Theor Appl Genet (2003) 106:1368–1378
DOI 10.1007/s00122-002-1175-y

Z. Liu · Z. Yan · Y. Wan · K. Liu · Y. Zheng · D. Wang

Analysis of HMW glutenin subunits and their coding sequences in two diploid Aegilops species

Received: 19 March 2002 / Accepted: 8 August 2002 / Published online: 7 February 2003

Abstract Considerable progress has been made in understanding the structure, function and genetic regulation of high-molecular-weight (HMW) glutenin subunits in hexaploid wheat. In contrast, less is known about these types of proteins in wheat related species. In this paper, we report the analysis of HMW glutenin subunits and their coding sequences in two diploid *Aegilops* species, *Aegilops umbellulata* (UU) and *Aegilops caudata* (CC). SDS-PAGE analysis demonstrated that, for each of the four *Ae. umbellulata* accessions, there were two HMW glutenin subunits (designated here as 1Ux and 1Uy) with electrophoretic mobilities comparable to those of the x- and y-type subunits encoded by the *Glu-D1* locus, respectively. In our previous study involving multiple accessions of *Ae. caudata*, two HMW glutenin subunits (designated as 1Cx and 1Cy) with electrophoretic mobilities similar to those of the subunits controlled by the *Glu-D1* locus were also detected. These results indicate that the U genome of *Ae. umbellulata* and the C genome of *Ae. caudata* encode HMW glutenin subunits that may be structurally similar to those specified by the D genome. The complete open reading frames (ORFs) coding for x- and y-type HMW glutenin subunits in the two diploid species were cloned and sequenced. Analysis of deduced amino acid sequences revealed that the primary structures of the x- and y-type HMW glutenin subunits of the two *Aegilops* species were similar to those of previously published HMW glutenin subunits. Bacterial expression of modified ORFs, in which the coding sequence for the signal peptide was removed, gave rise to proteins with electrophoretic mobilities identical to those of HMW glutenin subunits extracted from seeds, indicating that upon seed maturation the signal peptide is removed from the HMW glutenin subunit in the two species. Phylogenetic analysis showed that 1Ux and 1Cx subunits were most closely related to the 1Dx type subunit encoded by the *Glu-D1* locus. The 1Uy subunit possessed a higher level of homology to the 1Dy-type subunit compared with the 1Cy subunit. In conclusion, our study suggests that the *Glu-U1* locus of *Ae. umbellulata* and the *Glu-C1* locus of *Ae. caudata* specify the expression of HMW glutenin subunits in a manner similar to the *Glu-D1* locus. Consequently, HMW glutenin subunits from the two diploid species may have potential value in improving the processing properties of hexaploid wheat varieties.

Keywords HMW glutenin subunit · *Aegilops* · *Aegilops umbellulata* · *Aegilops caudata*

Communicated by F. Salamini

Z. Liu · Z. Yan · Y. Wan · K. Liu · D. Wang (✉)
The State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, The Chinese Academy of Sciences, Beijing 100101, China
e-mail: dwwang@genetics.ac.cn
Tel.: 00-86-10-64889380
Fax: 00-86-10-64854467

Z. Yan · Y. Zheng
The Triticeae Research Institute, Sichuan Agricultural University, Dujiangyan 611830, China

Introduction

HMW glutenin subunits are conserved storage proteins that are synthesized in the seeds of wheat and related species (Lawrence and Shepherd 1981; Shewry et al. 1995). In hexaploid wheat, HMW glutenin subunits are encoded by the x- and y-type genes in the *Glu-A1*, *Glu-B1* and *Glu-D1* loci located on the homoeologous group-1 chromosomes (Lawrence and Shepherd 1981; Payne 1987). Due to gene silencing and allelic variation, the composition of HMW glutenin subunits usually differs among hexaploid wheat varieties (Payne et al. 1981; Payne and Lawrence 1983). As seed storage proteins, the biological function of HMW glutenin subunits is to provide carbon, nitrogen and energy sources for seed germination and seedling growth. However, since the discovery that the composition of HMW glutenin subunits affects the baking properties of bread wheat varieties, considerable efforts have been devoted in understanding genetic regulation, structure and function relationships, and the potential of these proteins in improving the

processing properties of wheat varieties (Payne et al. 1981; Shewry et al. 1995). To date, genes coding for more than ten HMW glutenin subunits from wheat have been cloned and characterized (Forde et al. 1985; Sugiyama et al. 1985; Thompson et al. 1985; Halford et al. 1987; Anderson and Greene 1989; Anderson et al. 1989; Halford et al. 1992a; Reddy and Appels 1993; Wan et al. 2002). Amino acid sequence comparisons suggest that different HMW glutenin subunits share a similar primary structure composed of a signal peptide (which is removed from the mature protein upon seed maturation), a N-terminal region, a central repetitive domain and a C-terminal region (Shewry et al. 1995). Within the N- and C-terminal regions, there exist cysteine residues that are highly conserved in both numbers and positions (Shewry et al. 1995). The repetitive domain consisted of short repeated peptide motifs. The size of the repetitive domain usually differs among different subunits, which is caused mainly by changes in the number of repeated peptide motifs (Shewry et al. 1995). A few HMW glutenin subunits (such as 1Ax1 and 1Dx5) have been found to confer good processing properties to hexaploid wheat varieties in conventional breeding programs (Payne 1987). The application of the coding sequences of these subunits in improving grain quality through crop genetic transformation is therefore being actively pursued (Altpeter et al. 1996; Blechl and Anderson 1996; Barro et al. 1997).

In contrast to above progress, our knowledge on orthologous subunits in wheat related species is still limited. In barley, the D-hordeins, which are encoded by the locus *Hor3* on chromosome 5, have been postulated to be structurally related to HMW glutenin subunits from wheat (Halford et al. 1992b). In rye, the *Glu-R1* locus has recently been shown to resemble the *Glu-1* locus in specifying one x- and one y-type HMW glutenin subunit during seed development (De Bustos et al. 2001). In *Elytrigia elongata* and *Dasypyrum villosum*, HMW glutenin subunits are found to be encoded by genes located on chromosome 1E and 1V, respectively (Lawrence and Shepherd 1981; Dvorák et al. 1986; Montebove et al. 1987; De Pace et al. 1988; Blanco et al. 1991; Zhong and Qualset 1993). In the *Aegilops* genus, SDS-PAGE studies reveal that many species encode HMW glutenin subunits and that the expression of these subunits may be controlled by chromosomal loci similar to *Glu-1* (Lawrence and Shepherd 1981; Fernández-Calvín and Orellana 1990; Peña et al. 1991; Urbano et al. 1993; William et al. 1993; Mackie et al. 1996a; Wan et al. 2000; Xie et al. 2001; Pflüger et al. 2001). The coding sequences for several *Aegilops* HMW glutenin subunits have also been isolated and characterized (Mackie et al. 1996b; Xie et al. 2001; Wan et al. 2002). The results show that the primary structure of *Aegilops* HMW glutenin subunits is similar to the one shared by wheat subunits, but can possess novel modifications that are not found in the wheat subunits. For example, compared to the y-type subunits from wheat, the 1Dy subunit from *Aegilops cylindrica* possesses an extra cysteine residue towards the end of its repetitive domain (Wan et al. 2002). Because it has been shown that the 1Dx5 subunit with an additional cysteine residue in its structure exerts a positive influence on dough properties (Lafiandra et al. 1993; Gupta and MacRitchie 1994), structure variants of HMW subunits from *Aegilops* species (like the 1Dy subunit from *Ae. cylindrica*) may represent novel forms of functional proteins that have potential value in improving the processing properties of wheat. Therefore, studies of HMW glutenin subunits in *Aegilops* species are important for not only achieving a better understanding of the structure, function and evolution of these proteins across the species in *Triticeae* but also identifying additional genetic resources that may be useful in wheat grain quality improvement.

