Theor Appl Genet (2006) 113:631–641
DOI 10.1007/s00122-006-0327-x

ORIGINAL PAPER

Characterization of the HMW glutenin subunits from *Aegilops searsii* and identification of a novel variant HMW glutenin subunit

Xia Sun · Shanglian Hu · Xin Liu · Weiqiang Qian · Shanting Hao · Aimin Zhang · Daowen Wang

Received: 3 January 2006 / Accepted: 24 May 2006 / Published online: 1 July 2006

Abstract High molecular weight (HMW) glutenin subunits are conserved seed storage proteins in wheat and related species. Here we describe a more detailed characterization of the HMW glutenin subunits from *Aegilops searsii*, which is diploid and contains the S^s genome related to the S genome of *Aegilops speltoides* and the A, B and D genomes of hexaploid wheat. SDS-PAGE experiments revealed two subunits (one x and one y) for each of the nine *Ae. searsii* accessions analyzed, indicating that the HMW glutenin subunit gene locus of *Ae. searsii* is similar to the *Glu-1* locus found in wheat in containing both *x* and *y* genes. The primary structure of the four molecularly cloned subunits (from two *Ae. searsii* accessions) was highly similar to that of the previously reported x and y subunits. However, in one accession (IG49077), the last 159 residues of the x subunit ($1S^s$x49077), which contained the sequence element GHCPTSPQQ, were identical to those of the y subunit ($1S^s$y49077) from the same accession. Consequently, $1S^s$x49077 contains an extra cysteine residue located at the C-terminal part of its repetitive domain, which is novel compared to the x-type subunits reported so far. Based on this and previous studies, the structure and expression of the *Glu-1* locus in *Ae. searsii* is discussed. A hypothesis on the genetic mechanism generating the coding sequence for the novel $1S^s$x49077 subunit is presented.

Communicated by J. W. Snape

X. Sun · X. Liu · W. Qian · S. Hao · A. Zhang · D. Wang (✉)
The State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
e-mail: dwwang@genetics.ac.cn

X. Sun
College of Agriculture, Northeast Agricultural University, Harbin 150030, China

X. Sun · S. Hu
Southwest University of Science and Technology, Mianyang 621010, China

Introduction

High molecular weight (HMW) glutenin subunits constitute one of the major classes of seed storage proteins in wheat and related species (Payne 1987; Shewry et al. 1989). Their expression, structure and function have been intensively studied in hexaploid wheat because they play an important role in determining the end use qualities of bread wheat varieties (Payne et al. 1981; Shewry et al. 2003c). Genetic investigations have shown that, in hexaploid wheat, HMW glutenin subunits are encoded by the *x* and *y* genes contained in the *Glu-A1*, *Glu-B1* and *Glu-D1* loci (Brown et al. 1979; Lawrence and Shepherd 1981; Shewry et al. 2003b). Owing to allelic variation and gene silencing, bread wheat varieties usually express three to five subunits, and the compositions of the HMW glutenin subunits often differ among different varieties (Shewry et al. 2003c). The x and y subunits share a highly similar primary structure, which consists of a signal peptide (removed from the mature protein), a N-terminal domain, a central repetitive domain and a C-terminal domain (Shewry et al. 1995). Most of the reported x-type subunits possess four conserved cysteine residues (three in the N-terminal domain, one in the C-terminal domain), and the majority of the y-type subunits

characterized thus far contain seven conserved cysteine residues (five in the N-terminal domain, one in the repetitive domain and one in the C-terminal domain) (Shewry et al. 1995). These cysteine residues are involved in the formation of disulphide bonds within and between subunits and are thus important for the high order structure and the functionality of these proteins in shaping the elastic properties of the gluten complex in wheat dough (Shewry et al. 1995). In both x- and y-type subunits, the repetitive domains are composed of short and repeated peptide (tripeptide, hexapeptide, nanopeptide) motifs, with the presence of the tripeptide motif being unique to the repetitive domain of x-type subunits (Shewry et al. 1995). Based on available studies, the HMW glutenin subunits from various *Aegilops* species and rye have a primary structure similar to that of the wheat subunits (Mackie et al. 1996; Xie et al. 2001; Wan et al. 2002, 2005; Yan et al. 2002; Liu et al. 2003). In contrast, the D-hordeins from barley are more distantly related to the HMW glutenin subunits from wheat and *Aegilops* species (Shewry et al. 1989; Halford et al. 1992).

Although the general organization of the primary structure of different HMW glutenin subunits is well conserved, novel variations do occur to their amino acid sequences. One of the most common variations is changes in the number of cysteine residues, and these variations have been exploited for both applied and basic studies on these proteins. For example, the 1Dx5 subunit has an extra cysteine residue located at the N-terminal part of its repetitive domain and this subunit has frequently been found to associate with improved processing quality in bread wheat varieties (Lafiandra et al. 1993; Gupta and MacRitchie 1994). More recently, a novel variant of the 1Ax2* subunit, 1Ax2*^{B}, is found to contain an extra cysteine residue located in the middle of its repetitive domain (Juhász et al. 2003). This subunit exerts a positive effect on the gluten properties. In contrast to 1Dx5 and 1Ax2*^{B}, the 1Bx14 and 1Bx20 subunits have reduced numbers of cysteine residues in their N-terminal domains (Shewry et al. 2003a; Li et al. 2004). This property has been employed in studying the evolution of 1Bx subunits and led to the division of two evolutionary lineages for the *1Bx* alleles (Li et al. 2004). Although the size of the N- and C-terminal domains is highly conserved in either x- or y-type subunits, that of the repetitive domains is often variable. For instance, the repetitive domains of the 1Dx2.2 and 1Dx2.2* from bread wheat and the 1Ux subunit from *Aegilops umbellulata* subunits are much larger than those of other reported x subunits (Payne and Lawrence 1983; Payne et al. 1983; Liu et al. 2003). Molecular analysis has shown that the extra portion in the 1Ux repetitive domain contains 111 amino acid residues that form 9 tripeptides, 11 hexapeptides and 2 nanopeptides (Liu et al. 2003). The larger size of the 1Dx2.2* repetitive domain has been deduced to possibly associate with an insertion, which may be caused by unequal crossing over in its coding sequence (D′Ovidio et al. 1996). More recent molecular analysis has shown that the larger size of 1Dx2.2 and 1Dx2.2* is owing to the occurrence of internally duplicated regions within their repetitive domains (Wan et al. 2005).

