ labeling kit (Applied Biosystems, Foster City, CA, USA). Unincorporated dideoxynucleotides were removed by ethanol precipitation before electrophoresis on a 16-capillary DNA analyzer (ABI 3100 Genetic Analyzer, Applied Biosystems). Internal size standard GeneScan LIZ-500 (size range from 35 to 500 bp) was used in each sample, although the 250-bp size fragment was removed from the subsequent GeneScan analysis as recommended by Luo et al. (2003). Restriction fragments ranging in size from 75 to 500 nucleotides were exported to GenoProfiler (available at http://wheat.pw.usda.gov/physicalmapping/) for conversion into FPC-compatible files and removal of vector and false-positive fingerprint fragments.

Contig assembly was performed using FPC version 7 (available at http://www.genome.arizona.edu/fpc/) using the stringent conditions of tolerance 5 as recommended by Luo et al. (2003) and a Sulston score of 1×10^{-10}.

LMW-GS gene-specific primer design and PCR

LMW-GS gene-specific primers were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky 2000). All primer sequences and amplicon information are shown in Table 1. The first five primer pairs from LMWF1/LMWR1 to LMWF10/LMWR9 were designed based on the coding regions of contig consensus sequences of full-length LMW-m and -i-type cDNA clones, whereas the forward primer LMWF11 and the reverse primer LMWR11 were from the promoter region and the repetitive domain of genomic sequence of LMW-s-type AY542898 (Özdemir and Cloutier 2005). These primers were used to amplify LMW-GS genes using the BAC clones identified by hybridization and Glenlea genomic DNA as templates.

PCR reactions used either 5–10 ng BAC DNA or 200 ng genomic DNA as template, 1× PCR buffer, 1 mM $MgCl_2$ (to increase the stringency of the reactions), 0.8 mM dNTPs, 0.6 µM of each primer, 0.1 µl of 10× BSA (1 mg/ml) and 1 Unit *Taq* DNA polymerase made up to a final volume of 25 µl with sterile ddH_2O. PCR reactions were carried out in a PTC-100 thermocycler (MJ Research, Inc., Waltham, MA, USA). Following 5 min of denaturation at 94°C, 35 cycles were performed with 30 s at 94°C, 30 s at 60°C, 90 s at 72°C and a final extension step of 10 min at 72°C. PCR products were separated on a 1.5% agarose gel in 0.5× TBE buffer. Gels were photographed using an AlphaImager HP imaging system (Alpha Innotech Corp., San Leandro, CA, USA).

Table 1 LMW-GS gene specific primers designed based on contig consensus sequences of full-length LMW-GS cDNA clones and the genomic clone AY542898

Primer	Primer sequence	Contig no.[a]	T_m (°C)	Product size (bp)
LMWF1	CTTTGCTCTCCTTGCCATTG	8,1,4	60	907
LMWR1	CTGGTGTGGCTGCAAGAATA	8,1,4		
LMWF3	GACAAGTGCCATTGCACAGA	5,3	60	820
LMWR4	GCTGTACAACGGCACATTGA	5,3		
LMWF6	TCCTCGTCTTTGCCCTTCTA	6	60	826
LMWR6	GGCACATTGACACTGCACAT	6		
LMWF7	TTGCGCAGATGGAGACTAGA	2	60	806
LMWR7	CAACGGCACATTAACACTGC	2		
LMWF10	TCACAGCAACAACAACCACA	7	60	784
LMWR9	CTATCTGGTGTGGCTGCAAA	7		
LMWF11	CCAAACTCGGTTGCAAAAGT	AY542898	60	805
LMWR11	TGGTGGTTGTTGCGGTAGTA	AY542898		

[a] Contig number from Özdemir and Cloutier (2005)

Purification and direct sequencing of PCR products

Purification of PCR products was performed using Multi-Screen$_{384}$-PCR filter plates following manufacturer's instructions with some modifications (Millipore Corp., Billerica, MA, USA), as previously described by Huang and Cloutier (2007). Direct sequencing of PCR products was performed according to Huang and Cloutier (2007). Sequencing reactions contained 40–50 ng PCR product as template, 1 µl of 5× sequencing buffer, 1 µl of 5.2 µM primer (primers as used in PCR reactions) and 0.4 µl of BigDye reaction mix in a total volume of 6 µl. The PCR products were sequenced on a 3100 Genetic Analyser (Applied Biosystems).

Chromosome walking using hemi-nested touchdown (TD) PCR

We initially attempted to obtain and/or extend the LMW-GS sequences by primer walking using the BAC clone DNA as template but this approach was not always successful because of the nature of the sequence or the presence of multiple copies on a single BAC clone. To palliate this difficulty, we developed a hemi-nested touchdown (TD) PCR for chromosome walking that enabled us to efficiently perform sequencing of flanking regions from BAC clones (Huang and Cloutier 2007). Primers used for primer walking are listed in Table S1 (Supplementary data). When the first round of TD PCR produced a single strong band, the PCR product was used for direct sequencing. If the first round of TD PCR generated multiple faint bands, a second round of hemi-nested TD PCR was employed to produce a single strong PCR amplicon that was then directly sequenced.

Nucleotide sequence analysis

DNA sequences were processed and assembled using our in-house developed software called SOOMOS v0.6 (T. Banks, personal communication). Base calling, quality assessment and low quality sequence removal were performed with the software PHRED (Ewing et al. 1998) using the '-trim_alt' option. The software CAP3 (Huang and Madan 1999) was used to assemble the processed reads into contiguous sequences (contigs). DNA sequence multiple alignment, translation and identification of open reading frames (ORFs) were conducted using CLUSTAL W v1.82 (Higgins et al. 1994) and DNAMAN v3.2 (Lynnon Corporation, QC, Canada). MEGA3.1 (Kumar et al. 2004) was used to construct a phylogenetic tree using the Neighbor-joining method. Bootstrap tests were performed using 1,000 replications. The rate of nonsynonymous (K_a) versus synonymous (K_s) substitutions was calculated for different domains of LMW-GS genes with DnaSP (Rozas et al. 2003).

Results

BAC library screening, BAC fingerprinting and contig assembly

A total of 383 positive BAC clones were identified by screening of the 24 high-density filters of the 'Glenlea' BAC library with the LMW-GS composite probe, indicating the complexity of LMW-GS gene loci in hexaploid wheat.

Of the 383 BAC clones, 129 clones assembled into 42 contigs. Seventeen contigs consisted of three or more BAC clones, while the remaining contigs had only two BAC clones. The two largest contigs contained seven BAC clones each. However, 254 BAC clones remained as singletons. Using LMW-GS-gene specific primers, 82 BAC clones were confirmed by PCR to contain LMW-GS sequences (see below). Fingerprinting of this subset of clones resulted in the assembly of 30 BAC clones into eight contigs, while 52 BAC clones remained as singletons. A subset of six assemblies is illustrated in Fig. 1. Five contigs had four or more BAC clones (Fig. 1, contig a–e). Contig length was estimated to range between 95 kb for a contig of two BAC clones (Fig. 1, contig f) and 219 kb for a contig of five BAC clones (Fig. 1, contig a).