Aegilops umbellulata (2n = 2*x* = 14, UU) and *Aegilops caudata* (2n = 2*x* = 14, CC) are two diploid species whose genomes are involved in the formation of many polyploid *Aegilops* species (Miller 1987; Friebe et al. 1992; Friebe et al. 1995; Zhang et al. 1998). Past studies have shown that the 1U chromosome of *Ae. umbellulata* containing genes encoding HMW glutenin subunits with a composition similar to that controlled by the *Glu-D1* locus (Brown et al. 1979; Lawrence and Shepherd 1981; Payne et al. 1983). In our investigations, the C genome of *Ae. caudata* has also been found to code for a x- and a y-type subunit displaying electrophoretic mobilities comparable to those of the 1Dx and 1Dy subunits of hexaploid wheat (Wan et al. 2000). In the work reported in this paper, we cloned and sequenced the open reading frames (ORFs) encoding complete sets of HMW glutenin subunits from both species. These results have, for the first time, enabled us to compare the primary structure of HMW glutenin subunits from the two species with that of previously published subunits at the molecular level.

Materials and methods

Plant materials

Four accessions of *Ae. umbellulata* (IG46953, Y39, Y137, and Y139) and one accession of *Ae. caudata* (Y588) were used in this study. IG46953 was obtained from ICARDA (The International Center for Agricultural Research in the Dry Areas). The remaining four accessions were obtained from the Institute of Crop Germplasm Resources (ICGR) of the Chinese Academy of Agricultural Sciences. HMW glutenin subunits from hexaploid wheat varieties, Chinese Spring (1Bx7 + 1By8, 1Dx2 + 1Dy12) and MG7249 (1Ax2*, 1Bx7 + 1By8, 1Dx2.2 + 1Dy12), were used as standards for assessing the electrophoretic mobilities of the HMW glutenin subunits of *Ae. umbellulata* and *Ae. caudata* in SDS-PAGE analysis.

SDS-PAGE and Western-blot analysis

Seed proteins were prepared for SDS-PAGE and Western-blot analysis as described previously (Wan et al. 2000). For studying the composition of HMW glutenin subunits in each accession, at least five individual seeds were examined. In Western-blot experiments, the primary antibody was either a polyclonal antibody recognizing all types of wheat HMW glutenin subunits (Wan et al. 2000) or the

Table 1 Sets of PCR primers used in bacterial expression experiments

Set	Sequence[a]	Restriction sites introduced	Mutant ORF amplified
1	5'-ACC*CATATG*GAAGGTGAGGCCTCTG-3' 5'-CTA*GAATTC*CTATCACTGGCTGGCC-3'	*Nde*I *Eco*RI	1Ux
2	5'-ACC*CATATG*AAAGGTGAGGCCTCTA-3' 5'-CTA*GAATTC*CTATCACTGGCTGGCC-3'	*Nde*I *Eco*RI	1Uy
3	5'-ACC*CATATG*GAAGGTGAGGCCTCTG-3' 5'-TTC*GAATTC*CTATCACTGGCTGGCC-3'	*Nde*I *Eco*RI	1Cx
4	5'-ACC*CATATG*GAAGGTGAGGCCTCTA-3' 5'-TTC*GAATTC*CTATCACTGGCTAGCC-3'	*Nde*I *Eco*RI	1Cy

[a]The italicized nucleotides constitute the restriction sites listed in the third column of the table

monoclonal antibody IFRN 1602 that is specific for the 1Ax- and 1Dx-type HMW glutenin subunits (Fido et al. 1995). The results of SDS-PAGE and Western-blot experiments were recorded using a digital camera (Coolpix 990, Nikon).

Cloning and sequencing of HMW glutenin gene ORFs

The seedlings of the accessions IG46953 and Y588 were grown in darkness for 5 to 7 days at 24 °C. Genomic DNA was prepared from the etiolated seedlings using a previously published procedure (Guidet et al. 1991). For amplifying the complete ORFs of HMW glutenin genes of *Ae. umbellulata* and *Ae. caudata* in genomic PCR reactions, a pair of degenerate primers was designed according to published DNA sequences of HMW glutenin subunits. The sequences of the two primers are P1 (5'-ATGGCTAAGCGGC/TTA/GGTCCTCTTTG-3'), and P2 (5'-CTATCACTGGCTG/AGC-CGAC-AATGCG-3').

Primer P1 contains the start codon of the HMW ORF whereas primer P2 possesses the two tandem stop codons that are conserved in the HMW glutenin ORFs characterized so far. Genomic PCR reactions were carried out using the Ex *Taq* polymerase (TaKaRa Biotechnology Co., Ltd). The parameters for the reaction were one cycle at 94 °C for 5 min, followed by 35 cycles of 94 °C for 1 min and 68 °C for 3 min, and a final extension step at 68 °C for 7 min. PCR products were separated in 0.8% agarose gels. DNA fragments of expected sizes were recovered from agarose gels, and were subsequently cloned into the pGEM-T Easy vector (Promega). By restriction enzyme digestion mapping and partial DNA sequencing, the inserts in four positive clones, p1Ux-3.0, p1Uy-1.9, p1Cx-2.5 or p1Cy-1.9, were deduced to represent the ORF for the x- and y-type HMW glutenin subunits of *Ae. umbellulata* and *Ae. caudata*, respectively. To determine the complete DNA sequences of selected inserts, series of subclones were prepared for each insert using the nested deletion method (Sambrook et al. 1989). The inserts in the resultant subclones were sequenced by a commercial company (TaKaRa). For each insert, the full-length nucleotide sequence was assembled from sequence data generated from six to eight subclones.

Bacterial expression of cloned ORFs

For bacterial expression of the mature proteins of HMW glutenin subunits from *Ae. umbellulata* and *Ae. caudata*, four sets of PCR primers (Table 1) were designed for (1) amplifying mutant ORFs from which the sequences coding for signal peptides were removed, and (2) introducing appropriate restriction enzyme sites for the mutant ORFs to facilitate subsequent cloning experiments (Table 1). PCR conditions for amplifying mutant ORFs were identical to those described above except that the template was plasmid DNA purified from the clone p1Ux-3.0, p1Uy-1.9, p1Cx-2.5 or p1Cy-1.9. After cloning the mutant ORFs into the expression vector pET-30a (Invitrogen), four constructs, pET-1Ux, pET-1Uy, pET-1Cx and pET-1Cy, were selected for expressing the mature proteins of the 1Ux, 1Uy, 1Cx and 1Cy subunits, respectively, in bacterial cells. The conditions for inducing bacterial expression of pET constructs were the same as those detailed by Wan et al. (2002). For SDS-PAGE analysis, protein extracts were prepared by either directly dissolving cells in SDS-PAGE sample buffer (Wan et al. 2000), or treating cells with a method that was designed to preferentially extract HMW glutenin subunits from complex protein mixtures (Mackie et al. 1996b). To confirm the identity of the bacterially expressed proteins directed by pET-1Ux, pET-1Uy, pET-1Cx and pET-1Cy constructs, Western-blot experiments using the polyclonal antibody specific for HMW glutenin subunits were conducted as described in the above section.