Aegilops searsii ($S^{s}S^{s}$, $2n = 2x = 14$) belongs to section *Sitopsis* of the genus *Aegilops* (Feldman and kislev 1977; van Slageren 1994). In addition to *Ae. searsii*, the *Sitopsis* section also contains *Aegilops bicornis* ($S^{b}S^{b}$, $2n = 2x = 14$), *Aegilops longissima* ($S^{l}S^{l}$, $2n = 2x = 14$), *Aegilops sharonensis* ($S^{l}S^{l}$, $2n = 2x = 14$) and *Aegilops speltoides* (SS, $2n = 2x = 14$) (van Slageren 1994). The genomes in the five species have been found closely related based on cytogenetic and molecular genetic investigations (Kihara 1954; Yen and Kimber 1990; Zhang et al. 2001, 2002; Salina et al. 2004; Sasanuma et al. 2004). The expression of HMW glutenin subunits in *Ae. searsii* has been investigated in the past using SDS-PAGE analysis (Fernández-Calvín and Orellana 1990; Urbano et al. 1993). In the present work, we studied the HMW glutenin subunits from nine *Ae. searsii* accessions using both electrophoretic and molecular cloning approaches. The complete open reading frames (ORFs) of the *x* and *y* genes from two different *Ae. searsii* accessions were characterized. Comparative analysis of the deduced amino acid sequences of the cloned *Ae. searsii* subunits with those of previously published subunits led to the discovery of a novel x-type subunit from the *Ae. searsii* accession IG49077 that has hereto not been reported in wheat and other *Triticeae* species.

Materials and methods

Plant materials

Nine accessions of *Ae. searsii* (IG47307, IG47353, IG47388, IG47395, IG47423, IG47461, IG48586, IG48798 and IG49077), all obtained from The International Center for Agricultural Research in the Dry Areas, ICARDA) were used in this study. The HMW glutenin subunits from the bread wheat variety Chinese Spring (having the HMW glutenin subunits 1Bx7+1By8, 1Dx2+1Dy12) were used as standards for comparing the electrophoretic mobilities of the HMW glutenin subunits from the different *Ae. searsii* accessions in SDS-PAGE experiments.

SDS-PAGE experiments and N-terminal protein sequencing

General conditions for preparing seed protein extracts and SDS-PAGE experiments were those described previously (Wan et al. 2002). To examine the composition of HMW glutenin subunits in *Ae. searsii* more vigorously, at least five individual seeds were analyzed for each of the nine accessions. The running buffer for SDS-PAGE was either Tris–borate–SDS (124 mM Tris, 50 mM boric acid, 0.1% pH 8.9) or Tris–glycine–SDS (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). The results of the SDS-PAGE experiments were recorded using a digital camera (Coolpix 990, Nikon). N-terminal protein sequencing of the HMW glutenin subunits from selected *Ae. searsii* accessions (IG47388, IG48586, IG49077) was carried out as described previously (Wan et al. 2002).

Cloning and sequencing *Ae. searsii* HMW glutenin gene ORFs

Seeds of the *Ae. searsii* accessions IG48586 and IG49077 were germinated in the dark at 24°C. Genomic DNA samples were prepared from the etiolated seedlings following a previously published procedure (Guidet et al. 1991). For amplifying the complete ORFs of the HMW glutenin genes of *Ae. searsii* in genomic PCR reactions, two degenerate primers, P1 (5′-ATGGCTAAGCGGC/TTA/GGTCCTCTTTG-3′) and P2 (5′-CTATCACTGGCTG/AGCCGACAA TGCG-3′), were designed according to the conserved sequences located at the 5′ and 3′ ends of the published ORF sequences of HMW glutenin subunits. P1 contains the start codon of the genomic ORF whereas P2 possesses the two tandem stop codons that are present in the HMW glutenin ORFs characterized thus far. Genomic PCR reactions were conducted using the high fidelity DNA polymerase ExTaq (TaKaRa Biotechnology Co. Ltd, Kyoto, Japan). The parameters for the reactions were those reported previously (Liu et al. 2003). The PCR fragments of the expected size were recovered from agarose gels, followed by cloning into the pGEM-T Easy vector (Promega, Madison, WI, USA). Based on restriction enzyme digestion analysis and DNA sequencing, the inserts in the two plasmid clones p1S^{s}x48586 and p1S^{s}y48586 were found to contain the complete ORF sequences for the x- and y-type subunits of accession IG48586, respectively. Using the same strategy, the inserts in the two plasmid clones p1S^{s}x49077 and p1S^{s}y49077 were determined to harbor the complete ORF sequences for the x- and y-type subunits of accession IG49077, respectively. The complete ORF sequences were determined using overlapping subclones created by the nested deletion method (Sambrook et al. 1989). The correctness of the assembled ORF sequences was verified by partial sequencing of additional clones from independent PCR amplification and cloning experiments, and by in vitro expression in the bacterial cells (see below).

Comparison of deduced amino acid sequences and phylogenetic analysis

The HMW glutenin gene ORFs from the *Ae. searsii* accessions IG48586 and IG49077 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 previously published HMW glutenin subunits through multiple alignment using the Clustal W program (Thompson et al. 1994). For phylogenetic analysis using the amino acid sequence of the signal peptide plus the N-terminal domain, the MEGA program (http://www.oup-usa.org/sc/0195135857, Nei and Kumar 2000) was employed. During the phylogenetic analysis, the complete deletion option was used to deal with gaps in the aligned sequences, and the PC distance was calculated for each pair of the aligned sequences. The bootstrap values in the phylogenetic tree were obtained based on 500 replications. The D-hordein protein from barley (GenBank accession number Ay268139) was used as a control outgroup. The GenBank accession numbers for the published HMW glutenin subunits under comparison were M22208 (1Ax2*), X13927 (1Bx7), X12928 (1Dx5), AJ306977 (1Ay from *Triticum timopheevi*), X61026 (1By9) and X12929 (1Dy10). The ORF sequences of 1S^{s}x48586, 1S^{s}y48586, 1S^{s}x49077 and 1S^{s}y49077 determined in this work had been submitted to the GenBank with the accession numbers Ay611722, Ay611721, Ay611723 and Ay611724, respectively.

