

Novel x-Type High-Molecular-Weight Glutenin Genes From *Aegilops tauschii* and Their Implications on the Wheat Origin and Evolution Mechanism of *Glu-D1-1* Proteins

Yanzhen Zhang,^{*,1} Xiaohui Li,^{*,1} Aili Wang,^{*,†} Xueli An,^{*} Qian Zhang,^{*,†} Yuhe Pei,^{*} Liyan Gao,^{*} Wujun Ma,^{†,1} Rudi Appels[†] and Yueming Yan^{*,2}

^{*}Key Laboratory of Genetics and Biotechnology, College of Life Science, Capital Normal University, Beijing 100037, China and

[†]State Agriculture Biotechnology Centre, Western Australian Department of Agriculture and Food, Murdoch University, Perth, Western Australia 6150, Australia

Manuscript received August 1, 2007

Accepted for publication October 7, 2007

ABSTRACT

Two new x-type high-molecular-weight glutenin subunits with similar size to *IDx5*, designated *IDx5^{5'}* and *IDx5.1^{5'}* in *Aegilops tauschii*, were identified by SDS-PAGE, RP-HPLC, and MALDI-TOF-MS. The coding sequences were isolated by AS-PCR and the complete ORFs were obtained. Allele *IDx5^{5'}* consists of 2481 bp encoding a mature protein of 827 residues with deduced M_r of 85,782 Da whereas *IDx5.1^{5'}* comprises 2526 bp encoding 842 residues with M_r of 87,663 Da. The deduced M_r 's of both genes were consistent with those determined by MALDI-TOF-MS. Molecular structure analysis showed that the repeat motifs of *IDx5^{5'}* were correspondingly closer to the consensus compared to *IDx5.1^{5'}* and *IDx5* subunits. A total of 11 SNPs (3 in *IDx5^{5'}* and 8 in *IDx5.1^{5'}*) and two indels in *IDx5^{5'}* were identified, among which 8 SNPs were due to C-T or A-G transitions (an average of 73%). Expression of the cloned ORFs and N-terminal sequencing confirmed the authenticities of the two genes. Interestingly, several hybrid clones of *IDx5^{5'}* expressed a slightly smaller protein relative to the authentic subunit present in seed proteins; this was confirmed to result from a deletion of 180 bp through illegitimate recombination as well as an in-frame stop codon. Network analysis demonstrated that *IDx5^{5'}*, *IDx2*, *IDx1.6*, and *IDx2.2* represent a root within a network and correspond to the common ancestors of the other *Glu-D-1-1* alleles in an associated star-like phylogeny, suggesting that there were at least four independent origins of hexaploid wheat. In addition to unequal homologous recombination, duplication and deletion of large fragments occurring in *Glu-D-1-1* alleles were attributed to illegitimate recombination.

WHEAT is the most important grain crop in the world, with total annual yields of almost 600 million tonnes (SHEWRY *et al.* 2001). The seed storage proteins, mainly including gliadins and glutenins that initially deposit in ER-derived protein bodies, have the ability to form gluten polymers linked by disulfide bonds, and these are among the largest protein molecules in nature (WRIGLEY 1996). Gliadins are monomeric and confer dough tractility whereas glutenins are polymeric and consist of high-molecular-weight (HMW) and low-molecular-weight (LMW) subunits, which contribute to the visco-elasticity of dough (SHEWRY *et al.* 1992). Although the HMW glutenin subunits (HMW-GSs) compose only 8–10% of the total extractable flour protein, they play an important role in flour-processing quality due to network formation in dough by gluten polymerization, allowing wheat flour to be processed

into bread, pasta, noodles, and a range of other food products (SHEWRY *et al.* 1992; SHEWRY and HALFORD 2002; MA *et al.* 2005).