Amino-acid sequence comparisons and phylogenetic analysis

The HMW glutenin ORFs were translated into amino acid sequences using the ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The deduced amino acid sequences were compared to those of the representatives of published HMW glutenin subunits with the aid of the Clustal W program (Thompson et al. 1994). For studying evolutionary distances and constructing phylogenetic trees, the MEGA program (Version 2, http://www.oup-usa.org/sc/0195135857, Nei and Kumar 2000) was followed. During the analysis, the complete deletion option was adopted with respect to gaps in the aligned sequences, and the evolutionary distances were measured by calculating PC distances for each pair of aligned sequences. The bootstrap values in the phylogenetic tree were estimated based on 500 replications.

Results

HMW glutenin subunits of *Ae. umbellulata*

In SDS-PAGE analysis of seed protein extracts, two protein bands (Fig. 1A, marked with short and long arrows, respectively) with electrophoretic mobilities comparable to those of hexaploid wheat HMW glutenin subunits (Fig. 1A, lanes 1 and 2) were found for each of the four *Ae. umbellulata* accessions. In Western-blot experiments, the bands indicated by the arrows (Fig. 1A) reacted strongly with a polyclonal antibody specific for HMW glutenin subunits (Fig. 1B), indicating that they were likely to represent HMW glutenin subunits encoded by the U genome of *Ae. umbellulata*. In further Western-blot experiments using the monoclonal antibody IFRN1602, the subunits indicated by the short arrows (Fig. 1A, C), but not the ones by the long arrows (Fig. 1A), showed positive signals (Fig. 1C), suggesting that the reactive proteins were the x-type subunits (designated

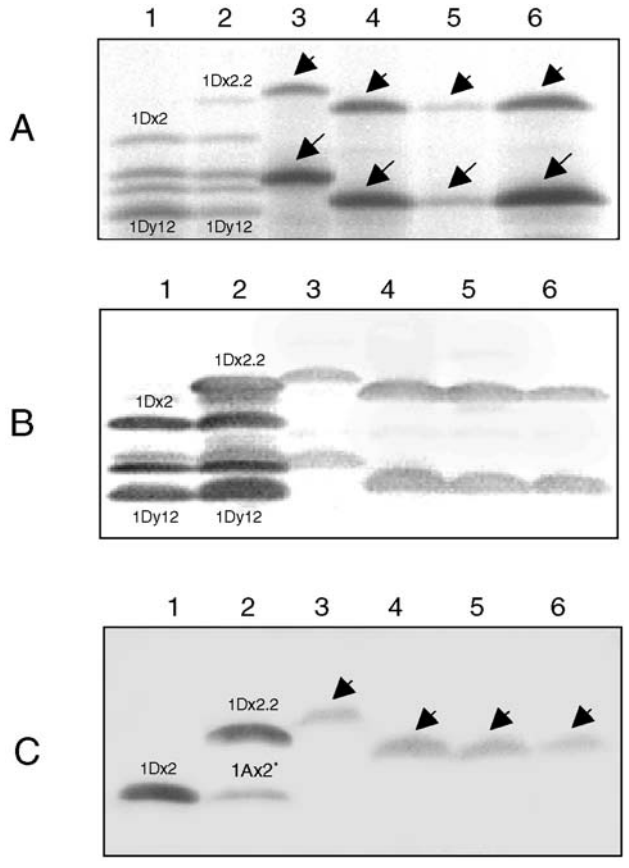

Fig. 1A–C SDS-PAGE (**A**) and Western-blot (**B, C**) analysis of HMW glutenin subunits of *Ae. umbellulata*. *Lanes 1 and 2* are HMW glutenin subunits from hexaploid wheat varieties Chinese Spring and MG7249, respectively. *Lanes 3 to 6* are HMW glutenin subunits of four *Ae. umbellulata* accessions IG46953, Y39, Y137 and Y139, respectively. 1Dx2 and 1Dy12 are subunits encoded by the *Glu-D1* locus of Chinese Spring, 1Dx2.2 and 1Dy12 are encoded by an orthologous locus in MG7249. In **A**, two putative HMW subunits of different electrophoretic mobilities (indicated by *short and long arrows*, respectively) are detected for each of the four *Ae. umbellulata* accessions. In **B**, the two types of HMW subunits from *Ae. umbellulata* accessions (lanes 3 to 6) react strongly to a polyclonal antibody specific for known HMW glutenin subunits (lanes 1 and 2). In **C**, HMW glutenin subunits from *Ae. umbellulata* accessions show differential reactions to the monoclonal antibody IFRN 1602. The subunits with a slower electrophoretic mobility (indicated by *short arrows*), but not the ones with a faster electrophoretic mobility (indicated by *long arrows* in **A**), exhibit positive signals. In the same experiment, the 1Dx and 1Ax subunits of hexaploid wheat (lanes 1 and 2) are specifically recognized by IFRN 1602

hereafter as 1Ux) whereas the non-reactive proteins were the y-type subunits (designated hereafter as 1Uy). The electrophoretic mobilities of the 1Ux subunits were generally similar to that displayed by the 1Dx2.2 subunit, whereas the electrophoretic mobilities of the 1Uy subunits from most accessions were similar to that of the 1Dy12 subunit (Fig. 1A), except for the y-type subunit of the accession IG46953, whose electrophoretic mobility was close to that of 1By8 (Fig. 1A, lane 3).

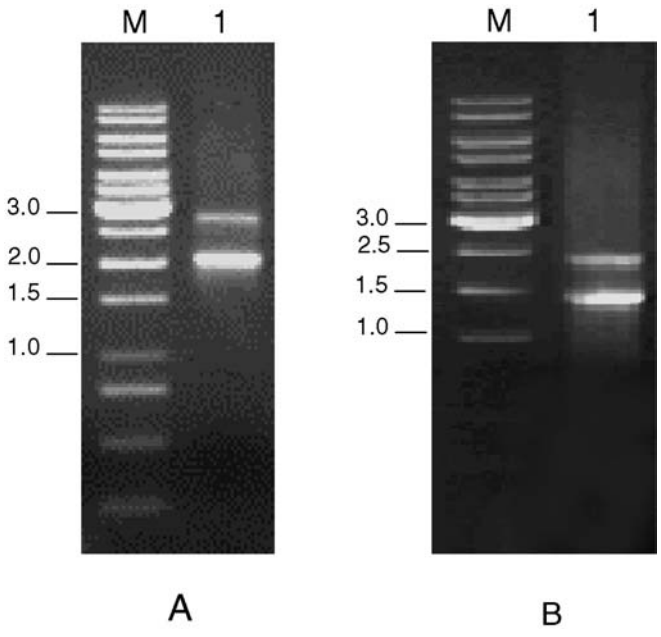

Fig. 2A, B Amplification of complete ORFs coding for HMW glutenin subunits of *Ae. umbellulata* accession IG46953 (**A**, lane 1) and *Ae. caudata* accession Y588 (**B**, lane 1) in genomic PCR experiments. The size of DNA markers (kb, lane M) is indicated on the left side of the figure

