



# Asian Journal of **Biochemistry**

ISSN 1815-9923



Academic  
Journals Inc.

[www.academicjournals.com](http://www.academicjournals.com)

## **Molecular Classification of Members in LMW Glutenin Family of Wild Emmer Wheat (*Triticum dicoccoides*, AABB)**

<sup>1,2</sup>Li-Juan Yang, <sup>1</sup>Ya-Xi Liu, <sup>3</sup>Bi-Ling Xu, <sup>4</sup>Wei Li and <sup>1</sup>Guo-Yue Chen

<sup>1</sup>Triticaceae Research Institute, Chengdu Academy of Science, Sichuan Agricultural University, Wenjiang, Sichuan 611130, China

<sup>2</sup>Xinxiang Institute of Agriculture, Xinxiang, Henan 453600, China

<sup>3</sup>Crop Institute, Sichuan Agricultural Academy of Science, Chengdu, Sichuan 610066, China

<sup>4</sup>College of Agronomy, Sichuan Agricultural University, Yaan, Sichuan 625014, China

*Corresponding Author: Guo-Yue Chen, Triticaceae Research Institute, Sichuan Agricultural University, Wenjiang, Sichuan 611130, China Tel: 86-28-82650337 Fax: 86-28-82650350*

### **ABSTRACT**

One hundred and thirteen low-molecular-weight glutenin subunits encoding sequences and parts of upstream were characterized from *Triticum dicoccoides*. The encoded proteins of 113 genes had similar structures to previously characterized LMW-GS. These sequences had 856~1402 nucleotides in length with 2~26 repeat motifs. Most of the sequences were typical LMW-m glutenin subunits genes and the frequency of SNPs was 1.6 out of 10 bases and A-G mutation was the most frequent. Fourteen deduced amino acid sequences were found to be possessed with an additional cysteine residue in C-ter I. Thirty-six haplotypes were detected and phylogenetic analysis indicated that the 36 haplotypes could be classified into 3 haplotype groups. Individually classifications based on the four main domains of LMW-GS DNA sequences, 5'flanking, single peptide, N-terminals and C-terminal, were in agreement with the classification based on the coding regions. Consilient evolution was found between domains of LMW-GS as well as each domain and the whole coding region. The results revealed the important information of low-molecular-weight glutenin subunit gene family and contributed to our understanding of functional aspects of the low-molecular-weight glutenin subunit genes.

**Key words:** *Triticum dicoccoides*, LMW-GS, classification, viscoelasticity

### **INTRODUCTION**

Wild emmer wheat *T. dicoccoides* (*Triticum turgidum* var. *dicoccoides* Koern., AABB, 2n = 4x = 28) is the tetraploid and predominantly self-pollinated progenitor of cultivated wheats (Zohary, 1970). It is the donor of the A and B genome of common wheat and with which they can make fertile hybrids. It is a valuable source of genes for diverse agronomic traits, photosynthetic yield, high quality proteins and amino acids and for tolerance against abiotic and biotic stresses (Nevo, 1983, 1989, 1995, 2001), disease, insect and biotic stress resistances and also quality traits (Nevo and Payne, 1987). Abundant genetic variability of glutenins between and within wild emmer groups revealed a highlighted relationship between patterns of genetic diversity and ecogeographical factors, but little variability were found in cultivar wheat (Pagnotta *et al.*, 1995; Nevo and Payne, 1987).

Glutenins are the major protein fraction in the endosperm of bread wheat and related species. They mainly consisted of High Molecular Weight Glutenin Subunits (HMW-GSs) and Low Molecular Weight Glutenin Subunits (LMW-GSs). LMW-GS represent about one-third of total seed storage proteins and 60% of glutenins (Bietz and Wall, 1973).

The structures of typical LMW-GSs were known as several regions as deduced from their encoding genes: signal peptide, N-terminal region, repetitive domain (small boxes indicate repeat motifs), C-terminal regions (D'Ovidio and Masci, 2004). On the basis of N-terminal amino acid sequences, three LMW-GS subgroups can be recognized as LMW-s, LMW-m and LMW-i types, according to the first amino acid residue of the mature protein: serine, methionine, or isoleucine, respectively. Long *et al.* (2005) have classified 69 known low-molecular-weight glutenin subunit (LMW-GS) genes into 9 groups by the deduced amino acid sequence of the highly conserved N-terminal domain.

Many results showed that glutenins determine the nutritional and processing properties of bread wheat, especially in dough viscoelasticity (Shewry *et al.*, 1992). Low molecular weight glutenins are closely related to dough resistance and extensibility and also play an important role in determining wheat flour properties (Payne, 1987; Gupta *et al.*, 1994; Masci *et al.*, 1998; D'Ovidio *et al.*, 1999). Though many LMW-GS gene sequences have been reported, comprehensive research on the LMW-GS gene family is very essential to realize the relationship between the gene family and wheat quality.

## **MATERIALS AND METHODS**

**Plant materials:** The accessions of *T. dicoccoides* used in this study were kindly provided by Triticeae Research Institute of Sichuan Agricultural University, China which was originally obtained from Israel.