MALDI-TOF-MS analysis of native 1S^{s}x49077

The seed protein sample prepared from the accession IG49077 was separated using 10% SDS-PAGE with the Tris–glycine–SDS running buffer. After Coomassie staining, the protein band representing the x-type subunit of IG49077 (i.e., 1S^{s}x49077) was manually excised from the gel, followed by in gel digestion with trypsin (Kumarathasan et al. 2005). A sample of the digested protein (0.5 μl) was analyzed in an Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics, Billerica, MA, USA) over a mass range of m/z 1,000–23,000.

Mass spectra were calibrated externally using cytochrome c $(M+H)^+$ (*m*/*z* 12,361.09 Da) and cytochrome c $(M+2H)^{2+}$ (*m*/*z* 6,181.05 Da). The MALDI-TOF measured peptide mass spectra were compared to the calculated mass spectra of trypsin digested 1S^{s}x49077, which were predicted based on the deduced amino acid sequence of 1S^{s}x49077 using the bioinformatic program PeptideMass (http://www.expasy.ch/tools/peptide-mass.html). The MS analysis was repeated three times using the protein samples prepared from separate batches of seeds.

In vitro expression of the cloned *Ae. searsii* HMW glutenin gene ORF sequences in the bacterial cells

To further verify the correctness of the cloned *Ae. searsii* HMW glutenin gene ORF sequences, they were expressed in the bacterial cells after removing the coding sequence for the signal peptide. This ensured the synthesis of the mature proteins of the cloned subunits, which should possess electrophoretic mobilities comparable to those of the native subunits from the seeds. The removal of the coding sequence for the signal peptide from each of the four cloned ORF sequences was accomplished using PCR mutagenesis with the primers listed in Table 1. The conditions for amplifying the mutant ORFs were identical to those described above, but the template was plasmid DNA purified from the clones p1S^{s}x48586, p1S^{s}y48586, p1S^{s}x49077 and p1S^{s}y49077, respectively. Four bacterial expression constructs, pET-1S^{s}x48586, pET-1S^{s}y48586, pET-1S^{s}x49077 and pET-1S^{s}y49077, were prepared by cloning the mutant ORFs individually into the bacterial expression vector pET-30a (Invitrogen, San Diego, CA, USA). The induction of the pET constructs by IPTG (Isopropyl-β-D-thiogalactopyranoside, Sigma, St Louis, MO, USA) was carried out as described by Wan et al. (2002). For SDS-PAGE analysis, protein extracts were prepared by directly dissolving the bacterial cells in the SDS-PAGE sample buffer. The running buffer was Tris–glycine–SDS (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). To confirm the identity of the bacterially expressed subunits, Western blotting analysis was conducted using a polyclonal antibody specific for HMW glutenin subunits as reported previously (Wan et al. 2000).

Results

Composition of HMW glutenin subunits in nine *Ae. searsii* accessions

Seed protein extracts were prepared from nine accessions of *Ae. searsii* and the bread wheat variety Chinese Spring. They were then separated using SDS-PAGE. Two subunits (Fig. 1, indicated by arrowheads and arrows, respectively) were reproducibly detected for each of the nine *Ae. searsii* accessions. By analogy with the electrophoretic patterns of the orthologous wheat subunits, the subunits displaying slower electrophoretic mobilities (marked by arrowheads) were likely to be of x-type, whereas those showing faster electrophoretic mobilities (marked by arrows) were probably of y-type. Interestingly, the electrophoretic mobilities of the x and y subunits from the nine accessions were more similar to those of the B genome encoded subunits than to those of the D or A genome derived subunits (Fig. 1, data not shown). Using direct protein sequencing, the N-terminal amino acid residues were determined for several subunits with slower electrophoretic mobilities. For those from the accessions IG48586 and IG49077, the obtained N-terminal amino acid sequence was KGEAQGQLQCERELQ. For IG47388, the corresponding sequence was EGEAQGQLQCERELQ. These sequences were highly similar to the first 15 amino acid residues (EGEASG/EQLQCERELQ) in the N-terminal domains of published 1Ax and 1Dx subunits, thus confirming that the slow-moving subunits of the *Ae. searsii* accessions were indeed of x-type. The fast-moving proteins were subsequently confirmed to be y-type subunits by molecular cloning experiments (see below).

Table 1 PCR primers used for preparing the bacterial expression constructs

Primer	Sequence (5′ to 3′)[a]	Use
E1	ACCCATATGAAAGGTGAGGCCTCT	Preparing pET-1S^{s}x48586
E2	CTAGAATTCTCACTAGTGAATTCGC	
E3	ACCCATATGGAAGGTGAGGCCTCTG	Preparing pET-1S^{s}y48586
E4	CTAGAATTCTCACTAGTGAATTCGC	
E5	ACCCATATGAAAGGTGAGGCCTCTG	Preparing pET-1S^{s}x49077
E6	CTAGAATTCTCACTAGTGAATTCGC	
E7	ACCCATATGGAAGGTGAGGCCTCTG	Preparing pET-1S^{s}y49077
E8	GGGGAATTCTCACTAGTGAATTCGC	

[a] The underlined nucleotides form *Eco*RI (GAATTC) or *Nde*I (CATATG) restriction sites

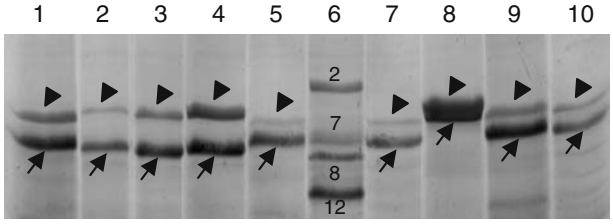

Fig. 1 SDS-PAGE analysis of the HMW glutenin subunits in the seed extracts of the *Ae. searsii* accessions IG47307 (*lane 1*), IG47353 (*lane 2*), IG47388 (*lane 3*), IG47395 (*lane 4*), IG47423 (*lane 5*), IG47461 (*lane 7*), IG48586 (*lane 8*), IG48798 (*lane 9*) and IG49077 (*lane 10*). By analog with the HMW glutenin subunits from the bread wheat variety Chinese Spring (1Dx2, 1Bx7, 1By8, 1Dy12, *lane 6*), the protein species with slower electrophoretic mobilities (indicated by *arrowheads*) represented the x-type subunits from the nine *Ae. searsii* accessions. On the other hand, the protein bands with faster electrophoretic mobilities (indicated by *arrows*) were the y-type subunits from the nine *Ae. searsii* accessions. The gel was run using the Tris–borate–SDS buffer. Under this condition the two subunits from IG48586 did not separate completely. However, better separation of the two subunits was achieved with the Tris–glycine–SDS buffer (see Fig. 5)