Analysis of HMW glutenin gene ORFs and derived amino acid sequences

In genomic PCR experiments using the primers P1 and P2, two DNA fragments were amplified for both *Ae. umbellulata* (accession IG46953, Fig. 2A) and *Ae. caudata* (accession Y588, Fig. 2B). After cloning the four fragments into the plasmid vector pGEM-T Easy, four types of clones, with insert sizes ranging from approximately 1.9 to 3.0 kb, were obtained. The inserts in four plasmid clones, p1Ux-3.0, p1Uy-1.9, p1Cx-2.5 and p1Cy-1.9, were selected to represent the complete ORF of the 1Ux, 1Uy, 1Cx and 1Cy subunits, respectively. Using subclones prepared by nested deletion, the 1Ux, 1Uy, 1Cx and 1Cy ORFs were completely sequenced. Analysis of amino acid sequences derived from the cloned ORFs showed that HMW glutenin subunits from *Ae. umbellulata* and *Ae. caudata* possessed a primary structure that was identical to the one shared by published HMW glutenin subunits (Fig. 3A, B). In each case, a conserved structure consisting of a signal peptide, a N-terminal region, a central repetitive domain and a C-terminal region was observed (Fig. 3A, B). Some properties of the 1Ux, 1Uy, 1Cx and 1Cy subunits, in comparison to those of the representatives of published HMW glutenin subunits, were summarized in Table 2. Except for the extra cysteine residues in the 1Dx5 and 1Ry subunits (underlined, single cysteine residues, Fig. 3A, B) and the substitution of the cysteine residue with a phenylalanine residue in the repetitive domain of the 1Ay subunit (Fig. 3B), the numbers and positions of other cysteine

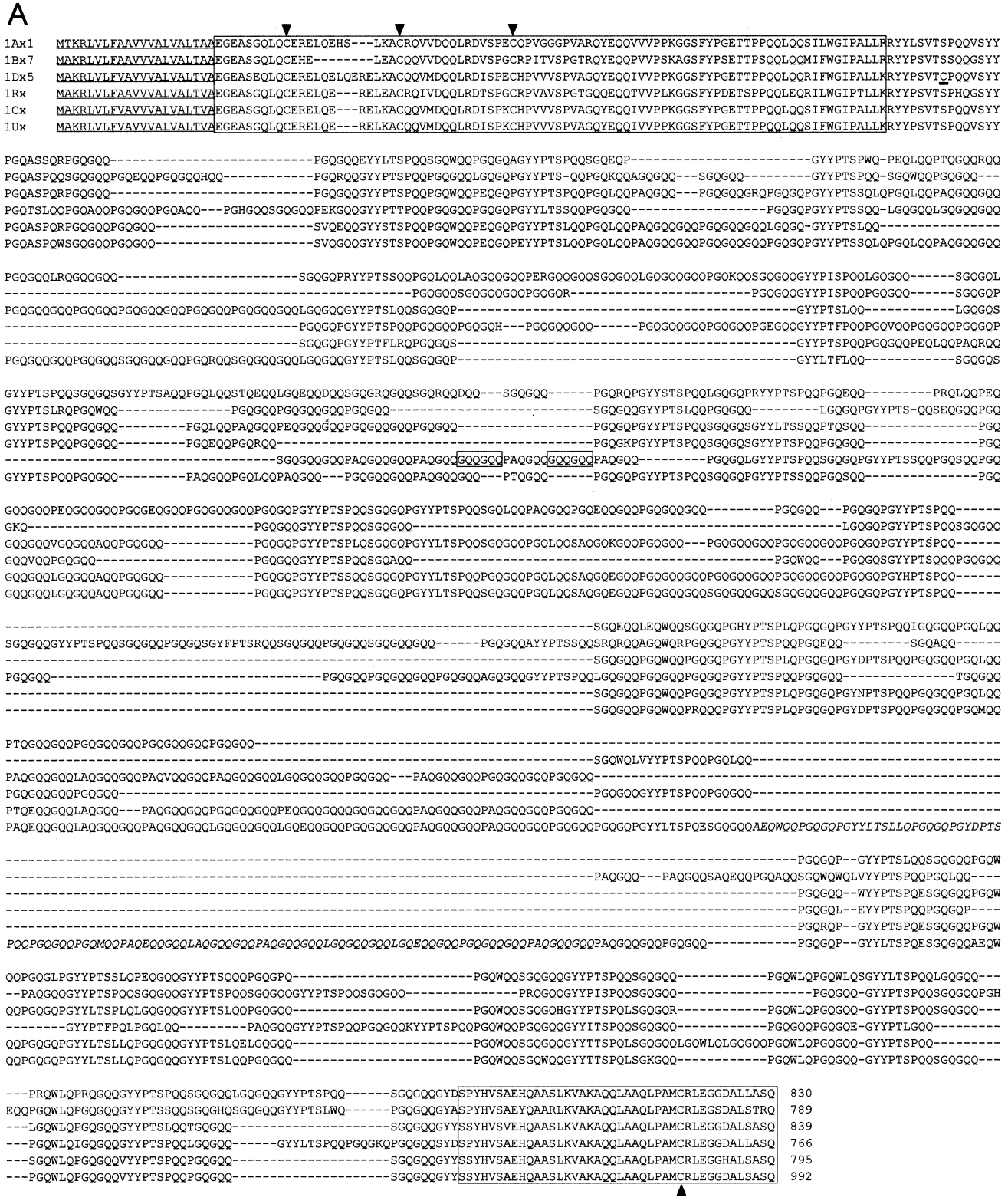

Fig. 3A, B Comparison of the primary structure of 1Cx and 1Ux to that of published x-type HMW glutenin subunits (**A**) and the primary structure of 1Cy and 1Uy to that of published y-type HMW glutenin subunits (**B**). For reason of space, the amino acid sequences of 1Ax1, 1Ay (*T. timopheevi*), 1Bx7, 1By9, 1Dx5, 1Dy10, 1Rx and 1Ry are used as representative of published HMW glutenin subunits. The region for the signal peptide is *underlined*. The N- and C-terminal regions are *boxed*. The repetitive domain is situated between the *boxed* N- and C-terminal regions. *Arrowheads* indicate cysteine residues conserved in x- or y-type subunits. The *underlined*, *single letters* represent the extra cysteine residues in the 1Dx5 and 1Ry subunits, respectively. Tandem duplications of the tripeptide GQQ in the 1Cx subunit are also *boxed*. The extra portion of the amino acid sequence in the repetitive domain of the 1Ux subunit is *italicized*. EMBL accession numbers for 1Ax1, 1Ay (*T. timopheevi*), 1Bx7, 1By9, 1Dx5, 1Dy10, 1Rx and 1Ry are X61009, AJ306977, X13927, X61026, X12928, X12929, AF216868 and AF216869, respectively