The HMW subunits are encoded by the *Glu-1* loci located on the long arms of chromosomes 1A, 1B, and 1D, and each locus consists of two closely linked genes, designated x- and y-types, with higher and lower molecular weights at two tightly linked loci, *Glu-1-1* and *Glu-1-2*, respectively (PAYNE 1987). In general, the number of cysteine residues is four in x-type and seven in y-type subunits; repetitive motifs with hexapeptides and nanopeptides are present in both, and tripeptides are present only in x-type subunits (SHEWRY *et al.* 1992). Consequently, three loci encoding up to six HMW-GSs are present in hexaploid bread or common wheat (*Triticum aestivum*, AABBDD). However, silencing of specific genes leads to variation in the number of subunits from three to five while allelic variation in the subunits encoded by active genes results in proteins with different electrophoretic mobilities (PAYNE 1987; SHEWRY *et al.* 2001).

It is generally accepted that *Aegilops tauschii* ($2n = 2x = 14$, DD) is the D-genome donor of hexaploid wheat,

Sequence data from this article have been deposited with the GenBank Data Libraries under accession nos. DQ681076 and DQ681077.

¹These authors contributed equally to this work.

²Corresponding author: Key Laboratory of Genetics and Biotechnology, College of Life Science, Capital Normal University, Beijing 100037, China. E-mail: yanyem@hotmail.com

which is presumed to have arisen from interspecific hybridization between *T. dicoccum* (AABB) and *Ae. tauschii*, with subsequent chromosome doubling, in south-western Asia 8000–12,000 years ago (MCFADDEN and SEARS 1946a,b; DVORAK *et al.* 1998; GILES and BROWN 2006). Recent investigations suggest that this polyploidization event occurred at least twice (LELLEY *et al.* 2000; CALDWELL *et al.* 2004; GILES and BROWN 2006), suggesting multiple origins of hexaploid wheats (DVORAK *et al.* 1998; ALLABY *et al.* 1999; HUANG *et al.* 2002; YAN *et al.* 2003a,b; GU *et al.* 2004).

Although a considerable amount of information is already available for the evolutionary origins of common wheat, some aspects need to be verified independently. For example, *Ae. tauschii* possesses extensive allelic variation in seed storage proteins (LAGUDAH and HALLORAN 1988; YAN *et al.* 2003a) that should provide useful evidence for insights into the evolution of hexaploid wheat, but only a few genes at the *Glu-D1-I* locus of *Ae. tauschii* have been characterized (WAN *et al.* 2005). Evidence from molecular analysis demonstrated that the HMW glutenin subunits from wheat and related species have highly conserved structures, consisting of a signal peptide (21 residues), an N-terminal domain (86–89 residues in x-type and 104 in y-type subunits), a C-terminal domain (42 residues), and a central repetitive domain (630–830 residues) that is mainly responsible for differences in molecular weight of the subunits (ANDERSON *et al.* 1989; WAN *et al.* 2002, 2005; YAN *et al.* 2004; SUN *et al.* 2006; ZHANG *et al.* 2006). The HMW glutenin genes, therefore, could derive from a common ancestor. The main molecular mechanisms for the evolution of glutenin genes at the *Glu-I* loci appear to be single nucleotide polymorphism (SNP) and insertion/deletion (indel) variations, duplications, and deletions of large repeats, probably resulting from events such as unequal crossover and slip-mismatching (ANDERSON and GREENE 1989; D'OVIDIO *et al.* 1996; ZHANG *et al.* 2006).

In this work, we identified and isolated two new x-type HMW subunit genes in *Ae. tauschii* accessions, and their molecular characteristics provided new evidence for multiple origins of hexaploid wheat. In particular, a large fragment deletion in the repetitive domain occurring in the *Escherichia coli* expression system suggested illegitimate recombination as a possible mechanism for duplication and deletion of large fragments at the *Glu-D1-I* locus.

MATERIALS AND METHODS

Plant materials: Two accessions of *Ae. tauschii* (Coss.) Schmal., TD81 and TD130, were kindly provided by GenBank (Braunschweig, Germany). The Yugoslav common wheat cultivar Dunav and Chinese Spring (CS) were used as standards for HMW-GS identification.