Isolation of HMW glutenin gene ORFs and analysis of derived amino acid sequences

Using genomic PCR with degenerate oligonucleotide primers, the DNA fragments covering the complete ORFs of the two subunits in IG48586 (tentatively named as 1S^{s}x48586 and 1S^{s}y48586) and their orthologs in IG49077 (1S^{s}x49077 and 1S^{s}y49077) were amplified (Fig. 2a). The four DNA fragments were cloned and sequenced, and were found not to contain any introns, which is a general feature of the HMW glutenin gene ORFs characterized so far (Shewry et al. 1989, 2003c). Comparison of the amino acid sequences deduced from the cloned fragments indicated that the four subunits from *Ae. searsii* shared a primary structure identical to that of previously described x and y subunits, which is composed of a signal peptide, a N-terminal domain, a central repetitive domain and a C-terminal domain (Fig. 2b; Table 2). The size of the N- and C-terminal domains in the four *Ae. searsii* subunits was typical of that of previously reported x and y subunits (Fig. 2b; Table 2). Except for 1S^{s}x49077, the remaining three subunits all contained the expected numbers of cysteine residues at the anticipated positions (Fig. 2b; Table 2). There was an extra cysteine residue in 1S^{s}x49077 (Fig. 2b; Table 2). This residue was located within the sequence element GHCPTSPQQ, which is, however, not present in any of the previously characterized x-type subunits (Shewry et al. 1995, 2003c). Further examination revealed that the last 159 residues of 1S^{s}x49077 were identical to those of 1S^{s}y49077 (Fig. 2c). The 159 residues encompassed the C-terminal part of the repetitive domain (117 residues) and the entire C-terminal domain (42 residues) (Fig. 2c). Moreover, the nucleotide sequences encoding the corresponding portions of the two subunits were also 100% identical (data not shown). Like their orthologous subunits, the repetitive domains in the four *Ae. searsii* subunits were composed of repetitive motifs, with the presence of the tripeptides only found in the two x subunits (Table 3). The size of the repetitive domains in the four subunits varied to some extent, which was caused by differences in the numbers of repetitive amino acid motifs forming the individual domains (Tables 2, 3). In general, the amino acid sequences of the four subunits showed higher identities with those of the subunits encoded by the D genome of hexaploid bread wheat. Phylogenetic analyses using the amino acid sequences of the signal peptides plus the N-terminal domains revealed that 1S^{s}x48586 and 1S^{s}x49077 clustered with 1Dx2 and 1Dx5, and 1S^{s}y48586 and 1S^{s}y49077 with 1Dy10 and 1Dy12 (Fig. 3).

Peptide mass fingerprint of native 1S^{s}x49077

The unusual feature toward the C-terminal region of the deduced amino acid sequence of 1S^{s}x49077 prompted us to investigate the primary structure of the native 1S^{s}x49077 subunit. Because mass spectrometry has been found effective in gaining structural information of the HMW glutenin subunits directly isolated from seeds (Cunsolo et al. 2002, 2003), native 1S^{s}x49077 was subject to MALDI-TOF-MS analysis. Using the trypsin digested 1S^{s}x49077, the molecular mass of nine peptides could be reliably determined in repeated MS experiments (Table 4). The molecular mass values of the nine peptides matched closely with their calculated mass values (Table 4). Together, the nine peptides covered a substantial portion of the deduced amino acid sequence of 1S^{s}x49077 (Fig. 4). Consequently, the deduced sequence of 1S^{s}x49077 was likely to be a true representation of that of the native subunit. A closer inspection of the peptide mass fingerprint data revealed that the N-terminal domain of native 1S^{s}x49077 possessed typical features of x-type subunits (such as the presence of three conserved cysteine residues, Fig. 4), whereas its C-terminal domain was characteristic of that of y-type subunits (compare Figs. 2c, 4).

Bacterial expression of the HMW glutenin gene ORFs from the *Ae. searsii* accessions IG49077 and IG48586

After removing the coding sequence for the signal peptide from each of the four cloned ORFs, four bacterial expression constructs (pET-1S^{s}x48586, pET-1S^{s}y48586,