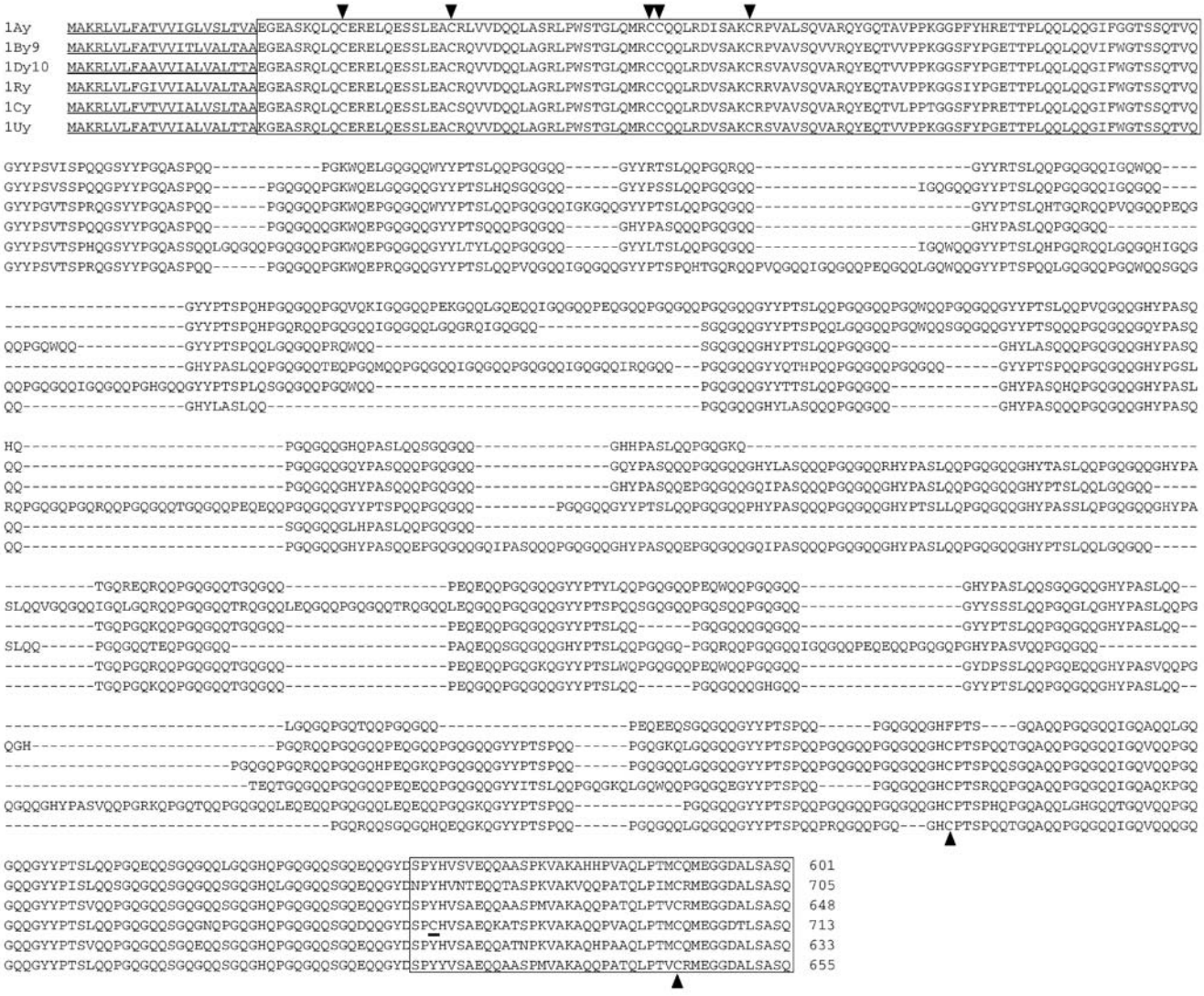

Fig. 3B

residues in the x- and y- type subunits compared showed strong type-specific conservation (Fig. 3A, B, Table 2).

From Fig. 3A, B and Table 2, it was also clear that HMW glutenin subunits from *Ae. umbellulata* and *Ae. caudata* possessed novel modifications in their amino acid sequences. The 1Ux, 1Uy, 1Cx and 1Cy subunits differed from each other and from the representatives of published HMW glutenin subunits by substitutions, insertions and/or deletions involving single or more amino acid residues (Fig. 3A, B). For example, the size of the repetitive domain of the 1Ux subunit was substantially larger than that found for previously reported subunits. This was reflected by a rise in the number of repeated peptide motifs forming the central domain (Fig. 3A, the extra portion of the sequence in the repetitive domain of 1Ux is italicized) and an increase in the size of the central domain of the subunit (Table 2). In the repetitive domain of the 1Cx subunit there were tandem duplications of the tripeptide GQQ (boxed sequences, Fig. 3A). These duplications were located in a region composed of regular arrangements of hexapeptides (SGQGQQ, PAQGQQ) and tripeptides (GQQ).

Expression of the HMW glutenin gene ORFs of *Ae. umbellulata* and *Ae. caudata* in bacterial cells

For bacterial expression, the nucleotide sequence coding for the signal peptide was removed from the cloned ORF by PCR mutagenesis. After cloning the modified ORFs into the pET-30a vector, four expression constructs, pET-1Ux, pET-1Uy, pET-1Cx and pET-1Cy, were chosen for

Table 2 A summary of some of the properties of the primary structure of the 1Ux, 1Uy, 1Cx and 1Cy subunits in comparison with those of the representatives of published HMW glutenin subunits

Subunits	N-terminus		C-terminus		Repetitive domain		Total	
	Size[a]	Cysteine residues	Size[a]	Cysteine residues	Size[a]	Cysteine residues	Size[a]	Cysteine residues
1Ux	86	3	42	1	843	0	971	4
1Cx	86	3	42	1	646	0	774	4
1Ax1	86	3	42	1	681	0	809	4
1Bx7	80	3	42	1	648	0	770	4
1Dx5	89	3	42	1	696	1	827	5
1Rx	86	3	42	1	617	0	745	4
1Uy	104	5	42	1	488	1	634	7
1Cy	104	5	42	1	464	1	610	7
1Ay	104	5	42	1	420	0	566	6
1By9	104	5	42	1	538	1	684	7
1Dy10	104	5	42	1	481	1	627	7
1Ry	104	5	42	2	546	1	692	8