**DNA isolation and PCR amplification:** Genomic DNA was isolated from 5 g of leaves from single plant. A pair of primers (F4:CCAAAAGTACGCTTGTAGCTAGT R3: TTTCTTAT CAGTAGA /GCA CCAACTCC) was designed to amplify the upstream coding regions and downstream based on previously cloned LMW glutennin gene sequences. The forward primer was at the -210 site of start site and the reverse primer was in the end site of C-terminal domain. Therefore, the PCR products should contain the whole N-terminal and repetitive domain. Owing to the high sequence identity in conserved domains of LMW-GS genes, an additional primer-template mismatched base was added at position 11 of 3'end of R3 primer sequence following the strategy described by Kwok *et al.* (1990).

Polymerase Chain Reaction (PCR) amplifications were performed in 25  $\mu$ L reaction volume, consisting of 1 U Taq DNA polymerase (TaKaRa), 2.5  $\mu$ L PCR buffer (supplied with Taq DNA polymerase), 50 ng genomic DNA, 1.5 mM MgCl<sub>2</sub> and 100 mM of each dNTP. PCR amplifications were conducted according to the following program: 95°C for 5 min denaturation followed by 40 cycles of 45 sec at 94°C, 50 s at 56°C and 1 min at 72°C. PCR products were separated in 1.5% agarose gels.

**Molecular cloning and DNA sequencing:** The amplification products of about 1.1 kb were purified from the gels using E.Z.N.A. Gel Extraction kit (OMEGA). Subsequently purified products were ligated into pMD18-T Easy vector (Promega) and transformed into cells of *E. coli* JM109 strain. The recombinant colonies were analyzed to verify the presence of the 1.1 kb fragment using

PCR amplification with universal primers M13. Then these white clones which had inserted fragments consist with expected fragments in length were chosen and DNA sequences were using primer walking performed by Shanghai Sangon company.

**Identification of SNPs, haplotypes and phylogenetic analysis:** Multiple alignment of LMW glutenin nucleotide and protein sequences were completed by DNAMAN 6.0 (Lynnon Biosoft). The repetitive domain and some microsatellite were wiped off when we identified SNPs and haplotypes. The repetitive domain and some microsatellite were wiped off to concentrated on the conservative domains when phylogenetic analysis of coding domain. MEGA program (Kumar *et al.*, 2001) was used to setup phylogenetic trees based on clustering of several domains of nucleotide sequences using UPGMA (unweighted pair group method with arithmetic mean) method.

## RESULTS

**Nucleotide and deduced amino-acid sequences:** One hundred and thirteen nucleotide sequences have been cloned by the genomic DNA from 24 accessions. Through the BLAST software on the web site of NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>), they were confirmed as the similar sequences of LMW-GS genes. All of the 113 sequences were submit to NCBI Gene Bank. Each of them consisted of 200 or 201 bp upstream and a single complete Open Reading Frame (ORF) and parts of the promoter sequences were shown in Fig. 1.

The coding regions of these genes were all terminated by double stop codons like other prolamin genes. Each one of the encoded proteins of 113 genes comprised four or three main structural regions including a 20 amino acid signal peptide, a short (or missing) N-terminal region, a repetitive domain rich in glutamine and proline residues and a C-terminal domain.



Fig. 1: Comparison of the 5' flanking DNA sequences of several sequences. TATA box and CAAT box were boxed by lines

The longest sequence was 1402 bp and the shortest one was just 856 bp, in size with the numbers of the deduced amino acid residues ranging from about 285 to 467.

According to the method of Long *et al.* (2005), 113 sequences were classified into five groups (i METSCIP-, ii MDTSCIP-, iii METSHIP-, iv METICNPS-, v ISQQQ-). Most of the sequences were typical LMW-m typed glutenin subunits genes, with a methionine started N-terminal of deduced amino acid residues. The other 15 sequences were ascribed to LMW-i typed glutenin subunits genes (group v, isoleucine started). Among five groups, group i was found to be the biggest in quantity and 72 sequences fell into this group. Group ii included 21 and group iii included 4. Notably, sequence 10-2 had a exceptional N-terminal of METICNPS- not reported previously. At least 108 sequences came from 1AS chromosome based on Long's method.

**SNPs and haplotypes:** Identify of the deduced amino-acid sequences without varied repetitive domains calculated by the tool of multiple sequences alignment of DNAMAN was only 46.96% and more than 300 sites were polymorphic. The frequency of SNPs was about 1.6 out of 10 bases. Most of SNPs distributed in the vastest C-terminal region. Among 6 sorts of mutations, A-G transition were the most frequent. We determined the position of each SNP and found about half of SNPs occur at the third codon position and most of them were synonymous.

Palpable diversity among different groups were found though the result of multiple sequences alignment especially between the LMW-i typed and LMW-m typed (data not shown). Many group-specific SNP and In/Del sites were detect in 19 sequences attributed to group iii and group v, so we exclude them to analysis the distributing of SNP sites in most sequences.