Identification of BAC clones containing LMW-GS genes

Considering that LMW-GS and gliadin sequences share approximately 60% similarities (D'Ovidio and Masci 2004), positive hybridization signals may include BAC clones containing gliadin sequences. Specific primers were designed from the α-gliadin sequence U08287 to amplify BAC clones assembled in contigs. PCR amplification indicated that BAC clones in the two largest contigs with seven clones also contained gliadin genes (data not shown).

In order to confirm the identity of positive BAC clones containing LMW-GS genes, six LMW-GS-specific primers were designed. After PCR screening, LMWF1/LMWR1 identified only four positive BAC clones, whereas LMWF3/LMWR4 identified 52 positive BAC clones. In total, 82 out of the 383 BAC clones produced one amplicon with at least one of the six primer pairs. Thirty-eight BAC clones were identified by a single primer pair, while 44 BAC clones were identified using two or more primer pairs. Figure S1 (Supplementary data) shows the screening of the BAC clones with primer pair LMWF3/LMWR4 that identified nine positive BAC clones.

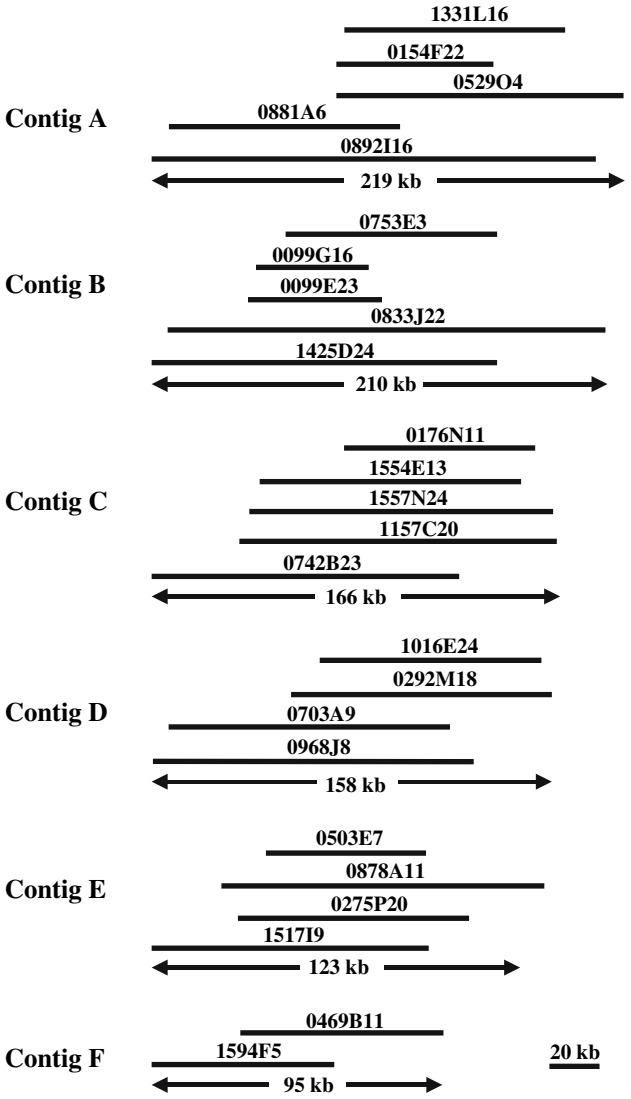

Fig. 1 Diagram illustrating six of the eight wheat BAC contig assemblies containing LMW-GS gene(s)

Isolation and molecular characterization of LMW-GS genes in Glenlea

PCR products obtained using LMW-GS specific primers of the 82 BAC clone subset were directly sequenced. Flanking sequence information was generated until complete open reading frames (ORFs) as well as 5' and 3' untranslated sequence information was obtained for all the different LMW-GS sequences. Sequence analysis revealed that the 82 BAC clone subset contained 12 different LMW-GS genes and seven pseudogenes (Table 2). Of the 82 BAC clones, 46 clones contained a single LMW-GS gene and eight clones had a single LMW-GS pseudogene. The remaining clones contained two or more LMW-GS genes and/or pseudogenes (Supplementary data, Table S2). As expected, identical LMW-GS sequences were obtained from BAC clones belonging to the same contig (data not shown). Identical sequences were also obtained from different BAC clones that remained singletons or that were part of an assembly and present on singletons. Failure to assemble these BAC clones by fingerprinting could be due to their short length or small overlap. Alternatively, they could represent recent duplications, in which case, assembly would likely not be possible due to the nature of the BAC sequences.

Twelve different LMW-GS genes in 11 BAC clones were chosen for further extension of sequences in their 5′ and 3′ non-coding flanking regions using hemi-nested touchdown PCR (Huang and Cloutier 2007). These genes were designated as clones 0072P19-m, 0099E23-m, 0154A5-m, 0154F22-s, 0275P20-m, 0359D24-m, 0359D24-s, 0703A9-m, 0877L13-m, 1238L16-m, 1557N24-m and 1594F5-i, respectively, based on the BAC clone name and the first amino acid residue of the mature protein. The GenBank accession numbers for these sequences are as follows: 0072P19-m, EU189090; 0099E23-m, EU189091; 0154A5-m, EU189092; 0154F22-s, EU189088; 0275P20-m, EU189093; 0359D24-m, EU189094; 0359D24-s, EU189095; 0703A9-m, EU189096; 0877L13-m, EU189097; 1238L16-m, EU189098; 1557N24-m, EU189089; and 1594F5-i, EU189087. Their total sequence length ranged from 1,455 bp for 0072P19-m to 2,755 bp for 1238L16-m. The 5′ non-coding regions ranged in size from 125 bp to 1,325 bp. In this region, the endosperm boxes with the E-motif at the 5′ end and N-motif at the 3′ end of the boxes, CAAT boxes and TATA boxes were identified (Supplementary data, Fig. S2). The E-motif (TGTAAAGT) is conserved and present in all the genes. In the 3′-UTRs of the 12 genes, three polyadenylation signals (AATAAA) were identified. In 0275P20-m, the last polyadenylation signal was AATGAA. Many single nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs) were observed in the coding regions of the genes, and in the 5′ and 3′ non-coding flanking regions as well (Supplementary data, Fig. S2). These LMW-GS genes are intronless and the length of their coding region ranged from 897 to 1,167 bp. The pairwise percentage sequence similarities between the coding regions of the genes are shown in Table S3 (Supplementary data). The largest similarity score was 99.1% between 0099E23-m and 0072P19-m and the lowest similarity score was 72.5% between 1594F5-i and 0703A9-m.