[a]Represented by the number of amino acid residues

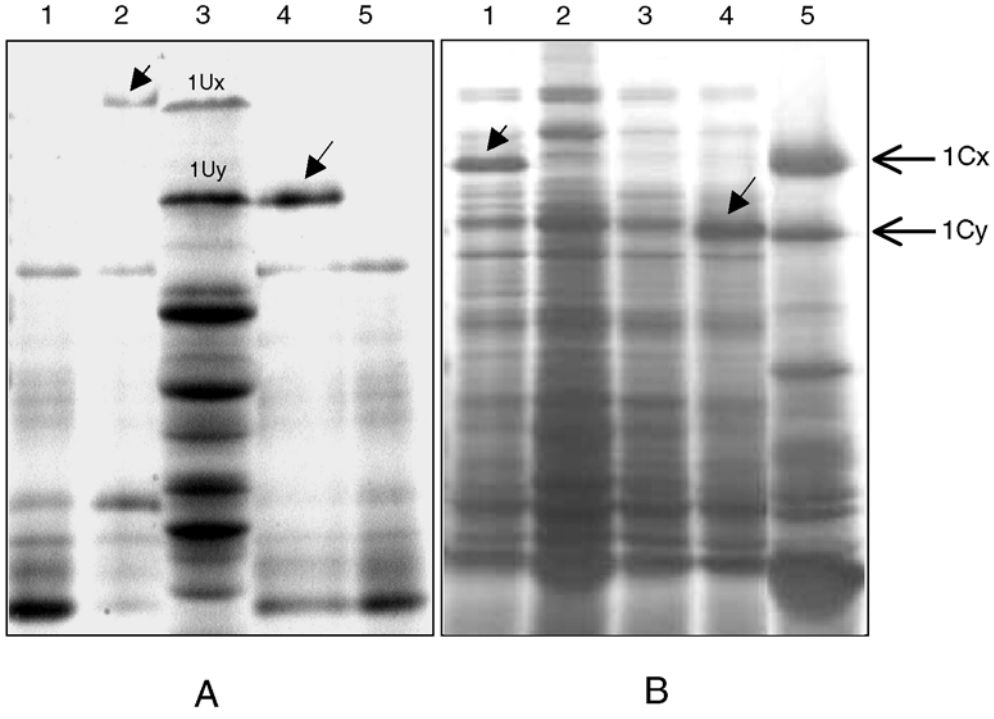

Fig. 4A, B Expression of the modified ORFs of 1Ux, 1Uy, 1Cx and 1Cy genes in bacterial cells and SDS-PAGE analysis of expressed products. The modified ORFs are prepared by removing the coding sequence for the signal peptide from each of the four complete ORFs via PCR mutagenesis. In **A**, protein extracts are prepared by the method that results in preferential extraction of HMW glutenin subunits from complex protein mixtures (Mackie et al. 1996b). IPTG induction of the bacterial culture harboring pET-1Ux (lane 2) or pET-1Uy (lane 4) leads to the synthesis of 1Ux (indicated by the *short arrow* in lane 2) or 1Uy (marked by the *long arrow* in lane 4) subunit with an electrophoretic mobility identical to that of the x- or y-type subunit originated from seed (lane 3). In the absence of IPTG induction, no HMW glutenin subunit is produced in the culture harboring either pET-1Ux (lane 1) or pET-1Uy (lane 5). In **B**, protein extracts are prepared by dissolving cells directly in SDS-PAGE sample buffer. IPTG induction of the bacterial culture harboring pET-1Cx (lane 1) or pET-1Cy (lane 4) gives rise to the production of 1Cx (indicated by the *short arrow* in lane 1) or 1Cy (marked by the *long arrow* lane 4) subunit with an electrophoretic mobility identical to that of the x- or y-type subunit derived from seed (lane 5). The lack of IPTG induction abrogates the synthesis of HMW glutenin subunit in the culture harboring either pET-1Cx (lane 2) or pET-1Cy (lane 3)

expressing the mature proteins of the 1Ux, 1Uy, 1Cx and 1Cy subunits, respectively, in bacterial cells. In SDS-PAGE analysis of protein extracts prepared from induced bacterial cultures, the electrophoretic mobility of the two proteins directed by pET-1Ux and pET-1Uy (Fig. 4A, lanes 2 and 4) was similar to that of the 1Ux and 1Uy subunits extracted from seeds (Fig. 4A, lane 3), respectively. In a similar analysis, the electrophoretic mobility

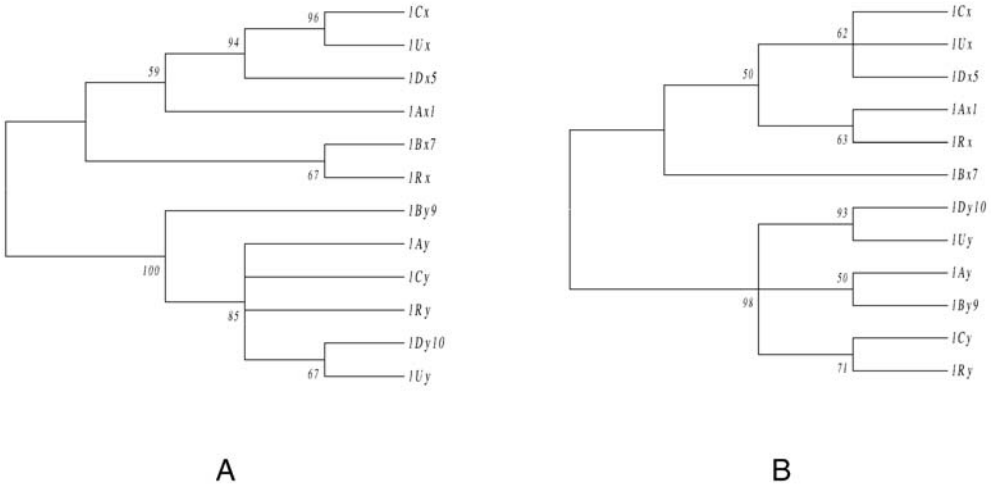

Fig. 5A, B Analysis of evolutionary relationships among 1Ux, 1Uy, 1Cx, 1Cy and published HMW glutenin subunits. The phylogenetic trees in **A** and **B** are constructed based on N- and C-terminal amino acid sequences, respectively. The bootstrap values are obtained using 500 replications

of the two proteins directed by pET-1Cx and pET-1Cy (Fig. 4B, lanes 1 and 4) was found to correspond to that of the 1Cx and 1Cy subunits extracted from seeds (Fig. 4B, lane 5), respectively. The identity of the four bacterially expressed subunits was confirmed by Western-blot experiments. In these assays, all four bacterially expressed subunits exhibited strong reactions to the polyclonal antibody that was specific for HMW glutenin subunits (data not shown).

Evolutionary relationship studies

To investigate evolutionary relationships among the subunits characterized in this study and the representatives of published HMW glutenin subunits, evolutionary distances were calculated for each pair of subunits based on either N- or C-terminal amino acid sequences. In both types of calculations, the distance values for the 1Ux and 1Dx5 pair, the 1Cx and 1Dx5 pair, and the 1Uy and 1Dy10 pair were among the smallest (data not shown). In the phylogenetic tree constructed based on N-terminal amino acid sequences, the 1Ux, 1Cx and 1Dx5 subunits were clustered together with high bootstrap values (Fig. 5A). In contrast, the 1Uy, 1Cy and 1Dy10 subunits did not form a clade of their own (Fig. 5A). Instead, the 1Uy and 1Dy10 subunits constituted an interior branch that was parallel with that of the 1Ay, 1Cy and 1Ry subunits (Fig. 5A). In the phylogenetic tree derived from the C-terminal amino acid sequences, a similar pattern of clustering was found (Fig. 5B), except that the bootstrap value for the 1Uy and 1Dy10 branch was much higher than the corresponding value in the tree based on N-terminal amino acid sequences (Fig. 5A).

Discussion

In the work reported in this paper we studied HMW glutenin subunits and their coding sequences in two diploid *Aegilops* species, *Ae. umbellulata* and *Ae. caudata*. SDS-PAGE and Western-blot analysis of seed proteins prepared from multiple accessions demonstrated that the U genome of *Ae. umbellulata* was likely to encode one x- and one y-type HMW glutenin subunit. Further experiments on cloning, sequencing and in vitro expression of two different types of HMW glutenin gene ORFs from *Ae. umbellulata* confirmed that there existed two types of HMW glutenin genes in the U genome of this species. Our results, together with previous finding on the location of HMW glutenin genes on the 1U chromosome of *Ae. umbellulata* (Brown et al. 1979; Lawrence and Shepherd 1981), suggest that there is a genetic locus on the 1U chromosome (designated here as *Glu-U1*) that contains x- and y-type HMW glutenin genes. Both types of genes are actively expressed, leading to the synthesis of the 1Ux and 1Uy subunits in the seeds of the *Ae. umbellulata* accessions analyzed in this study.

The C genome of *Ae. caudata* has been found to encode one x- and one y- type HMW glutenin subunit in a previous investigation (Wan et al. 2000). In the current study, the results on cloning, sequencing and in vitro expression of two different types of HMW glutenin gene ORFs from *Ae. caudata* suggest that the C genome of this species contain two types of HMW glutenin genes, which specify the synthesis of the 1Cx and 1Cy subunits in the seeds, respectively. The data from our experiments also imply that there may be a chromosomal locus in the C genome (designated as *Glu-C1*) that functions in a manner similar to the *Glu-1* locus in other related species. But the chromosomal location of the putative *Glu-C1* locus is unclear at present. Chromosome addition lines of *Ae. caudata* having the background of the wheat variety Chinese Spring have been prepared (Friebe et al. 1992). These lines will be useful in further investigations on chromosomal location of the *Glu-C1* locus of the C genome.