Then, 91 variable positions defining 27 haplotypes (H01-H27) were detected (Fig. 2) among the 855 nucleotides of the rest 94 sequences. Site 482, concerns the substitution of Arginine by Cysteine, was located in C-terminal domain and brought the mutants contained one more Cysteine in deduced protein than the others and may put up change in secondary and three-dimensional structure of low molecular weight glutenin subunits.

We examined the 15 sequences ascribed to group v individually and found 55 variable positions among 842 nucleotides. 8 haplotypes (H28-H35, Fig. 3) were detected. The 42 SNPs in coding region are predicted to result in 39 amino acid changes. Four remained sequences belong to group iii were ascribed to a same haplotype (H36, data not shown).

Thus, 36 haplotypes were detected through individual analysis. 36 DNA sequences were chosen as delegates of 36 haplotypes and they were used for alignment using DNAMAN (not shown). Phylogenetic analysis indicated that the 36 haplotypes could be classified into 3 haplotype groups (Fig. 4) which tallied with the preamble classification based on N-terminal of deduced amino acid sequences.

**Gene classification:** MEGA program (Ver. 2, Kumar *et al.*, 2001) was used to setup phylogenetic trees based on clustering of several domains of nucleotide sequences using UPGMA method. The whole cloned sequences, from -210bp to the end of coding region, about 1100 bp in lengths, were classified into eight group (w1-w8, Table 1, column 5). I-typed LMW-GS were clearly distinguished out and attributed to group w1 and group w2. Group w3-w5 and group w7-w8 were clearly diversified by different N-terminals. A single sequence 10-2 with a special N-terminal as METICNPS- was its own group w5.



Table 1: The whole list of the results of several methods of classification

	Chro-Loc	Group by Long	Deduced N-ter	w1-w8	r1-r7	c1-c7	s1-s7	5' flanking	Single peptide	C-ter	N-ter
1-4	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
1-5	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
1-6	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
1-7	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
1-8	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
1-9	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
1-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
1-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
15-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
23-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
26-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
45-8	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
46-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
46-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
46-6	A	i	METSCIP	w6	r7	c7	s8	2	4	4	7
51-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
51-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
51-4	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
52-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
52-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
52-3	A	i	METSCIP	w8	r7	c7	s7	5	4	4	8
52-4	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
52-5	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
52-6	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
52-7	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
52-8	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
61-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
61-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
61-4	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
61-5	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
61-11	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
62-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
62-5	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
62-7	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
70-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
70-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
71-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
192-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-4	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-6	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-7	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-8	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-9	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-10	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-11	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
92-12	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
101-4	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7

Table 1: Continued

	Chro-Loc	Group by Long	Deduced N-ter	w1-w8	r1-r7	c1-c7	s1-s7	5' flanking	Single peptide	C-ter	N-ter
105-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
105-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
105-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
131-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
131-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
132-1	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
132-2	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
132-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
141-3	A	i	METSCIP	w8	r7	c7	s8	5	4	4	7
61-1	A	i	METSCIP	w7	r7	c7	s8	5	4	4	7
102-2	A	i	METSCIP	w7	r7	c7	s8	5	4	4	7
102-3	A	i	METSCIP	w7	r7	c7	s8	5	4	4	7
3-Oct	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
23-2	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
23-3	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
46-2	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
46-4	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
46-5	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
46-7	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
46-8	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
61-6	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
61-7	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
61-9	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
61-10	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
62-1	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
62-3	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
62-4	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
62-6	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
144-1	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
142-1	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
101-3	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
102-1	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
102-4	A	ii	MDTSCIP	w7	r7	c7	s8	5	4	4.1	7
92-1	A	i	METSCIP	w8	r7	c7	s8	5	4	3	7
101-1	A	i	METSCIP	w4	r5	c5	s5	5	4	4	5
14-3	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
14-4	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
15-3	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
45-5	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
92-5	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
92-13	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
93-1	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
93-2	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
93-3	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
93-4	A	i	METSCIP	w4	r5	c5	s5	4	4	3	5
14-1	A	v	ISQQQ	w2	r2	c3	s2	2	2	5	2
14-2	A	v	ISQQQ	w2	r2	c3	s2	2	2	5	2
101-5	A	v	ISQQQ	w2	r2	c3	s2	2	2	5	2
26-1	A	v	ISQQQ	w2	r2	c3	s2	2	1	5	2



Table 1: Continued

	Chro-Loc	Group by Long	Deduced N-ter	w1-w8	r1-r7	c1-c7	s1-s7	5' flanking	Single Peptide	C-ter	N-ter
45-1	A	v	ISQQQ	w2	r2	c3	s2	2	1	5	2
61-8	A	v	ISQQQ	w2	r2	c3	s2	2	1	5	2
131-4	A	v	ISQQQ	w2	r3	c3	s3	2	2	5	3
51-3	A	v	ISQQQ	w1	r2	c3	s2	2	2	5	2
93-5	A	v	ISQQQ	w1	r2	c3	s2	2	2	5	2
45-3	A	v	ISQQQ	w1	r2	c3	s2	2	1	5	2
45-4	A	v	ISQQQ	w1	r2	c1	s2	2	1	5	2
141-2	A	v	ISQQQ	w1	r2	c1	s2	2	1	5	2
131-1	A	v	ISQQQ	w1	r2	c1	s2	2	1	5	2
45-6	A	v	ISQQQ	w1	r2	c2	s2	2	1	5	2
45-7	A	v	ISQQQ	w1	r1	c2	s1	2	1	5	1
15-2	B	iii	METSHIPS-	w3	r4	c4	s4	1	3	1	4
45-2	B	iii	METSHIPS-	w3	r4	c4	s4	1	3	1	4
27-2	B	iii	METSHIPS-	w3	r4	c4	s4	1	3	1	4
141-1	B	iii	METSHIPS-	w3	r4	c4	s4	1	3	1	4
10-2		Others	METICNPS	w5	r6	c6	s6	3	4	2	6