The deduced amino-acid sequences of the 12 genes showed that they all had a single open reading frame (ORF). Proteins encoded by these genes had from 288 amino acid residues for 0154A5-M to 388 amino acid residues for 1594F5-I (Table 3). The predicted proteins of the 12 genes had similar structures to previously characterized LMW-GS. Each comprised four main structural regions,

Table 2 Classification of LMW-GS genes based on their deduced N-terminal amino acid sequence

N-terminal sequence	BAC clones containing LMW-GS genes		BAC clones containing LMW-GS pseudogenes	
	BAC clone id	No. of different sequences	BAC clone id	No. of different sequences
ISQQQQ–	0469B11[a], 1594F5 [a]	1		
METSHIPS	0176N11[a], 0742B23[a], 1157C20[a], 1554E13[a], 1557N24[a], 0356P4, 0634H15, 0877F13, 0879G22, 0879H9, 0880L1, 0887A10, 0892A5, 0895D15, 0899O18, 1113E3, 1296C6, 1619A4	1		
MENSHIPG	0154F22[a], 0529O4[a], 0881A6[a], 0892I16[a], 1331L16[a], 0659B17, 1113E3, 1126H23, 1127G19	1	1468H7	1
METSHIPG	0292M18[a], 0703A9[a], 0968J8[a], 1016E24[a], 0122N17, 0258A13, 0428D3, 0527O8 // 0877L13[b]	2	0881A6, 0881F19, 0893K5	1
METSRVPG	1234D3, 1238L16, 1013M1, 1275F23	1		
METSCIPG	0099E23[a], 0099G16[a], 0753E3[a], 0833J22[a], 1425D24[a], 0082J4, 0196L16, 0199J21, 0201M1, 0212J1, 0323C2, 0634H15, 0644A2, 0753C6, 0828D8, 0879B18, 0879G22, 0879H9, 0880L1, 0892A5, 0898H1, 1013M1, 1156C11, 1296C6, 1619A4 // 0072P19 // 0154A5	3	1013M1	1
METSCISG	0275P20[a], 0503E7[a], 0878A11[a], 1517I9[a], 0122N17, 0199J21, 0212J1, 0356P4, 0428D3, 0496H7, 0704A17, 0742B23, 0877F13, 0878P3, 0879B18, 0881F19, 0885F3, 0887A10, 0893K5, 0893P12, 0895D15, 0898J12, 1275F23, 1490B22	1		
METRCIPG	0359D24	1	0258A13, 0348N15, 0356P4, 0493I20, 0885P5, 1157C20, 1169F2 // 0407K7 // 0412G23, 0425C13 // 0496H7	4
IENSHIPG	0359D24	1		
Total number		12		7

[a] BAC clones assembled into contigs

[b] Underlined BAC clones separated by '//' contained a LMW-GS sequence that differed from the others in the same group

including a 20 amino acid signal peptide (Sig), a short N-terminal region of 13 amino acids, a repetitive domain rich in glutamine and proline residues and a C-terminal domain consisting of three sub-regions of cysteine-rich (I), glutamine-rich (II) and a final conserved domain (III) (Fig. 2). Based on the first amino acid residue of the mature protein, 1594F5-i is LMW-i-type gene with ISQQQQ in the N-terminal domain. 0154F22-s and 0359D24-s are LMW-s-type gene, because they have isoleucine and methionine at position 1, respectively, and the same asparagine at position 3 and should show the N-terminal amino-acid sequence of the mature protein as the LMW-s-type (SHIPG) (Masci et al. 1998; Ikeda et al. 2002). The remaining genes are classified into LMW-m-type gene.

The predicted molecular weight of the 12 deduced amino acid sequences was between 32.7 kDa for 0154A5-M to 44.7 kDa for 1594F5-I. They not only showed a high level of similarity but also displayed unique features. All 12 LMW glutenin subunits contained eight cysteine residues (Fig. 2). Five cysteine residues in their C-terminal domain I and one cysteine in their C-terminal domain III were conserved and present at the same positions (Figs. 2, 3). The first cysteine in N-terminal domain or repetitive domain and the seventh cysteine in C-terminal domain II were not conserved for LMW-i- and -s-type genes.

The average number of synonymous substitutions (K_s) and non-synonymous substitutions (K_a) per site between different domains of the 12 LMW-GS genes is given in Table 4. Different nucleotide substitution rates were observed among the different domains of LMW-GS genes.

BLASTn searches of the coding region sequences of the 12 genes against the NCBI database were performed. The percentage identities and the accession numbers of these genes are listed in Table S4 (Supplementary data). The coding region of the gene 1594F5-i had 100% identity to genes AY453160 (Zhang et al. 2004) and AY831863 of 'Glenlea'. The latter was a full length cDNA sequence (Özdemir and Cloutier 2005). The gene 0154F22-s showed 99%

Table 3 Comparative analysis of the composition of the mature proteins of 12 LMW-GS isolated from 'Glenlea'

LMW-GS	Predicted Mw (kDa)	pI	No of AA[a]	N-terminal Sequence	Number of amino acid (AA) residues				Number of cysteine residues					
					N-terminal domain	Repetitive domain	C-terminal domain	Total	N-terminal domain	Repetitive domain	C-ter domain I	C-ter domain II	C-ter domain III	Total
1594F5-i	44.7	9.04	388	ISQQQQ-	0	186	182	368	0	0	6	1	1	8
1557N24-m	39.8	8.69	350	METSHIPS	13	133	184	330	0	1	5	1	1	8
0703A9-m	41.7	8.68	365	METSHIPG	13	148	184	345	0	1	5	1	1	8
0877L13-m	39.8	8.69	350	METSHIPG	13	133	184	330	0	1	5	1	1	8
1238L16-m	39.9	8.48	351	METSRVPG	13	132	186	331	0	1	5	1	1	8
0099E23-m	33.8	9.03	298	METSCIPG	13	83	182	278	1	0	5	1	1	8
0072P19-m	33.7	8.69	298	METSCIPG	13	83	182	278	1	0	5	1	1	8
0154A5-m	32.7	9.04	288	METSCIPG	13	83	172	268	1	0	5	1	1	8
0275P20-m	34.1	8.16	303	METSCISG	13	86	184	283	1	0	5	1	1	8
0359D24-m	34.6	8.71	304	METRCIPG	13	99	172	284	1	0	5	1	1	8
0359D24-s	40.1	8.48	354	IENSHIPG	13	149	172	334	0	1	5	1	1	8
0154F22-s	39.0	8.91	343	MENSHIPG	13	138	172	323	0	1	5	1	1	8

[a] Including amino acid residues of the signal peptide

identity to gene AB119007, which was also isolated from 'Glenlea' (Maruyama-Funatsuki et al. 2005). Two silent SNPs (A/G) were found at position 120 and 157 (data not shown). The genes 1238L16-m, 0275P20-m and 039D24-m showed 100% identity to two (AY831781, AY831781), four (AY831789, AY831790, AY831791, AY831792) and two (AY831858, AY831859) full length cDNA sequences from Glenlea, respectively (Özdemir and Cloutier 2005).