Amino acid sequence comparisons showed that the primary structure of 1Ux, 1Uy, 1Cx and 1Cy was similar to that of previously published HMW glutenin subunits. In each of the four subunits, the typical HMW glutenin subunit structure composed of a signal peptide, a N-

terminal region, a central repetitive domain and a C-terminal region was found. In past studies on HMW glutenin subunits from other *Aegilops* species (such as *Aegilops tauschii* and *Aegilops cylindrica*) and rye, conservation in the primary structure is also noted. Analysis of the amino acid sequence derived from a barley D hordein gene indicates that the D hordein protein may also adopt a primary structure related to that of HMW glutenin subunits (Halford et al. 1992b). The preservation of the HMW glutenin subunit-like structure in subsets of seed storage proteins from different *Triticeae* species discussed above may reflect the fact that these proteins reside in closely related species and play similar and important roles in seed development and during seed germination.

Parallel to the conservation of the HMW glutenin subunit structure, the 1Ux, 1Uy, 1Cx and 1Cy subunits also differed from each other and from published HMW glutenin subunits by variations in their amino acid sequences. In the 1Ux subunit, the size of its repetitive domain was substantially larger than that of published HMW glutenin subunits for which full-length amino acid sequences are available for comparison. In hexaploid wheat, the size of two 1Dx type subunits, 1Dx2.2 and 1Dx2.2*, has been found to be much larger than that of other wheat subunits (Payne and Lawrence 1983; Payne et al. 1983). In the case of 1Dx2.2*, molecular analysis has shown that the larger size of the subunit is associated with an insertion in the repetitive domain (D'Ovidio et al. 1996). However, owing to the lack of the full-length amino acid sequence for the 1Dx2.2* subunit, a direct comparison of the repetitive domain of this subunit with that of the 1Ux subunit is presently not possible. The presence of tandem duplications of the tripeptide GQQ in the repetitive domain of the 1Cx subunit represented another form of sequence modification. In the x-type HMW glutenin subunits that have been molecularly characterized so far, the GQQ tripeptides are usually separated from each other by the presence of one or more hexapeptides in between. Because the tandem GQQ duplications in the 1Ux subunit were situated in a region composed of regular arrangements of hexapeptides and tripeptides, we deduce that the tandem duplication may have been caused by deletion of the hexapeptide that was originally present in between the tripeptides. Apart from the two modifications described, there were also variations involving substitution, insertions and/or deletions of single or more amino acid residues among the amino acid sequences of the 1Ux, 1Uy, 1Cx and 1Cy subunits. All of these variations may have contributed to the uniqueness of the four different proteins, but have not changed the nature of the four proteins as HMW glutenin subunits.

Based on evolutionary distances that were calculated from aligned N- or C-terminal amino acid sequences, the 1Ux and 1Uy subunits were found most closely related to the 1Dx5 and 1Dy10 subunits, respectively. In contrast to the *Glu-U1* locus, the relationship between the subunits encoded by the *Glu-C1* and *Glu-D1* loci was more complex. While the 1Cx subunit was closely related to the 1Dx5 subunit, the relationship between the 1Cy and 1Dy10 subunits was more distant. In a similar phylogenetic analysis, the relationship between the Cy subunit of *Ae. cylindrica* and 1Dy10 was also found to be relatively distant (data not shown). The reason underlying this pattern of relationship between HMW subunits encoded by the *Glu-C1* and *Glu-D1* loci is unclear at present and will be a subject for further investigations.

In addition to yield a better understanding of HMW glutenin subunits of *Ae. umbellulata and Ae. caudata*, the present study also highlights that it may be useful to examine the potential values of the subunits of the two species in improving the processing properties of hexaploid wheat varieties. By constructing mutant HMW glutenin proteins that differ in the size of the repetitive domain, Anderson and colleagues found that an increase of a suitable degree in the size of the repetitive domain might enhance the functionality of HMW glutenin proteins in controlling dough properties (Anderson et al. 1996; US Patent 6174725). In this respect, it will be interesting to test if the 1Ux subunit can confer good processing properties when expressed in hexaploid wheat varieties. One way to do this would be to compare the processing properties of Chinese Spring with those of the 1U chromosome addition line having the Chinese Spring background, but the results may be complicated by the presence and function of other protein loci carried on the 1U chromosome. A more ideal strategy would be to transgenically express only the 1Ux gene in hexaploid wheat followed by testing the processing properties of the resultant plants. To this end, the preparation of transgenic wheat plants designed to allow endosperm specific expression of the 1Ux gene is currently underway.

Acknowledgements The first two authors contributed equally to this work. We thank Professor Peter Shewry for providing the monoclonal antibody IFRN 1602, and Dr. Jianru Zuo for comments on the manuscript. This work is supported by grants (2001AA222091, G1998010208) from the Ministry of Science and Technology of China and a biotechnology grant from the Chinese Academy of Sciences. The GenBank accession numbers for the ORFs of 1Cx, 1Cy, 1Ux and 1Uy are AF476959, AF476960, AF476961 and AF476962, respectively.

References

Anderson OD, Greene FC (1989) The characterization and comparative analysis of HMW glutenin genes from genomes A and B of hexaploid wheat. Theor Appl Genet 77:689–700

Anderson OD, Kuhl JC, Tam A (1996) Construction and expression of a synthetic wheat storage protein gene. Gene 174:51–58

Anderson OD, Greene FC, Yip RE, Halford NG, Shewry PR, Malpica-Romero J-M (1989) Nucleotide sequences of the two high-molecular-weight glutenin genes from the D-genome of a hexaploid bread wheat, *Triticum aestivum* L. cv Cheyenne. Nucleic Acids Res 17:461–462

Altpeter E, Vasil V, Srivastava V, Vasil IK (1996) Integration and expression of the high-molecular-weight glutenin subunit 1Ax1 gene into wheat. Nature Biotechnol 14:1155–1159

Barro F, Rooke l, Békés, Gras P, Tatham AS, Fido R, Lazzeri PA, Shewry PR, Barceó P (1997) Transformation of wheat with

high electrophoretic mobility subunit genes results in improved functional properties. Nature Biotechnol 15:1295–1299

Blanco A, Resta P, Simeone R, Parmar S, Shewry PR, Sabelli PW, Lafiandra D (1991) Chromosomal location of seed storage protein genes in the genome of *Daspyrum villosum* (L.) Candargy. Theor Appl Genet 82:358–362

Blechl AE, Anderson OD (1996) Expression of a novel high-molecular-weight glutenin subunit gene in transgenic wheat. Nature Biotechnol 14:875–879

Brown JWS, Kemble RJ, Law CN, Flavell RB (1979) Control of endosperm proteins in *Triticum aestivum* (var. Chinese Spring) and *Aegilops umbellulata* by homoeologous group-one chromosomes. Genetics 93:189–200

De Bustos A, Rubio P, Jouve N (2001) Characterisation of two gene subunits on the 1R chromosome of rye as orthologs of each of the *Glu-1* genes of hexaploid wheat. Theor Appl Genet 103:733–742

De Pace C, Montebove L, Delre V, Jan CC, Qualset CO, Scarascia-Mugnozza GT (1988) Biochemical versatility of amphiploids derived from crossing *Dasypyrum villosum* (L.) Candargy and wheat: genetic control and phenotypical aspects. Theor Appl Genet 76:513–529

D'Ovidio R, Lafiandra D, Porceddu E (1996) Identification and molecular characterization of a large insertion within the repetitive domain of a high-molecular-weight glutenin subunit gene from hexaploid wheat. Theor Appl Genet 93:1048–1053

Dvorák J, Kasarda DD, Dietler MD, Lew EJL, Anderson OD, Litts JC, Shewry PR (1986) Chromosomal location of seed storage protein genes in the genome of *Elytrigia elongata*. Can J Genet Cytol 28:818–830