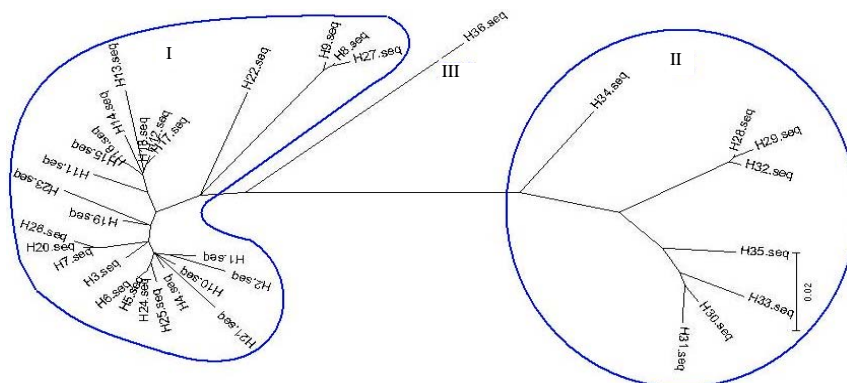


Fig. 4: Phylogenetic tree based on the alignment of 36 sequences including 5' flanking and coding regions, H01-H36 indicated the all 36 haplotypes. Numbers I, II, III in the blank is the serial number of haplotype-groups. Group I: H1-H27; Group II: H28-H35 and Group III consist just a single haplotype of H36

The sequences without 5'flanking and repetitive domains, about 840 bp in lengths, were classified into seven groups (r1-r7, Table 1, column 6) and the result revealed the similarity between group i and group ii were mostly in the person of length. The coding regions without single peptide, about 750 bp in lengths, were classified into eight groups (c1-c7, Table 1, column 7).

The coding regions without single peptide and repetitive domains, about 690 bp in lengths, were classified into eight groups (s1-s7, Table 1, column 8).

Sequence 52-3 was count out by SNPs. The 5'flanking sequences and single peptide and the N-terminals as well as C- terminal of 113 LMW-GS DNA sequences were separately taken into multiple alignments. Thus, four phylogenetic trees of those domains were built, respectively (Fig. 5-8).

All of the classification results were summed up in Table 1. Phylogenetic analysis based on the highly conserved N-terminal indicated that the 113 sequences could be divided into five groups (Table 1, column 12). Whereas, the classification based on the deduced amino acids residues of N-terminal domain (Table 1, column 4) were differed from above-mentioned. Twenty-one sequences with MDTSCIP-N-terminal were classified as the 4.1 group (Fig. 7) to interpret the inconsistent.

## DISCUSSION

All LMW-GS genes cloned in present study showed high homology with LMW-GS genes of other *Triticum* species. LMW-GS genes cloned in this study showed high diversity besides high homology with foregone LMW-GS genes what was attest by the various lengths and different N-terminal types and disparate locations of cysteine residues. One LMW-GS gene, AB209929, shorter than common was found though blasting in NCBI gene banks. Its length was 536 bp. The shortness of AB209929 could be attributed to the lack of C-terminal (Tanaka *et al.*, 2005). There was no homothetic genes could be found in gene banks shorter than 45-6, the shortest sequence cloned in present study.

Single Nucleotide Polymorphisms (SNPs) are the most frequent variations in the genome of any organism. The frequency of SNPs calculated in this study is higher than that in HMW-GS (Lu *et al.*, 2005), probably due to the fact that the total gene copy number of LMW-GS is highly variable from 10-15 (Harberd *et al.*, 1985) to 35-40 (Cassidy *et al.*, 1998) and the loci were more complex. Though the SNPs detected in LMW-GS may provide new tools for further insights into the mechanisms of quality variations, it should be study more to be used as reliable genetic markers for the complex classification and chromosome assignment of the LMW-GS gene.

The lengths of LMW-i typed genes varied intensely relative to LMW-m typed genes. The shortest and the longest of LMW-i typed genes, 45-6 and 141-2, contained 2 and 26 repeat motifs, respectively. That indicated that number of repeats present in the repetitive domain is mainly responsible for length variation which accord with the view of D'Ovidio and Masci (2004).

There were few differences between the results of phylogenetic analysis based on pristine sequences and repeat-domains-wiped sequences. At the same time, different types LMW genes particularly with distinct N-terminals were classified into distinct groups. That indicated the repeat-domain contributed rarely to the diversity of whole sequence but the type of LMW genes played an important role in classification.