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis, reverse-phase high-performance liquid chromatography,

and matrix-assisted laser desorption/ionization time of flight mass spectrometry: HMW glutenin subunits were extracted and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) based on the protocol of YAN *et al.* (2003a). Reverse-phase high performance liquid chromatography (RP–HPLC) was based on the method of ANDREWS *et al.* (1994) with slight modifications. The column was de-aerated and equilibrated; the mixture (8 μ l) was used and eluted with a linear 50-min solvent gradient of 21–48% acetonitrile containing trifluoroacetic acid (0.06%) at the flow rate of 1.00 ml/min on Agilent 1100. The column was maintained at 50° and was returned to the initial solvent composition and reequilibrated for 15 min before the next analysis. Eluted protein components were detected at 210 nm. Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI–TOF–MS) for determining the molecular mass of HMW-GS was as described by ZHANG *et al.* (2006).

PCR amplification, cloning, and sequencing: Genomic DNA was extracted as described by SUN *et al.* (2004). The complete coding region sequences of the genes encoding x-type subunit genes at the *Glu-D1'* locus were amplified by primers designed from the sequences of subunits *IDx5* and *IDx2* (SUGIYAMA *et al.* 1985; ANDERSON *et al.* 1989), namely P1:5'-ATGGCTAAGCGGTTAGTC-3' and P2:5'-GCTGCAGA GAGTTCTATC-3' (synthesized by Sangong, Shanghai, China). The expected amplified products covering the start and stop codons were ~2500 bp according to previously characterized x-type HMW subunit genes (ANDERSON *et al.* 1989).

PCR was carried out using a Perkin-Elmer Cetus DNA thermal cycler (PE Applied Biosystems, Foster City, CA). A 50- μ l reaction mix was used, including 100 ng of DNA, 25 μ l of 2 \times GC buffer II (MgCl₂ plus), 0.4 mM of each dNTP, 0.5 μ M of each primer, 2.5 units LA Taq polymerase (TaKaRa). The PCR reaction was performed at 94° for 2 min, followed by 35 cycles at 94° for 45 sec, at 58° for 60 sec, and at 72° for 150 sec, and then concluded at 72° for 10 min. The amplified products of expected size were cloned into pGEM-T vector (Promega, Madison, WI) or pET30a (Novagen, expression step as follows). DNA sequencing of three clones was performed on an automatic DNA sequencer (TaKaRa Biotech, DaLian City, China).

Identification of SNPs and indels: The identification of SNP and indel variations in the cloned ORFs was based on multiple alignments of DNA and amino acid sequences and performed by ClustalW (THOMPSON *et al.* 1994).

Heterologous expression in *E. coli* and N-terminal microsequencing: The cloned HMW subunit genes were amplified to remove the signal peptides using the primers Pbd-1:5'-ACC CATATG GAA GGT GAG GCC TCT-3' and Pbd-2:5'-CTA GAA TTC CTA TCA CTG GCT GGC-3' (the added *Nde*I and *Eco*RI restriction sites are underlined).

PCR products were cloned into the bacterial expression vector pET30a (Novagen), and the hybrid vector (pET30a-*IDx5'*/*IDx5.I''*) was transformed into *E. coli* strain BL21 (DE3) pLsS. BL21 (DE3) pLsS cells containing the hybrid vector were grown in 2 \times YT medium (containing 50 μ g ml⁻¹ kanamycin and 34 μ g ml⁻¹ chloramphenicol) in a shaking incubator at 37° until the OD₆₀₀ reached 0.6. The expression of HMW subunit proteins was induced by adding 1–1.2 mM isopropyl β -D-thiogalactopyranoside for 4–6 hr. The expressed proteins were extracted from 1.2 ml of bacterial cells according to the method of WAN *et al.* (2002) for SDS–PAGE analysis.

N-terminal amino acid microsequencing of the expressed proteins was performed by PROCISE cLC 491 protein sequence system (Applied Biosystems) after transferring the proteins from the SDS gel to a polyvinylidene fluoride microporous membrane (Millipore, Bedford, MA) with a tank system (Bio-Rad mini trans-blot cell).