Fernández-Calvín B, Orellana J (1990) High electrophoretic mobility glutenin subunit variation in the *Sitopsis* section of *Aegilops*. Implications for the origin of the B genome of wheat. Heredity 65:455–463

Fido RJ, Tatham AS, Shewry PR (1995) Western-blotting analysis. In: Jones H (ed) Methods in molecular biology, vol. 49: plant gene transfer and expression protocols. Humana press, Totowa, New Jersey, pp 423–437

Forde J, Malpica J-M, Halford NG, Shewry PR, Anderson OD, Greene FC, Miflin BJ (1985) The nucleotide sequence of a HMW subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.). Nucleic Acids Res 13:6817–6832

Friebe B, Schubert V, Blüthner WD, Hammer K (1992) C-banding pattern and polymorphism of *Aegilops caudata* and chromosomal constitutions of the amphiploid *T. aestivum–Ae. caudata* and six derived chromosome addition lines. Theor Appl Genet 83:589–596

Friebe B, Jiang J, Tuleen N, Gill BS (1995) Standard karyotype of *Triticum umbellulata* and the characterisation of derived chromosome addition and translocation lines in common wheat. Theor Appl Genet 90:150–156

Guidet F, Rogowsky P, Christopher T, Song W, Langridge P (1991) Cloning and characterization of a new rye-specific repeated sequence. Genome 34:81–87

Gupta RB, MacRitchie F (1994) Allelic variation at the glutenin subunit and gliadin loci, *Glu-1*, *Glu-3* and *Gli-1*, of common wheats. II. Biochemical basis of the allelic effects on dough properties. J Cereal Sci 19:19–29

Halford NG, Forde J, Anderson OD, Greene FC, Shewry PR (1987) The nucleotide and deduced amino-acid sequences of an HMW glutenin subunit gene from chromosome 1B of bread wheat (*Triticum aestivum* L.) and comparison with those of genes from chromosomes 1A and 1D. Theor Appl Genet 75:117–126

Halford NG, Field JM, Blair H, Urwin P, Moore K, Robert L, Thompson R, Flavell RB, Tatham AS, Shewry PR (1992a) Analysis of HMW glutenin subunits encoded by chromosome 1A of bread wheat (*Triticum aestivum* L.) indicates quantitative effects on grain quality. Theor Appl Genet 83:373–378

Halford NG, Tatham AS, Sui E, Daroda L, Dreyer T, Shewry PR (1992b) Identification of a novel beta-turn-rich repeat motif in the D hordeins of barley. Biochim Biophys Acta 1122:118–122

Lafiandra D, D'Ovidio R, Porceddu E, Margiotta B, Colaprico G (1993) New data supporting high Mr glutenin subunit 5 as the determinant of quality differences among the pairs 5+10 vs 2+12. J Cereal Sci 18:197–205

Lawrence GJ, Shepherd KW (1981) Chromosomal location of genes controlling seed protein in species related to wheat. Theor Appl Genet 59:25–31

Mackie AM, Sharp PJ, Lagudah ES (1996a) The nucleotide and derived amino-acid sequence of a HMW glutenin gene from *Triticum tauschii* and comparison with those from the D genome of bread wheat. J Cereal Sci 24:73–78

Mackie AM, Lagudah ES, Sharp PJ, Lafiandra D (1996b) Molecular and biochemical characterization of HMW glutenin subunits from *T. tauschii* and the D genome of hexaploid wheat. J Cereal Sci 23:213–225

Miller TE (1987) Systematics and evolution. In: Lupton FGH (ed) Wheat breeding: its scientific basis. Chapman and Hall, New York, pp 1–49

Montebove L, De Pace C, Jan CC, Scarascia-Mugnozza GT, Qualset CO (1987) Chromosomal location of isozyme and seed storage protein genes in *Dasypyrum villosum* (L.) Candargy. Theor Appl Genet 73:836–845

Nei N, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, UK

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

Payne PI, Lawrence GJ (1983) Catalogue of alleles for the complex gene loci *Glu-A1, Glu-B1* and *Glu-D1*, which code for the high-molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun 11:29–35

Payne PI, Corfield KG, Holt LM, Blackman JA (1981) Correlation between the inheritance of certain high-molecular-weight subunits of glutenin and bread-making quality in progenies of six crosses of bread-wheat. J Sci Food Agric 32:51–60

Payne PI, Holt LM, Lawrence GJ (1983) Detection of a novel high-molecular-weight subunit of glutenin in some Japanese hexaploid wheats. J Cereal Sci 1:3–8

Peña RJ, William MDHM, Amaya A, Mujeeb-kazi A (1991) Quality evaluation of some *Triticum turgidum* × *T. tauschii* synthetic hexaploids and their potential for wheat improvement. (abstract). In: Agronomy abstracts, Annu Meet Am Soc Agron. American Society of Agronomy, Madison, Wisconsin, USA, p 189

Pflüger LA, D'Ovidio R, Margiotta B, Peña R, Mujeeb-kazi A, Lafiandra D (2001) Characterisation of high- and low-molecular-weight glutenin subunits associated to the D genome of *Aegilops tauschii* in a collection of synthetic hexaploid wheats. Theor Appl Genet 103:1293–1301

Reddy P, Appels R (1993) Analysis of a genomic segment carrying the wheat high-molecular-weight glutenin Bx17 subunit gene and its use as an RFLP marker. Theor Appl Genet 85:616–624

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Shewry PR, Tatham AS, Barro P, Lazzeri P (1995) Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. Bio/technology 13:1185–1190

Sugiyama T, Rafalski A, Peterson D, Soll D (1985) A wheat HMW glutenin subunit gene reveals a highly repeated structure. Nucleic Acids Res 13:8729–8737

Thompson RD, Bartels D, Harberd NP (1985) Nucleotide sequence of a gene from chromosome 1D of wheat encoding a HMW-glutenin subunit. Nucleic Acids Res 13:6833–6846

Thompson JD, Higgins DG, Gibson TJ (1994) 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

Urbano M, Colaprico G, Margiotta B, Perrino P, Lafiandra D (1993) Gliadins and high-molecular-weight glutenin subunits variation in *Aegilops searsii*. In: Damania AB (ed) Biodiversity

and wheat improvement. John Wiley and Sons, Chichester, pp 299–308

Wan Y, Liu K, Wang D, Shewry PR (2000) High-molecular-weight glutenin subunits in the *Cylindropyrum* and *Vertebrata* section of the *Aegilops* genus and identification of subunits related to those encoded by the Dx alleles of common wheat. Theor Appl Genet 101:879–884

Wan Y, Wang D, Shewry PR, Halford NG (2002) Isolation and characterization of five novel high-molecular-weight subunit genes from *Triticum timopheevi* and *Aegilops cylindrica*. Theor Appl Genet 104:828–839

William MDHM, Peña RJ, Mujeeb-Kazi A (1993) Seed protein and isozyme variations in *Triticum tauschii* (*Aegilops squarrosa*). Theor Appl Genet 87:257–263

Xie R, Wan Y, Zhang Y, Wang D (2001) HMW glutenin subunits in multiploid *Aegilops* species: composition analysis and molecular cloning of coding sequences. Chinese Sci Bull 46:309–313

Zhang H, Jia J, Gale MD, Devos KM (1998) Relationships between the chromosomes of *Aegilops umbellulata* and wheat. Theor Appl Genet 96:69–75

Zhong GY, Qualset CO (1993) Allelic diversity of high-molecular-weight glutenin protein subunits in natural populations of *Dasypyrum villosum* (L.) Candargy. Theor Appl Genet 86:851–858