Among all of the 113 sequences in this study, group i, with the N-terminal of METSCIP, was found to be the biggest in quantity. Reversely, group v, with the N-terminal of ISQQQ- were less and no LMW-s typed genes were cloned by random. Which type or which group LMW-GS gene is the most popular in wild emmer wheat need more evidence to make certain.

Deduced amino acids of 14 sequences of H2, which resulted from a SNP variation (T-C transition) resulting in arginine-cysteine substitution, possessed nine cysteine residues with one more cysteine in C-ter I than the others (Fig. 9). The different distribution and the extra number of cysteine residues could lead to functional differences because the secondary and three-dimensional structures of LMW-GS would be quite different from the other subunits. Nine cysteines, one more than typical eight, has also been reported in the LMW-m gene AY263369 that is inferred to associate with the good properties of Chinese bread wheat cultivar Xiaoyan 6 (Zhao *et al.*, 2004). The additional cysteine residue present in subunits of H2 would promote a differential cross-linking and endow dough with increased strength and superior performance.

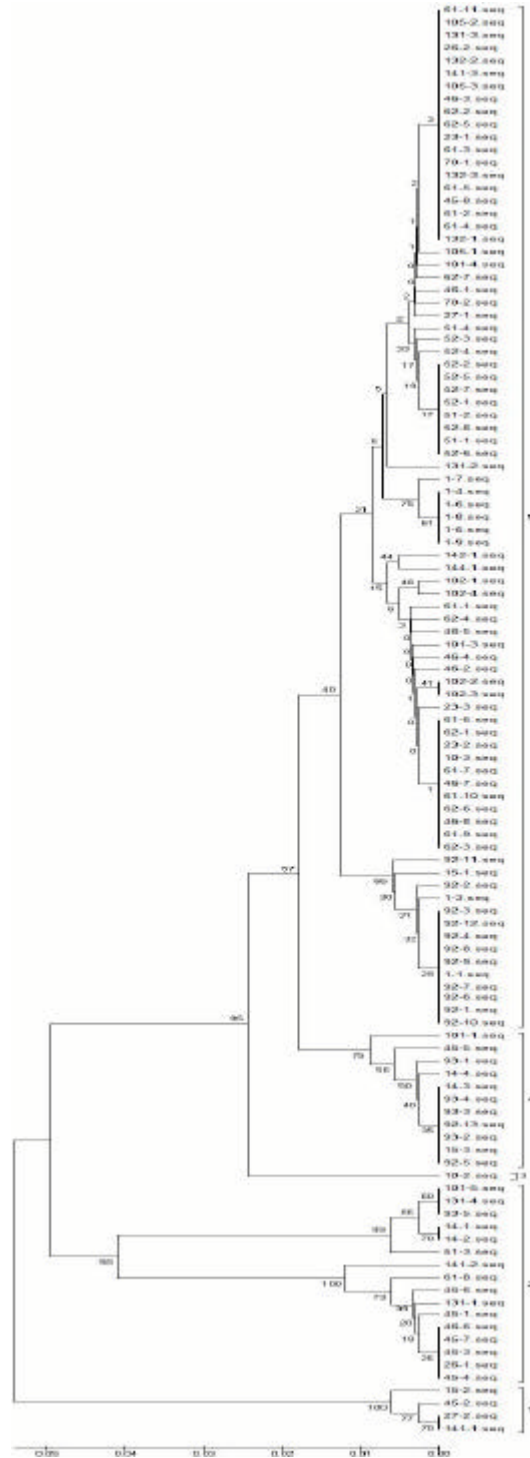


Fig. 5: Phylogenetic tree based on the 5'flanking sequences,number1-5 on the right of square bracket was the serial number of groups, this point will not be prompted in sequel