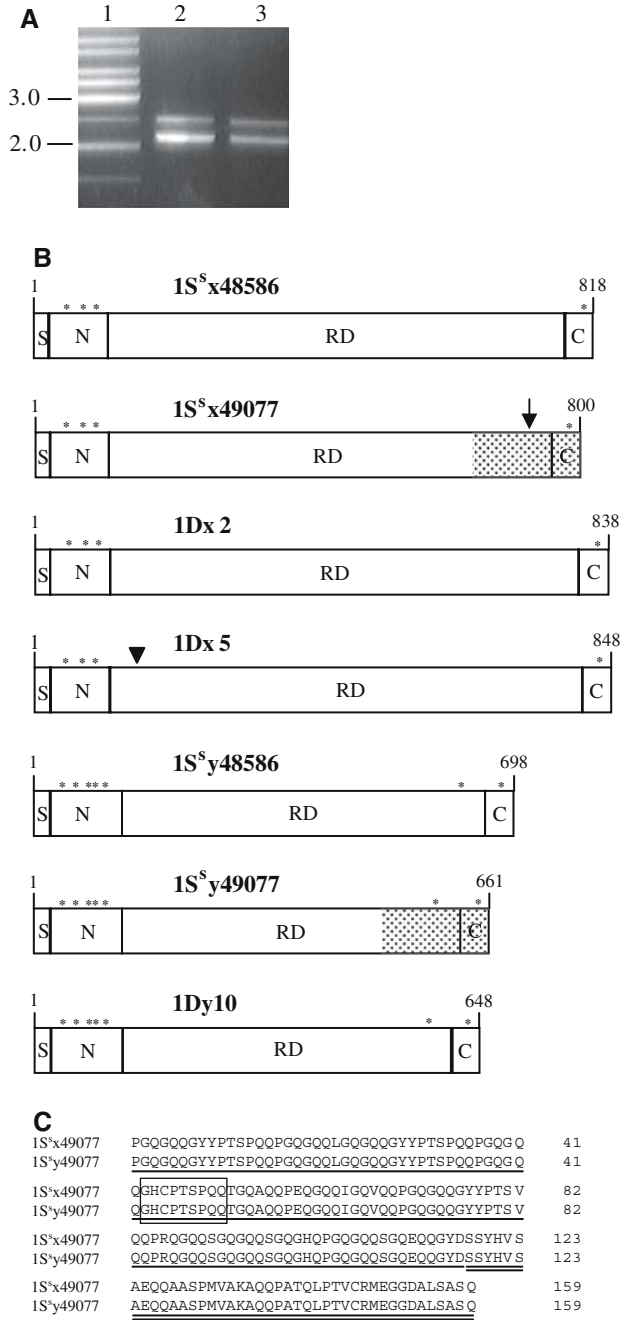

Fig. 2 Amplification of the complete genomic ORFs of the x- and y-type subunits from the *Ae. searsii* accessions IG48586 and IG49077 by genomic PCR and analysis of the primary structure of four cloned *Ae. searsii* subunits. **a** Two PCR fragments (representing the complete ORF sequences for the x- and y-type subunits, respectively) were specifically amplified from the genomic DNA samples of IG48586 (*lane 2*) and IG49077 (*lane 3*). The DNA size markers (kb) were contained in *lane 1*. **b** The primary structure of the four cloned *Ae. searsii* subunits (1S^{s}x48586, 1S^{s}x49077, 1S^{s}y48586, 1S^{s}y49077) was identical to that of previously reported x- and y-type subunits from wheat in possessing a signal peptide (S), a N-terminal domain (N), a central repetitive domain (RD) and a C-terminal domain (C). The conserved cysteine residues in the x- and y-type subunits were indicated by *asterisks*. The extra cysteine residue in the N-terminal part of the repetitive domain of 1Dx5 was represented by an *arrowhead*. Interestingly, the 1S^{s}x49077 and 1S^{s}y49077 subunits shared an identical region (*shaded area*) in their proteins. This caused the presence of an extra cysteine residue in the C-terminal part of the repetitive domain of 1S^{s}x49077 (indicated by *arrow*). **c** A comparison of the last 159 residues [117 in the repetitive domain (*underlined*), 42 in the C-terminal domain (*doubly underlined*)] of 1S^{s}x49077 and 1S^{s}y49077. This shows that the extra cysteine residue in the C-terminal part of the repetitive domain of 1S^{s}x49077 resides in the nanopeptide GHCPTSPQQ (*boxed region*)

pET-1S^{s}x49077 and pET-1S^{s}y49077) were prepared for synthesizing the 1S^{s}x49077, 1S^{s}y49077, 1S^{s}x48586 and 1S^{s}y48586 subunits, respectively, in the bacterial cells. In the cells harboring p1S^{s}x49077, IPTG induction led to the overexpression of a protein species with an electrophoretic mobility similar to that of the native 1S^{s}x49077 subunit from the seed extract of IG49077 (Fig. 5a, left panel). The overexpressed protein species was confirmed to be a bacterially synthesized 1S^{s}x 49077 in subsequent Western-blot experiment using a polyclonal antibody specific for HMW glutenin subunits (Fig. 5a, right panel). Using similar strategy, the induction of p1S^{s}y49077, p1S^{s}x48586 and p1S^{s}y 48586 constructs by IPTG was found to lead to the synthesis of the 1S^{s}y49077, 1S^{s}x48586 and 1S^{s}y 48586 subunits, respectively, in the bacterial cells (Fig. 5a, b). Combined, these results suggested that the ORF sequences isolated in this work were faithful representation of the native HMW glutenin subunit genes in the two *Ae. searsii* accessions.

Discussion

Owing to their role in determining the elastic properties of the gluten complex and their conservation in wheat and related species, continued mining of novel HMW glutenin subunits from cultivated and wild germplasms is beneficial for not only improving wheat end use qualities but also studying the structure variation and evolution of this important family of proteins. Although the available studies have indicated that *Aegilops* species is a rich source for novel variants of HMW glutenin subunits (Wan et al. 2002, 2005; Yan et al. 2002; Liu et al. 2003), the progress in characterizing the expression and primary structure of orthologous subunits from more *Aegilops* species and at the molecular level remains slow. To contribute to this area of research, the present study has investigated more systematically the expression and primary structure of the HMW glutenin subunits from *Ae. searsii*

Table 2 Selected properties of the mature proteins of the *Ae. searsii* subunits compared to the representatives of the previously reported HMW glutenin subunits from wheat

Subunit	Number of amino acid residues				Number of cysteine residues			
	N-terminal domain	C-terminal domain	Repetitive domain	Total	N-terminal domain	C-terminal domain	Repetitive domain	Total
1S^{s}x48586	86	42	669	797	3	1	0	4
1S^{s}x49077	86	42	651	779	3	1	1	5
1Ax2*	86	42	666	794	3	1	0	4
1Bx7	81	42	645	768	3	1	0	4
1Dx5	89	42	687	818	3	1	1	5
1S^{s}y48586	104	42	511	677	5	1	1	7
1S^{s}y49077	104	42	474	640	5	1	1	7
1Ay (*T. timopheevi*)	104	42	420	566	5	1	0	6
1By9	104	42	538	684	5	1	1	7
1Dy10	104	42	481	627	5	1	1	7

Table 3 Repetitive peptide motifs in the repetitive domains of the four *Ae. searsii* subunits characterized in this study

Subunit	Number of tripeptide	Number of hexapeptide	Number of nanopeptide
1S^{s}x48586	21	67	19
1S^{s}x49077	19	66	19
1S^{s}y48586	0	55	20
1S^{s}y49077	0	50	19

using complementary SDS-PAGE and molecular cloning approaches.

An analysis of the results obtained in this work enables us to draw the following conclusions. First, it is likely that, in most (if not all) *Ae. searsii* accessions, both x and y types of HMW glutenin subunits are expressed. This is supported by the SDS-PAGE analyses of the seed protein extracts in this work as well as the two previous reports (Fernández-Calvín and Orellana 1990; Urbano et al. 1993), and confirmed by the subsequent cloning of the complete ORFs for both x and y subunits from two different *Ae. searsii* accessions in the present study. This, plus the previous finding that the x and y subunits are specified by the 1S^s chromosome in *Ae. searsii* (Urbano et al. 1993), prompt us to propose that the chromosomal location and the expression of the *Glu-1* locus in *Ae. searsii* may be highly similar to that in wheat or other investigated *Aegilops* species. Second, the primary structure of the x and y types of HMW glutenin subunits from *Ae. searsii* is identical to that of the orthologous subunits studied previously. However, the x-type subunit (1S^{s}x49077) from the *Ae. searsii* accession IG49077 is unusual because the last 159 residues of its protein differed from those of previously isolated x-type subunits from wheat and related species, and were identical to the corresponding residues of the y-type subunit from the same accession. The unusual feature in the deduced amino acid sequence of 1S^{s}x49077 is unlikely to be an experimental artifact, because (1) MS analysis showed that the amino acid sequence of the C-terminal domain of native 1S^{s}x49077 was likely to be the same as the one predicted from the nucleotide sequence, and (2) the cloned ORF sequence of this subunit was found to direct the synthesis of the expected protein product. This suggests that important local variations may occur to the primary structure of the HMW glutenin subunits from some *Ae. searsii* accessions. These variations may not be visible when the subunits from the different accessions were compared in SDS-PAGE experiment, but may be uncovered in molecular investigations involving the cloning of complete ORFs and the comparison of deduced amino acid sequences. Third, phylogenetically, the HMW glutenin subunits from *Ae. searsii* may be more closely related to those encoded by the D genome of wheat. This is in line with the finding