Chromosomal assignment of LMW-GS genes

The 12 LMW-GS genes were classified into nine groups, which corresponded to nine different N-terminal sequences. There were three different LMW-GS genes in the group with the N-terminal sequence METSCIPG and two different LMW-GS genes in the group with the N-terminal sequence METSHIPG.

To determine chromosomal location of LMW-GS genes with the nine different N-terminal sequences, gene-specific primers were designed (Supplementary data, Table S5) and used to amplify genomic DNA of 'Chinese Spring' (CS) and CS group 1 nulli-tetrasomic lines. LMW-i-type genes amplified by primer pair LMWF10/LMWR9 were confirmed to be located on chromosome 1A. A LMW-m-type gene with the N-terminal sequence METSHIPS and a LMW-s-type gene with the N-terminal sequence (MEN)SHIPG were assigned to chromosome 1B. The other five LMW-m-type genes and a LMW-s-type gene with the N-terminal sequence (IEN)SHIPG were located on chromosome 1D (Fig. 4, Supplementary data Table S5).

Physical distance and phylogenetic analysis of LMW-GS genes

In wheat, Luo et al. (2003) reported that the average length of a contig assembly obtained using the SNaPshot® fingerprinting method, corresponded to 1.23 times the number of CB units (kb) obtained in FPC. Based on the fragment number of individual BAC clones after the SNaPshot® fingerprinting, the length of the 82 BAC clones was calculated. According to the LMW-GS gene number in the 82 BAC clones, the average physical distance between two adjacent LMW-GS genes was estimated to be 81 kb.

The phylogenetic analysis based on the coding region sequences of the 12 LMW-GS genes indicated that the 12 LMW-GS genes were clustered into three clear groups that represent LMW-s, LMW-m and LMW-i-type genes, respectively (Fig. 5a). The LMW-i gene showed greater differences to LMW-s and LMW-m genes with a homology of 81% while the identity between LMW-s and LMW-m genes was 86% (data not shown). On the basis of the deduced amino acids of the 12 LMW-GS genes, similar clusters were observed, but the two LMW-s-type genes, 0154F22-s and

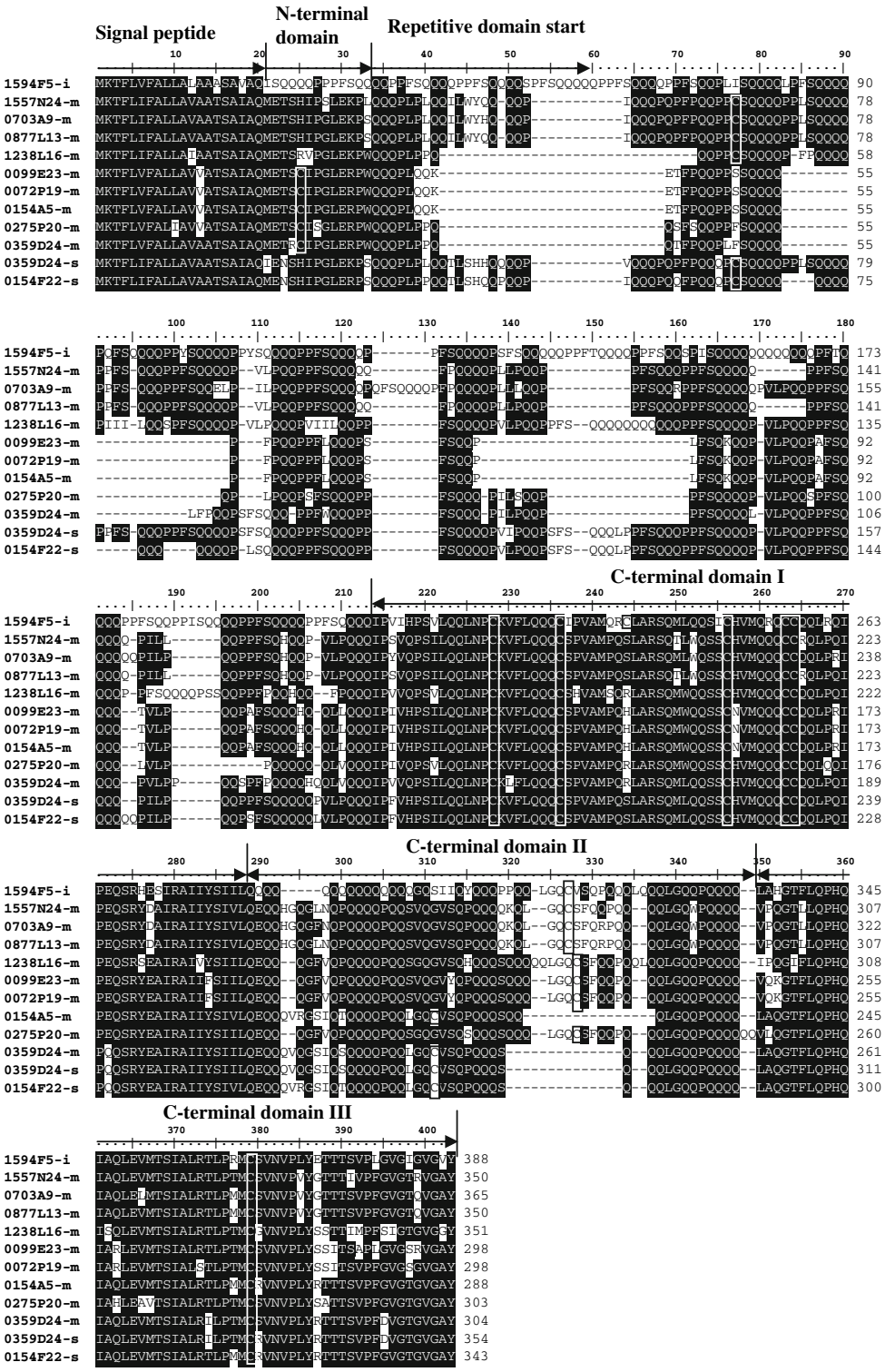

Fig. 2 Alignment of the deduced amino acid sequences of 12 LMW-GS genes from Glenlea. Cysteines are boxed by *white or black lines*. Deletions are indicated by *dashes*

0359D24-s, were closer to three LMW-m-type genes 0703A9-m, 0877L13-m and 1557N24-m (Fig. 5b).