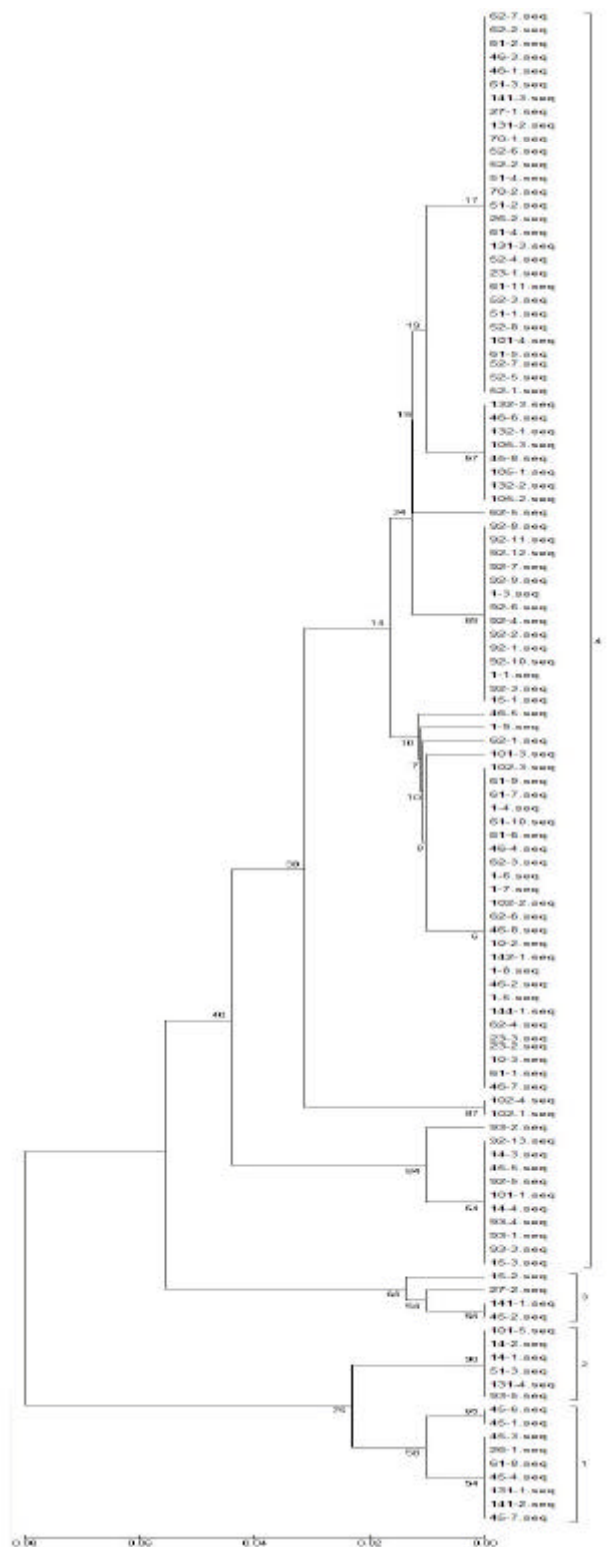


Fig. 6: Phylogenetic tree based on single peptide sequences

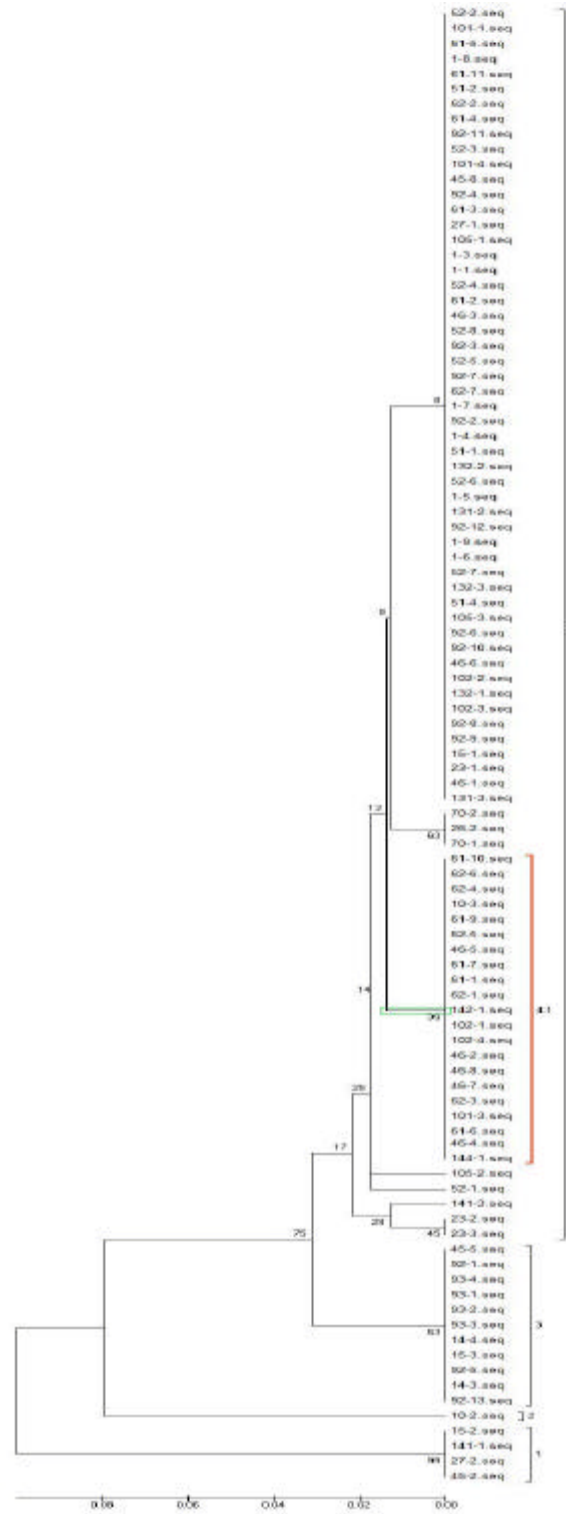


Fig. 7: Phylogenetic tree based on N-terminal sequences, the group 4.1 were included in group 4