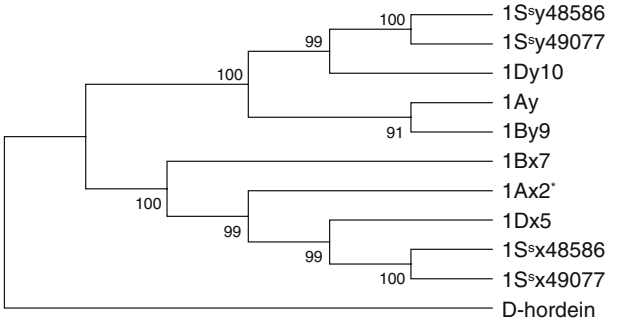

Fig. 3 Phylogenetic relationships of the four cloned *Ae. searsii* subunits (1S^{s}x48586, 1S^{s}x49077, 1S^{s}y48586, 1S^{s}y49077) with the subunits encoded by the A, B or D genomes of wheat. The phylogenetic tree was built by the neighbor joining method. Trees with identical topology were obtained with alternative methods (maximum likelihood, parsimony) (data not shown). The bootstrap values were obtained using 500 replications

Table 4 MALDI-TOF-MS analysis of peptide mass fingerprint of native 1S^{s}x49077 subunit

Measured mass $(M+H)^+$	Calculated mass $(M+H)^+$	Missed cleavage	Peptide sequence predicted using the bioinformatic program PeptideMass (with trypsin digestion)[a]
2,230.165	2,330.764	2	KGEASGQLQCERELQERELK
2,344.163	2,344.257	2	ELQERELKACQQVMDQQLR
2,373.251	2,374.497	0	CHPVVVSPVAGQYEQQIVVPPK
2,513.198	2,514.330	0	GGSFYPGETTSPQQLQQSIFWR
5,156.404	5,157.880	0	YYPSITSPQQVSYYPGQASPQRP
			GQGQQSGQGQQGYYPTSPQQPGQR
13,170.010	13,168.738	0	QGQQGQQSGQGQQPEQGQQGQQLGQGQQGYYATSLQQSGQGQ
			PGYYPTSLQQPGQGQSEYYPTSLQQPGQGQQPGQLQQPGQGQQGQ
			QSGQGQQGQQPGQGHQPGQGQPGYYPTSPQQSGQGQPR
6,491.962	6,495.174	1	QGQQSGQGQQSGQGHQPGQGQQSGQEQQGYDSSYHVSAEQQAA
			SPMVAKAQQPATQLPTVCR
1,411.724	1,413.075	0	AQQPATQLPTVCR
1,064.444	1,066.510	0	MEGGDALSASQ

[a] PeptideMass is available at http://www.expasy.ch/tools/peptide-mass.html

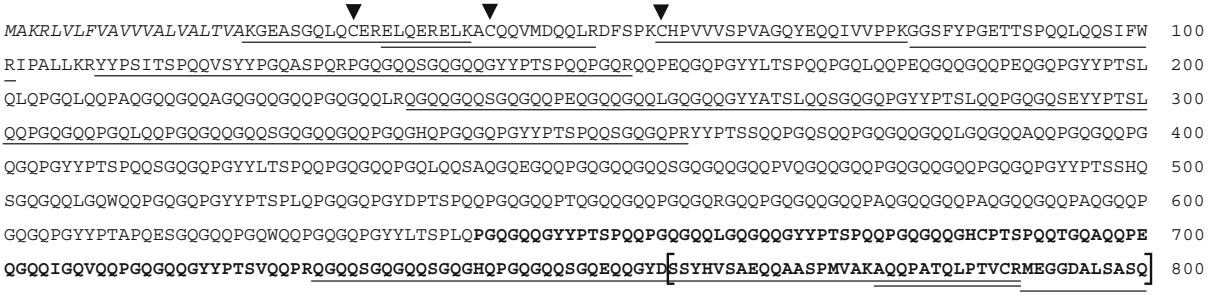

Fig. 4 Location of the nine peptides, which were detected by MALDI-TOF-MS analysis of native 1S^{s}x49077, in the deduced amino acid sequence of this subunit. The nine peptides are *underlined*. The last 159 residues of 1S^{s}x49077 are shown in *bold*. The 42 residues, which form a C-terminal domain characteristic of y-type HMW glutenin subunits, were bordered by *brackets*. The three cysteine residues (typical of x-type HMW glutenin subunits) are indicated by *filled arrowheads*. The signal peptide (*italicized region*) was not detected in the native 1S^{s}x49077 subunit by MS analysis

that the S^l genome of *Ae. longissima*, with which the S^s genome of *Ae. searsii* is closely related (Feldman et al. 1979; Yen and Kimber 1990; Salina et al. 2004; Sasanuma et al. 2004), is largely collinear with the D genome of common wheat (Zhang et al. 2001). In contrast to the closer phylogenetic relationship between the subunits encoded by the S^s and D genomes, in both our work and a previous investigation (Fernández-Calvín and Orellana 1990), the patterns of the electrophoretic mobilities of the *Ae. searsii* HMW glutenin subunits appear to be much more similar to those of the subunits encoded by the B genome of wheat. The electrophoretic patterns of the gliadin proteins of *Ae. searsii* are also highly similar to those specified by the B genome of common wheat (Urbano et al. 1993). The implications of these similarities are not clear at present, but will be a subject for future studies.