BAC end sequencing

The 82 BAC clones were end-sequenced using M13 forward and reverse primers to obtain partial sequence information of the *Glu-3* locus regions. A total of 162 quality reads with an average size of 450 bp were generated (Supplementary data, Fig. S3). After BLASTn searches against the TREP Database (http://wheat.pw.usda.gov/ITMI/Repeats/blastrepeats3.html), 105 sequences (65%) showed significant similarities to different types of repetitive DNA (cutoff < 10^{-10}, Table 5). Of the 96 LTR retrotransposons

(class I), Sabrina, Fatima and Wham elements were found 24, 11 and 8 times, respectively, suggesting that they are abundant at the *Glu-3* loci. Of the nine class II transposons, eight repetitive elements showed significant matches to the CACTA transposon group.

Discussion

In the present study, we report isolation and sequencing of 12 unique LMW glutenin genes and seven LMW pseudogenes from the hexaploid wheat 'Glenlea' BAC library via filter hybridization, primer-template mismatch PCR and subsequent primer walking using hemi-nested touchdown PCR (Huang and Cloutier 2007). This is the first comprehensive study of LMW-GS genes isolated from a hexaploid wheat BAC library.

Molecular characterization of LMW-GS genes

Cloning of LMW-GS can result in deletions of the repetitive domain (Masci et al. 1998; Ikeda et al. 2002). To avoid this artifact and insure high sequence reliability, we performed direct sequencing of PCR products on a minimum of three independent PCR reactions/reads. The 12 LMW glutenin genes reported here, like all other prolamin genes published to date, do not contain introns. Among the 12 LMW-GS genes, one gene encodes an LMW-i-type subunit, two genes encode LMW-s-type subunits and the remaining nine genes encode LMW-m-type subunits. They were classified into nine different groups based on the N-terminal sequences. There are two different LMW-GS genes in the group with the N-terminal sequence METSHIPG (Table 2). These two genes share 97.1% sequence similarity (Table S3), but have the same C-terminal sequence VGTQVGAY corresponding to that of group 1 (Ikeda et al. 2002). Similarly, three LMW-GS genes belong to the group with the N-terminal sequence METSCIPG and share more than 95.5% similarity (Supplementary data, Table S3). The LMW-GS genes with the N-terminal sequences MDTSCIPG and ISQQQQQP reported by Ikeda et al. (2002) were not identified in this study. The amino acid sequence encoded by the former gene was not detected in the variety Norin 61 using two-dimensional gel electrophoresis (2DE) and N-terminal amino acid sequencing (Ikeda et al. 2006).

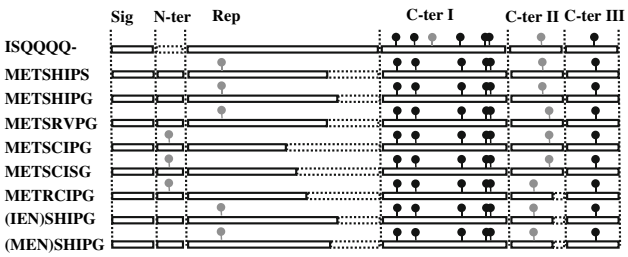

Fig. 3 Structure comparison of the deduced amino acid sequences of nine LMW-GS genes. Cysteine residues and their locations are depicted as pin heads. The *grey* and *black* pin heads indicate cysteines involved in inter-molecular and intra-molecular disulphide bonds, respectively. *Sig* signal peptide, *N-ter* N-terminal domain, *Rep* repetitive domain, *C-ter* C-terminal domain. *Solid line boxes* represent nucleotide sequences of LMW-GS genes while *dotted boxes* indicate deletions

The highest average number of K_s was found in the repetitive domain, followed by C-terminal II. The lowest average number of K_s was found in C-terminal III (Table 4). This indicates large variation in the repetitive domain and sequence conservation in the C-terminal III of LMW-GS genes. The large variation in the repetitive domain reflects the common occurrence of deletions and/or insertions of repeat units. Unequal crossing-over and/or slippage during replication could be responsible for the variations of the repetitive domain as suggested for the evolution of other prolamins (D'Ovidio and Masci 2004).

The predicted proteins of the 12 LMW-GS genes indicated the presence of eight cysteine residues. For the LMW-m- and LMW-s-type genes, it has been proposed that the first and seventh cysteine residues are involved in intermolecular disulfide bond formation and their positions are variable in their respective domains (Shewry and Thatam 1997; D'Ovidio et al. 1999; Masci et al. 1998; D'Ovidio and Masci 2004). For the LMW-m-type genes with the N-terminal sequences METSCIPG, METSCISG and METRCIPG, the first cysteine residue is present in the N-terminus. For the other three LMW-m-type and the two LMW-s-type genes, the first cysteine residue exists in the repetitive domain. The seventh cysteine residue was found in the

Table 4 Average number of synonymous (K_s) and non-synonymous (K_a) substitutions for pairwise comparisons of the 12 LMW-GS complete genes and respective values for the five different sub-domains

	Complete gene	Sequence for signal peptide and *N*-terminal domain	Sequence for repetitive domain	Sequence for C-terminal I	Sequence for C-terminal II	Sequence for C-terminal III
K_a	0.0947	0.0293	0.1408	0.0604	0.0808	0.0988
K_s	0.3166	0.2212	0.4448	0.2906	0.4104	0.1542

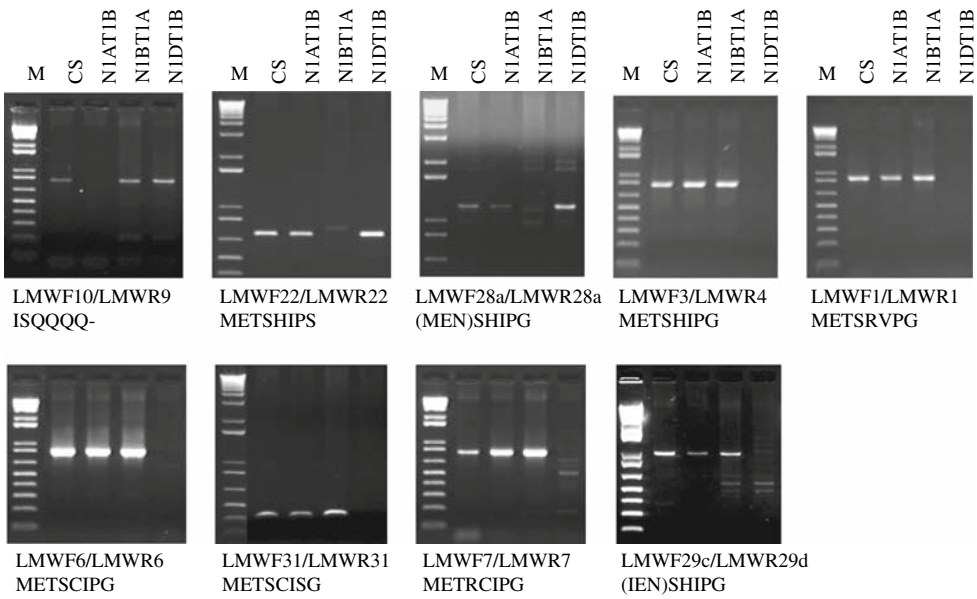

Fig. 4 Chromosomal location of nine different types of LMW-GS genes using group 1 nulli-tetrasomic lines and gene-specific primers