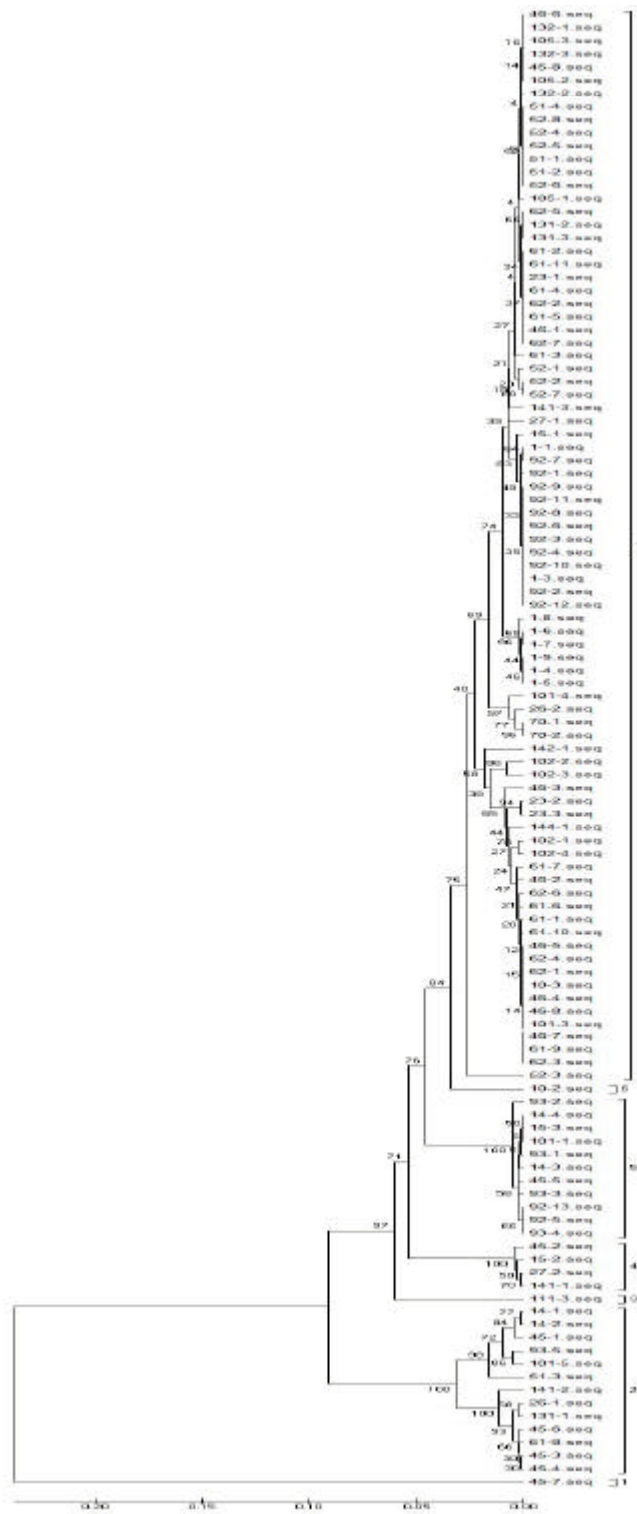


Fig. 8: Phylogenetic tree based on C-terminal sequences

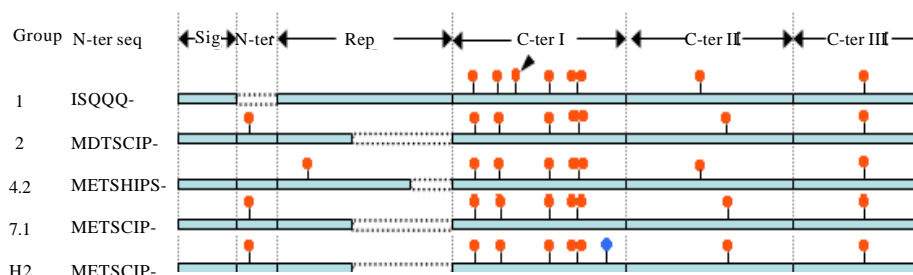


Fig. 9: Schematic model of the structure comparison of each group (based on the deduced N-terminal domain) of LMW-GS genes. Cysteine residues and their locations were shown as bullethead. Sig Signal peptide; N-ter N-terminal domain; Seq sequence; Rep repetitive domain; C-ter C-terminal domain. The bullethead with different color figure the location of arginine-cysteine substitution

Therefore, it could be considered that the 14 genes of H2 is a new candidate LMW glutenin gene that may have important effect on wheat quality.

Individually classifications based on the four main domains of LMW-GS DNA sequences, 5' flanking, single peptide, N-terminals and C-terminal, were in agreement with the classification based on the coding regions despite that there were several sequences mingled. These results not only attest to the conjecture of that the coding and 5' flanking regions of LMW-GS genes were likely to have evolved in a concerted fashion (Long *et al.*, 2006) but also suggested that the coding regions and every regions of LMW-GS genes were likely to have evolved in a concerted fashion.

The 113 sequences cloned from 24 wild accessions in this study supplied a great deal of information of the huge LMW glutenin family and therefore possess great value for further research in the diversity and evolution of LMW-GS genes. They were expected to be useful gene resources for quality improvement of bread wheat.