What might be the mechanism underlying the evolution of the novel x-type subunit in the *Ae. searsii* accession IG49077? Based on the finding that the last 159 residues of 1S^{s}x49077 were identical to those of the y-type subunit, we hypothesize that unequal crossing over between the two *Glu-1* loci located on the sister chromatids of the homologous chromosomes during meiosis might have led to the formation an unusual hybrid ORF, which underlies the novel feature in 1S^{s}x49077 (Fig. 6). Judging from the fact that the nucleotide sequences encoding the last 159 residues in the x and y subunits of IG49077 were 100% identical, the evolution of this hybrid HMW glutenin gene ORF might have occurred very recently. The model presented in Fig. 6 also predicts the occurrence of a hybrid ORF that would encode a y-type subunit with its C-terminal end identical to the corresponding portion of a x-type subunit. This possibility is currently being investigated by characterizing HMW glutenin gene ORFs from more *Ae. searsii* accessions.

Accompanying the unusual feature in the last 159 residues, 1S^{s}x49077 contained an extra cysteine residue at the C-terminal end of its repetitive domain. Because

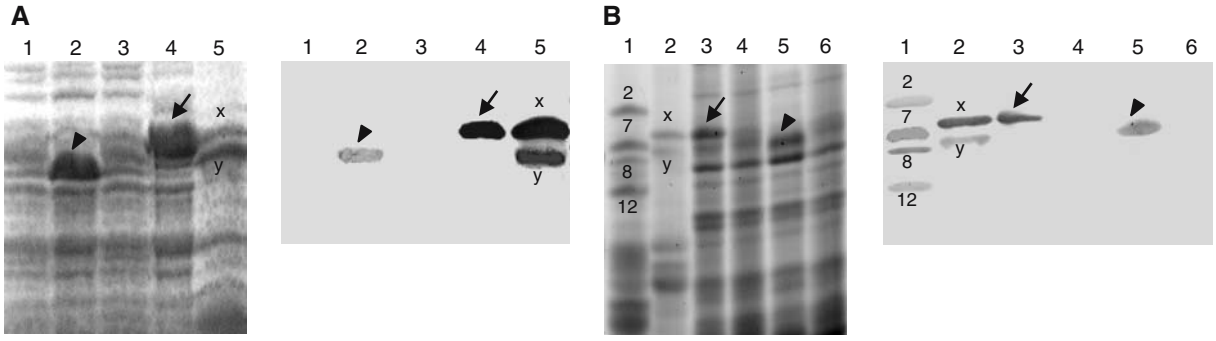

Fig. 5 In vitro expression of modified *Ae. searsii* HMW glutenin gene ORFs in the bacterial cells and SDS-PAGE and Western-blot analyses of the expressed products. The coding sequence for the signal peptide was removed from each of the four cloned *Ae. searsii* HMW glutenin gene ORFs via PCR mutagenesis. The modified ORFs were then used to generate the bacterial expression constructs pET-1S^{s}x48586, pET-1S^{s}y48586, pET-1S^{s}x49077 and pET-1S^{s}y49077. **a** SDS-PAGE and Western-blot analyses of the products expressed from the constructs pET-1S^{s}x48586 or pET-1S^{s}y48586. In the *left panel*, IPTG induction led to the overexpression of 1S^{s}x48586 (indicated by *arrow*, *lane 4*) and 1S^{s}y48586 (indicated by *arrowhead*, *lane 2*) in the bacterial cells. The bacterially synthesized subunits possessed similar electrophoretic mobilities as the native subunits from the seed extract (*lane 5*). The overexpressed products were not observed in the uninduced bacterial cells *(lanes 1* and *3*). In the *right panel*, the bacterially synthesized 1S^{s}x48586 (*lane 4*) and 1S^{s}y48586 (*lane 2*) subunits and the native subunits (*lane 5*) were recognized by a polyclonal antibody specific for HMW glutenin subunits in Western-blot experiment. **b** SDS-PAGE and Western-blot analyses of the products expressed from the constructs pET-1S^{s}x49077 or pET-1S^{s}y49077. In the *left panel*, overexpression of 1S^{s}x49077 (indicated by *arrow*, *lane 3*) and 1S^{s}y49077 (indicated by *arrowhead*, *lane 5*) was observed in the IPTG induced bacterial cells. The bacterially synthesized subunits and the native subunits from the seed extract (*lane 2*) displayed similar electrophoretic mobilities. No overexpressed products were found in the uninduced bacterial cells (*lanes 4* and *6*). In the *right panel*, Western-blot experiment showed that the bacterially synthesized 1S^{s}x49077 (*lane 3*) and 1S^{s}y49077 (*lane 5*) subunits, like the native subunits (*lane 2*) and the subunits of the hexaploid wheat variety Chinese Spring (*lane 1*), could be detected by a polyclonal antibody specific for HMW glutenin subunits

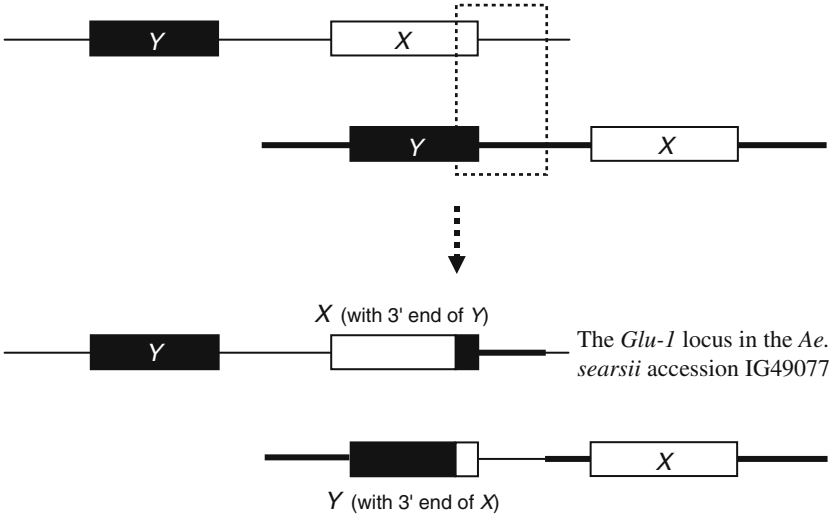

Fig. 6 A hypothesis on the mechanism that may generate the coding sequence for the 1S^{s}x49077 subunit in the *Ae. searsii* accession IG49077. Unequal crossing-over between the *Glu-1* loci located on the sister chromatids of the homologous chromosomes may have occurred, resulting in a double crossing over between the 3′ end regions of the *x* and *y* genes. This would generate two hypothetical recombinant *Glu-1* loci, one of which would resemble the *Glu-1* locus found in the *Ae. searsii* accession IG49077 in this work. In this hypothetical model, the *left boundary* involved in the crossing over resides in the coding sequences of the repetitive domains of the recombining *x* and *y* genes, whereas the *right boundary* is outside the coding sequences of the C-terminal domains of the two genes