C-terminal II in all LMW-m- and LMW-s-type genes (Figs. 2, 3). For the LMW-i-type gene, the N-terminal region is missing and the eight residues are all present in the C-terminal domain. As proposed by Cloutier et al. (2001) and D'Ovidio and Masci (2004), the third cysteine residue in the C-terminal I of the LMW-i-type gene is likely to participate in intermolecular disulfide bond formation (Fig. 3).

So far, most of the isolated LMW-GS genes from *T. aestivum* and *T. tauschii* were m-type (Johal et al. 2004; Ikeda et al. 2002; Cassidy et al. 1998; Zhao et al. 2006, 2007; this study). Only two LMW-s-type genes with the N-terminal sequences (MEN)SHIPG and (IEN)SHIPG were identified from genomic sequences of wheat (Ikeda et al. 2002; this study). By using 2DE and N-terminal sequencing, Ikeda et al. (2006) confirmed that LMW-m-type genes were predominant in *T. aestivum* variety 'Norin 61'. They also found that a single LMW-GS gene gave rise to two different N-terminal sequences SHIPGLERPS (s-type) and MENSHIPGL (m-type) due to post-translational cleavage by an asparaginyl endoprotease, although the latter was a minor fraction. Lew et al. (1992) reported that the LMW-s-type genes were predominant in *T. aestivum* variety 'Yecora Rojo' based on reverse-phase high-performance liquid chromatography (RP-HPLC) and N-terminal amino acid sequencing. Detailed comparisons of the N-terminal sequences of the LMW-s-type and the LMW-m-type subunits identified from the same fractions (peaks) indicated that the former had the same amino acid sequence as the latter in many cases, if the initial tripeptidyl unit M-E-T of the LMW-m-type subunits was omitted. If this is the case, LMW-s-type subunits are most likely identified as LMW-m-type subunits or derivatives of LMW-m-type subunits. Possible explanations could be that the LMW-s-type genes might be processed post-translationally by a threonine endoprotease in 'Yecora Rojo' or the chemical conditions used during sample preparation could catalyse the deamidation and cyclization of the N-terminal glutamine to pyroglutamic acid (Tao and Kasarda 1989). These assumptions need to be further investigated using 2DE and N-terminal sequencing developed by Ikeda et al. (2006). In fact, only one LMW-s-type genomic sequence with the N-terminal sequence (MEN)SHIPG encoding the 42 KDa LMW-GS was isolated from 'Yecora Rojo' (Masci et al. 1998). Therefore, classification between the LMW-s-type and LMW-m-type genes seems tenuous because LMW-s type genes are originally LMW-m-type genes and LMW-s-type glutenins have very similar primary structure with that of LMW m-type glutenins. In the present study, two LMW-s-type genes indicated a very close relationship at the amino acid sequence level with the LMW-m-type genes, in which the first cysteine residue was in the repetitive domain (Figs. 3, 5).

Among the nine groups of LMW-GS genes, the i-type genes have been found exclusively at the *Glu-A3* locus (Supplementary data Table S5, Fig. 4). This was the case in *T. monococcum*, durum and hexaploid wheats (Ikeda et al. 2002; Wicker et al. 2003; An et al. 2006; this study). It is known that the i-type genes are expressed as protein products (Cloutier et al. 2001; Maruyama-Funatsuki et al. 2004; Ikeda et al. 2006). At least one of them has been shown to play a role in improvement of bread making properties (McCartney et al. 2006). The s-type LMW-GS starting with (MEN)SHIPG and the m-type LMW-GS starting with METSHIPS are encoded at the *Glu-B3* locus (Supplementary data Table S5, Fig. 4). These two LMW-GS types play an important role in determining the pasta-making properties of durum wheat (D'Ovidio et al. 1997; 1999), and the s-type LMW-GS (42K) is associated with good bread-

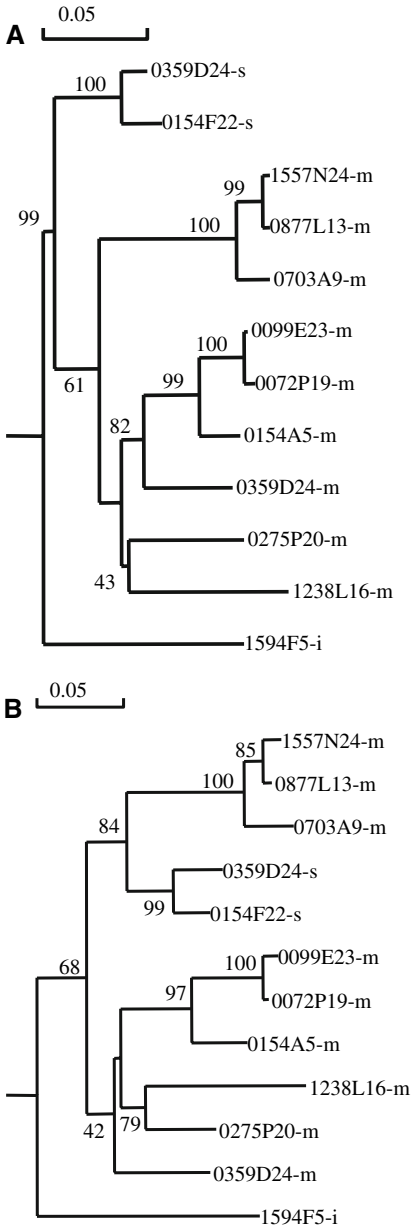

Fig. 5 Phylogenetic trees of 12 LMW-GS genes constructed using MEGA3.1 based (**a**) on the nucleotide sequence of the coding region and (**b**) on the deduced amino acid sequences. Bootstrap values are indicated and bootstrap percentages are based on 1,000 iterations. The *scale bar* indicates the level of sequence divergence

making quality (Masci et al. 1998; Maruyama-Funatsuki et al. 2005). The s-type LMW-GS gene with the N-terminal sequence (IEN)SHIPG and the other five m-type LMW-GS genes are located at the *Glu-D3* locus (Supplementary data Table S5, Fig. 4). This is the second report on the isolation of sequence of this s-type LMW-GS gene in hexaploid wheat after Ikeda et al. (2002). The impact on bread-making quality of individual LMW-GS encoded at the *Glu-D3* locus is unknown. Development of near-isogenic lines (NILs) or transgenic lines with these individual LMW-GS genes may help to elucidate their effects on bread-making quality. Since most of LMW-GS genes are located at the *Glu-D3* locus (Table S5, Ikeda et al. 2006), it is assumed that LMW-GS encoded at this locus could contribute positively to rheological properties.