## REFERENCES

- Bietz, J.A. and J.S. Wall, 1973. Isolation and characterization of gliadinlike subunits from glutenins. *Cereal Chem.*, 50: 537-547.
- Cassidy, B.G., J. Dvorak and O.D. Anderson, 1998. The wheat low-molecular-weight glutenin genes: Characterization of six genes and progress in understanding gene family structure. *Theor. Applied Genet.*, 96: 743-750.
- D'Ovidio, R. and S. Masci, 2004. The low-molecular-weight glutenin subunits of wheat gluten. *J. Cereal Sci.*, 39: 321-339.
- D'Ovidio, R., C. Marchitelli, L. Ercoli Cardelli and E. Porceddu, 1999. Sequence similarity between allelic Glu-B3 genes related to quality properties of durum wheat. *Theor. Appl. Genet.*, 98: 455-461.
- Gupta, R.B., J.G. Paul, G.B. Cornish, G.A. Palmer, F. Bekes and A.J. Rathjen, 1994. Allelic variation at glutenin subunits and gliadin loci, *Glu21*, *Glu23* and *Gli21* of common wheats I. Its additive and interaction effects on dough properties. *J. Cereal Sci.*, 19: 9-17.
- Harberd, N.P., D. Bartels and R.D. Thompson, 1985. Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. *Mol. Gen. Genet.*, 198: 234-242.
- Kumar, S., K. Tamura, I.B. Jakobsen and N. Masatoshi, 2001. MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics*, 17: 1244-1245.

- Kwok, S., D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson and J.J. Sninsky, 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.*, 18: 999-1005.
- Long, H., Y.M. Wei, Z.H. Yan, B. Baum, E. Nevo and Y.L. Zheng, 2005. Classification of wheat low-molecular-weight glutenin subunit genes and its chromosome assignment by developing LMW-GS group-specific primers. *Theor. Applied Genet.*, 111: 1251-1259.
- Long, H., Y.M. Wei, Z.H. Yan, B. Baum, E. Nevo and Y.L. Zheng, 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.
- Lu, C., W. Yang, W. Zhang and B. Lu, 2005. Identification of SNPs and development of allelic specific PCR marker for high molecular weight glutenin subunit Dtx1.5 from *Aegilops tauschii* through sequence characterization. *J. Cereal. Sci.*, 41: 13-18.
- Masci, S., R. D'Ovidio, D. Lafiandra and D.D. Kasarda, 1998. Characterization of a low-molecular-weight glutenin subunit gene from breadwheat and the corresponding protein that represents a major subunit of the glutenin polymer. *Plant Physiol.*, 118: 1147-1158.
- Nevo, E. and P.I. Payne, 1987. Wheat storage proteins: diversity of HMW glutenin subunits in wild emmer from Israel. *TAG Theor. Applied Genet.*, 74: 827-836.
- Nevo, E., 1983. Genetic resources of wild emmer wheat: Structure, evolution and application in breeding. *Proceedings of the 6th International Wheat Genetics Symposium, (IWGS'83)*, Kyoto University, Kyoto, Japan, pp: 421-431.
- Nevo, E., 1989. Genetic resources of wild emmer wheat revisited: Genetic evolution, conservation and utilization. *Proceedings of the 7th International Wheat Genetics Symposium, (IWGS'89)*, Cambridge, England, pp: 121-126.
- Nevo, E., 1995. Genetic resources of wild emmer wheat, *Triticum dicoccoides* for wheat improvement: News and views. *Proceedings of the 8th International Wheat Genetics Symposium, July 20-25*, Agricultural Sciencetech Press, Beijing, pp: 79-87.
- Nevo, E., 2001. Genetic resources of wild emmer wheat, *Triticum dicoccoides*, for wheat improvement in the third millennium. *Israel J. Plant Sci.*, 49: 77-91.
- Pagnotta, M.A., E. Nevo, A. Belies and E. Proceddu, 1995. Wheat storage proteins DNA diversity in wild emmer wheat, *Triticum dicoccoides*, in Israel and Turkey.2. DNA diversity detected by PCR. *Theor. Appl. Genet.*, 91: 409-414.
- Payne, P.I., 1987. Genetics of wheat storage protein and the effect of allelic variation in bread making quality. *Ann. Rev. Plant Physiol.*, 38: 141-153.
- Shewry, P.R., N.G. Halford and A.S. Tatham, 1992. High molecular weight subunits of wheat glutenin. *J. Cereal Sci.*, 15: 105-120.
- Tanaka, H., S. Toyoda and H. Tsujimoto, 2005. Diversity of low-molecular-weight glutenin subunit genes in Asian common wheat (*Triticum aestivum* L.). *Breed. Sci.*, 55: 349-354.
- Zhao, H.X., A.G. Guo, S.W. Hu, S.H. Fan and D.P. Zhang *et al.*, 2004. Development of primers specific for LMW-GS Genes at Glu-D3 and Glu-B3 loci and PCR Amplification. *Acta Agronomica Sinic*, 30: 126-130.
- Zohary, D., 1970. Centers of Diversity and Centers of Origin. In: *Genetic Resources in Plants: Their Exploration and Conservation*, Frankel, O.H. and E. Bennet (Eds.). Oxford, Blackwell, pp: 33-42.