the sequence element containing the extra cysteine residue (GHCPTSPQQ) is present in all y-type HMW glutenin subunits characterized so far (Shewry et al. 1995), it is very likely that this cysteine residue is originally evolved in a y-type subunit. The presence of the GHCPTSPQQ element in 1S^{s}x49077 makes it the third

reported x-type subunit that possesses cysteine residue in the repetitive domain. Because of the demonstrated associations between the presence of extra cysteine residue in the repetitive domain and superior functionality of the 1Dx5 and 1Ax2*^{B} subunits (Lafiandra et al. 1993; Gupta and MacRitchie 1994; Juhász et al. 2003), it will be important to study if 1S^{s}x49077 would confer improved processing property to wheat flour after being expressed in bread wheat varieties. For this purpose, we have prepared plant expression construct using the molecularly verified ORF sequence of 1S^{s}x49077, and are currently in the process of developing transgenic wheat lines that express 1S^{s}x49077 in the endospermic tissues.

Acknowledgments We thank Professor Moshe Feldman for constructive comments on this work and manuscript preparation. This work is supported by grants from Ministry of Science and Technology of China (2002CB111301, 2003AA222040) and the Chinese Academy of Sciences (KSCX2-SW-304).

References

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

Cunsolo V, Foti S, Saletti R, Gilbert S, Tatham AS, Shewry PR (2002) Investigation and correction of the gene-derived sequence of glutenin subunit 1Dx2 by matrix-assisted laser desorption/ionisation mass spectrometry. Rapid Commun Mass Spectrom 16:1911–1918

Cunsolo V, Foti S, Saletti R, Gilbert S, Tatham AS, Shewry PR (2003) Structural studies of glutenin subunits 1Dy10 and 1Dy12 by matrix-assisted laser desorption/ionisation mass spectrometry and high-performance liquid chromatography/electrospray ionisation mass spectrometry. Rapid Commun Mass Spectrom 17:442–454

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

Feldman M, Kislev M (1977) *Aegilops searsii*, a new species of section *Sitopsis* (*Platystachys*). Isr J Bot 26:190–201

Feldman M, Strauss I, Vardi A (1979) Chromosome pairing and fertility of F1 hybrids of *Aegilops longissima* and *Ae.* searsii. Can J Genet Cytol 21:261–272

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

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

Juhász A, Larroque OR, Tamás L, Hsam SL, Zeller FJ, Békés F, Bedő Z (2003) Bánkúti 1201—an old Hungarian wheat variety with special storage protein composition. Theor Appl Genet 107:697–704

Kihara H (1954) Considerations on the evolution and distribution of *Aegilops* species based on the analyzer-method. Cytologia 19:336–357

Kumarathasan P, Mohottalage S, Goegan P, Vincent R (2005) An optimized protein in-gel digest method for reliable proteome characterization by MALDI-TOF-MS analysis. Anal Biochem 346:85–89

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

Li W, Wan Y, Liu Z, Liu K, Liu X, Li B, Li Z, Zhang X, Dong Y, Wang D (2004) Molecular characterization of HMW glutenin subunit allele *1Bx14*: further insights into the evolution of *Glu-B1-1* alleles in wheat and related species. Theor Appl Genet 109:1093–1104

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

Mackie AM, Sharp PJ, Lagudah ES (1996) 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

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

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

Salina EA, Adonina IG, Vatolina TY, Kurata N (2004) A comparative analysis of the composition and organization of two subtelomeric repeat families in *Aegilops speltoides* Tausch. and related species. Genetica 122:227–237

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

Sasanuma T, Chabane K, Endo TR, Valkoun J (2004) Characterization of genetic variation in and phylogenetic relationships among diploid *Aegilops* species by AFLP: incongruity of chloroplast and nuclear data. Theor Appl Genet 108:612–618

Shewry PR, Halford NG, Tatham AS (1989) The high molecular weight subunits of wheat, barley and rye: genetics, molecular biology, chemistry and role in wheat gluten structure and functionality. In: Miflin BJ (ed) Oxford surveys of plant molecular and cell biology, vol. 6. Oxford University Press, Oxford, pp. 163–219

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

Shewry PR, Gilbert SMA, Savage WJ, Tatham AS, Wan YF, Belton PS, Wellner N, D'ovidio R, Békés F, Halford NG (2003a) Sequence and properties of HMW subunit 1Bx20 from pasta wheat (*Triticum durum*) which is associated with poor end use properties. Theor Appl Genet 106:744–750

Shewry PR, Halford NG, Lafiandra D (2003b) The genetics of wheat gluten proteins. In: Hall JC, Dunlap JC, Friedman T (eds) Advances in genetics, vol. 49. Academic, pp 111–184

Shewry PR, Halford NG, Tatham AS, Popineau Y, Lafiandra D, Belton PS (2003c) The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. Adv Food Nutr Res 45:221–302

van Slageren MW (1994) Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (*Poaceae*). Wageningen Agriculture University Papers, 513 pp

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. Wiley, 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

Wan Y, Yan Z, Liu K, Zheng Y, D'Ovidio R, Shewry PR, Halford NG, Wang D (2005) Comparative analysis of the D genome-encoded high-molecular weight subunits of glutenin. Theor Appl Genet 111:1183–1190

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

Yan Z, Wan Y, Liu K, Zheng Y, Wang D (2002) Identification of a novel HMW glutenin subunit and comparison of its amino acid sequence with those of homologous subunits. Chin Sci Bull 47:220–225

Yen Y, Kimber G (1990) Genomic relationships of *Triticum searsii* to other S-genome diploid *Triticum* species. Genome 33:369–373

Zhang H, Reader SM, Liu X, Jia JZ, Gale MD, Devos KM (2001) Comparative genetic analysis of the *Aegilops longissima* and *Ae. sharonensis* genomes with common wheat. Theor Appl Genet 103:518–525

Zhang P, Friebe B, Gill BS (2002) Variation in the distribution of a genome-specific DNA sequence on chromosomes reveals evolutionary relations in the *Triticum* and *Aegilops* complex. Plant Syst Evol 235:169–179