Table 5 Analysis of 162 BAC end sequences from 82 BAC clones containing LMW-GS sequences

Homology	Name	Number of sequences
Class I retrotransposon		96
LTR family		
Athila	Derami	1
	Egug	4
	Sabrina	24
	Wham	8
Ty1-copia	Angela	4
	Barbara	2
	BARE-1	1
	Georgia	1
	Inga	1
	Maximus	4
	TAR	5
	Usier	1
	WIS	2
Ty3-gypsy	Cereba	1
	Daniela	1
	Erika	2
	Fatima	11
	Gujog	3
	Hawi	3
	Heidi	1
	Ifis	1
	Latidu	1
	Laura	1
	Nusif	2
	Romani	3
	Wilma	3
Unknown	Sukkula	3
Non-LTR	LINE	2
Class II transposon		9
	CACTA	8
	MITE	1
No hits		57

Genomic organization of LMW-GS genes

It has been suggested that the α-gliadin genes have undergone frequent duplications leading to their large numbers in wheat genomes (Gu et al. 2004a). The same could be hypothesized for LMW-GS sequences. Identical LMW-GS

genes were found in many singletons that could not be assembled into contigs (Table 2). The lack of assembly of these BAC clones by fingerprinting could be the result of the small insert size of the BAC library (79 kb, Nilmalgoda et al. 2003) or truly represent non-overlapping genomic regions containing identical LMW-GS genes. The latter is highly probable considering the large number of singletons with identical LMW-GS sequences. These multiple duplication events may have phenotypic implications such as improvement of rheological properties as been shown for HMW glutenin gene duplication (Radovanovic et al. 2002). In the present study, we identified 12 individual LMW-GS genes and 7 pseudogenes. Considering duplicated LMW-GS genes, approximately 30 genes are estimated to be present in the hexaploid wheat genome supporting the findings of Cassidy et al. (1998) who estimated that the LMW-glutenin gene family contains 30–40 members in the hexaploid wheat. Second, *Glu-3* and *Gli-1* loci are adjacent. In fact, genes encoding for LMW-GS and gliadins are interspersed (Gao et al. 2007). Mechanisms leading to duplication of gliadins are likely to result in duplication of LMW-GS sequences as well. Such mechanisms can be driven by the numerous retroelements found at these loci (Table 5).

To date, little is known about the physical organization of the *Glu-A3*, *Glu-B3* and *Glu-D3* loci. Wicker et al. (2003) reported that two LMW-GS genes were separated from each other by 150 kb in *T. monococcum* (A^m genome), whereas Gao et al. (2007) found that two paralogous LMW-GS genes were approximately 100 kb apart. Recent study on LMW-GS genes in the D-genome of *T. tauschii* indicated that large physical distances separated LMW-GS genes because all seven LMW-GS genes in this study were located on different BAC clones (Johal et al. 2004). In the present study, the majority of the BAC clones isolated contained only one LMW-GS (Supplementary data Table S2). On the basis of the estimated BAC clone length using the SNaPshot fingerprinting method (Luo et al. 2003) and the number of LMW-GS sequences identified per BAC clone, the average physical distance between LMW-GS sequences is 81 kb. However, the large number of HMW-GS sequences found on singletons is indicative that this distance would vary greatly. Regarding the large number of LMW-GS genes located in the D-genome (Table S5), we can draw a conclusion that the *Glu-D3* locus is much larger than the *Glu-B3* locus and *Glu-A3* locus.

Recent studies of wheat genome structure and organization indicated that more than 80% of wheat genomic sequences are repetitive DNA. Among the repetitive DNA, long terminal repeat (LTR)-retrotransposons are the most common components and their insertions play a major role in genome expansion (Wicker et al. 2003; SanMiguel et al. 2002; Anderson et al. 2003; Gao et al. 2007). It is known that nested retroelements are present around LMW glutenin loci of the A^m, A and B genomes of wheat (Wicker et al. 2003; Gao et al. 2007). BAC end sequencing of the 82 BAC clones containing LMW-GS gene sequences indicated that repetitive DNA is likely to be the major component at the *Glu-3* loci. The most known retroelements in grasses were found in the BAC end sequences in the present study (Table 5). Of these retroelements, Sabrina, Wham and Fatima were found to be the most abundant ones.

It has been suggested that conservation of orthologous sequences at the *Glu-A3* locus is very limited between the A and A^m genomes (Wicker et al. 2003) and intergenic sequences in the *Glu-3* regions between homoeologous A and B genome of durum wheat are not conserved (Gao et al. 2007). We performed BLASTn searches against the NCBI database using the complete LMW-GS gene sequences including the 5'- and 3'-UTR regions. The gene 1594F5-i located on the A genome of hexaploid wheat shows strong sequence identity (99%, $E = 0.0$) to the gene AY146587 (i-type), which is located on the A genome of durum wheat cv Langdon (Wicker et al. 2003). Alignment of 2,269 bp sequences indicates one and four SNPs in the 5' UTR regions and ORF regions, and identical sequences in the 3'-UTR region. The gene 1557N24-m of the B genome of hexaploid wheat has high sequence similarity (98%, $E = 0.0$) to the gene Y14104 (m-type) in the B genome of durum wheat cv Langdon (D'Ovidio et al. 1997). Similarly, the remaining m-type LMW-GS genes shows strong conservation (96–99%, $E = 0.0$) to the genes AY585350, AY585351, AY585354, AY585355 and AY585356, isolated from *T. tauschii* (Johal et al. 2004), the D genome donor of hexaploid wheat. These results indicate that the same type (starting with the same N-terminal sequence) LMW-GS genes are highly conserved in the homologous A (B) genomes between durum and hexaploid wheats, as well as in the homologous D genomes between hexaploid wheat and *T. tauschii*. It would be interesting to determine whether intergenic sequences in the *Glu-3* regions of three homologous genomes are conserved between hexaploid wheat and its progenitors.

Acknowledgments We would like to thank Travis Banks for bioinformatics support, and Elsa Reimer and Andrzej Walichnowski for technical assistance. X.Q. Huang received a visiting fellowship from the National Science and Engineering Research Council of Canada (NSERC). This research was supported by the Canadian Crop Genomics Initiative (CCGI). This is Agriculture and Agri-Food Canada publication number 1962.

References

An X, Zhang Q, Yan Y, Li Q, Zhang Y, Wang A, Pei Y, Tian J, Wang H, Hsam SLK, Zeller FJ (2006) Cloning and molecular characterization of three novel LMW-i glutenin subunit genes from cultivated einkorn (*Triticum monococcum* L.). Theor Appl Genet 113:383–395

Anderson OD, Rausch C, Moullet O, Lagudah ES (2003) The wheat D-genome HMW-glutenin locus: BAC sequncing, gene distribution, and retrotransposon clusters. Funct Integr Genomics 3:56–68

Cassidy BG, Dvorak J, Anderson OD (1998) The wheat low-molecular weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. Theor Appl Genet 96:743–750

Cloutier S, Ramptisch C, Penner GA, Lukow OM (2001) Cloning and expression of a LMW-i glutenin gene. J Cereal Sci 33:143–154

Colot V, Bartels D, Thompson R, Flavell R (1989) Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol Gen Genet 216:81–90

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

D'Ovidio R, Simeone M, Masci S, Porceddu E (1997) Molecular characterization of a LMW-GS gene located on chromosome 1B and the development of primers specific for the *Glu-B3* complex locus in durum wheat. Theor Appl Genet 95:1119–1126

D'Ovidio R, Marchitelli C, Cardelli LE, Porceddu E (1999) Sequence similarity between allelic *Glu-B3* genes related to quality properties of durum wheat. Theor Appl Genet 98:455–461

Ewing B, Hillier L, Wendel MC, Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:175–185

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

Gianibelli MC, Larroque OR, Macritchie F, Wrigley CW (2001) Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. Cereal Chem 78:635–646

Gu YQ, Crossman C, Kong X, Luo M, You FM, Coleman-Derr D, Dubcovsky J, Anderson OD (2004a) Genomic organization of the complex α-gliadin gene loci in wheat. Theor Appl Genet 109:648–657

Gu YQ, Coleman-Derr D, Kong X, Anderson OD (2004b) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four Triticeae genomes. Plant Physiol 135:459–470

Gu YQ, Salse J, Coleman-Derr D, Dupin A, Crossman C, Lazo GR, Huo N, Belcram H, Ravel C, Charmet G, Charles M, Anderson OD, Chalhoub B (2006) Types and rates of sequence evolution at the high-molecular-weight glutenin locus in hexaploid wheat and its ancestral genomes. Genetics 174:1493–1504

Gupta RB, Shepherd KW (1990) Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutenin: 1. Variation and genetic control of the subunits in hexaploid wheats. Theor Appl Genet 80:65–74

Gupta RB, Singh NK, Shepherd KW (1989) The cumulative effect of allelic variation in LMW and HMW glutenin subunits on dough properties in the progeny of two bread wheats. Theor Appl Genet 77:57–64

Gupta RB, Paul JG, Cornish GB, Palmer GA, Bekes F, Rathjen AJ (1994) Allelic variation at glutenin subunit and gliadin loci, *Glu-1*, *Glu-3* and *Gli-1*, of common wheats. I. Its additive and interaction effects on dough properties. J Cereal Sci 19:9–17

Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Huang XQ, Cloutier S (2007) Hemi-nested touchdown PCR combined with primer-template mismatch PCR for rapid isolation and sequencing of low molecular weight glutenin subunit gene family from a hexaploid wheat BAC library. BMC Genetics 8:18

Ikeda TM, Nagamine T, Fukuoka H, Yano H (2002) Identification of new low-molecular-weight glutenin subunit genes in wheat. Theor Appl Genet 104:680–687

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

Johal J, Gianibelli MC, Rahman S, Morell MK, Gale KR (2004) Characterization of low-molecular-weight glutenin genes in *Aegilops tauschii*. Theor Appl Genet 109:1028–1040

Kong X, Gu YQ, You FM, Dubcovsky J, Anderson OD (2004) Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. Plant Mol Biol 54:56–69

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

Lew EJL, Kuzmicky DD, Kasarda DD (1992) Characterization of low molecular weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino acid sequencing. Cereal Chem 69:508–515

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

Luo C, Giffin WB, Branlard G, McNeil L (2001) Comparison of low and high molecular-weight wheat glutenin allele effects on flour quality. Theor Appl Genet 102:1088–1098

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378–389

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

Maruyama-Funatsuki W, Takata K, Funatsuki H, Tabiki T, Ito M, Nishio Z, Kato A, Saito K, Yahata E, Saruyama H, Yamauchi H (2005) Identification and characterization of a novel LMW-s glutenin gene of a Canadian Western Extra-Strong wheat. J Cereal Sci 41:47–57

Masci S, D'ovidio R, Lafiandra D, Kasarda DD (1998) Characterization of a low-molecular-weight glutenin subunit gene from bread wheat and corresponding protein that represents a major subunit of the glutenin polymer. Plant Physiol 18:147–1158

McCartney CA, Somers DJ, Lukow O, Ames N, Noll J, Cloutier S, Humphreys DG, McCallum BD (2006) QTL analysis of quality traits in the spring wheat cross RL4452 × 'AC Domain'. Plant Breed 125:565–575

Nieto-Taladriz MT, Perretant MR, Rousset M (1994) Effect of gliadins and HMW and LMW subunits of glutenin on dough properties in F_6 recombinant inbred lines from a bread wheat cross. Theor Appl Genet 88:81–88

Nilmalgoda S, Cloutier S, Walichnowski AZ (2003) Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L.) and validation of genome coverage using locus-specific primers. Genome 46:870–878

Özdemir N, Cloutier S (2005) Expression analysis and physical mapping of low molecular-weight glutenin loci in hexaploid wheat (*Triticum aestivum* L.). Genome 48:401–410

Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on breadmaking quality. Ann Rev Plant Physiol 38:141–153

Radovanovic N, Cloutier S, Brown D, Hymphreys DG, Lukow OM (2002) Genetic variance for gluten strength contributed by high molecular weight glutenin proteins. Cereal Chem 79:843–849

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct Integr Genomics 2:70–80

Shewry PR, Thatam AS (1997) Disulphide bonds in wheat gluten proteins. J Cereal Sci 25:207–227

Singh NK, Shepherd KW (1988) Linkage mapping of genes controlling endosperm storage proteins in wheat: 1. Genes on the short arms of group 1 chromosomes. Theor Appl Genet 75:628–641

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

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

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

Zhang W, Gianibelli MC, Rampling LR, Gale KR (2004) Characterization and marker development for low molecular weight glutenin genes from *Glu-A3* alleles of bread wheat (*Triticum aestivum* L). Theor Appl Genet 108:1409–1419

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

Zhao XL, Xia XC, He ZH, Lei ZS, Appels R, Yang Y, Sun QX, Ma WJ (2007) Novel DNA variations to characterize low molecular weight glutenin *Glu-D3* genes and develop STS markers in common wheat. Theor Appl Genet 114:451–460