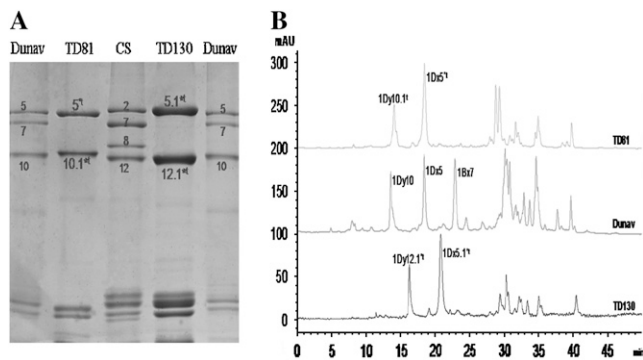


FIGURE 1.—Identification of HMW subunits from *Ae. tauschii* accessions TD81 and TD130. (A) SDS-PAGE. (B) RP-HPLC. Dunav and/or CS were used as references.

Network construction and phylogenetic analysis: Networks were constructed as described by ALLABY and BROWN (2001) using the nucleotide sequences of the signal peptide plus the N-terminal domain, which were considered phylogenetically informative (LI *et al.* 2004). A neighbor-joining tree was constructed by MEGA3 on the basis of the alignment of complete coding sequences using ClustalW. Bootstrap values were calculated as a percentage of 1000 trials.

RESULTS

Identification of novel HMW-GS in *Ae. tauschii*: SDS-PAGE analyses (Figure 1A) showed that *Ae. tauschii* accessions TD81 and TD130 possessed a pair of novel HMW glutenin subunits, designated $1Dx5^{5't}$ + $1Dy10.1^t$ and $1Dx5.1^{1't}$ + $1Dy12.1^{1't}$, respectively. The coding gene for the $1Dy10.1^t$ subunit was isolated and characterized in an earlier investigation (ZHANG *et al.* 2006). The subunits $1Dx5^{5't}$ and $1Dy10.1^t$ moved slightly faster and slower than the standard $1Dx5$ and $1Dy10$ subunits whereas $1Dx5.1^{1't}$ and $1Dy12.1^{1't}$ ran slightly slower and faster than the $1Dx5$ and $1Dy12$ subunits, respectively.

Figure 1B shows the RP-HPLC patterns of HMW glutenin subunits TD81 and TD130 as well as a control, Dunav (N, 7, 5 + 10); two subunits in each accession were separated. According to GIANIBELLI *et al.* (2002), the surface hydrophobicities of HMW-GS subunits were differ-

ent and the order was $1Dy < 1By < 1Dx < 1Bx < 1Ax$. Therefore, the x- and y-type subunits were readily identified as indicated in Figure 1B. The hydrophobicities of the seven HMW-GSs were $1Bx7 > 1Dx5.1^{1't} > 1Dx5^{5't}$ and $1Dx5 > 1Dy12.1^{1't} > 1Dy10.1^t > 1Dy10$.

The accurate molecular weights of the HMW glutenin subunits from TD81 and TD130 as well as the control, Dunav (N, 7, 5 + 10), were obtained by MALDI-TOF-MS (Figure 2). Both accessions possessed an x-type HMW-GS with molecular masses of 85,563.6 Da in TD81 and 87,480.2 Da in TD130, which correspond closely to those of the $1Dx5^{5't}$ and $1Dx5.1^{1't}$ subunits, respectively. Both subunits were smaller than the mature $1Dx5$ subunit (88,196.1 Da) from cultivar Dunav, with differences of 2632.5 and 715.9 Da, respectively. The molecular mass of the $1Dx5.1^{1't}$ subunit was different from that indicated for SDS-PAGE in Figure 1A. The anomalous electrophoretic behavior might result from fundamental conformational and structural differences between $1Dx5$ and $1Dx5.1^{1't}$, similar to those reported for $1Dy10$ and $1Dy12$ (GOLDSBROUGH *et al.* 1989), as well as gliadins (TATHAM and SHEWRY 1985).

Molecular characterization of $1Dx5^{5't}$ and $1Dx5.1^{1't}$ subunit genes: The designed degenerate oligonucleotide primers $P_1 + P_2$ were used to amplify the coding regions of the x-type HMW-GS genes in TD81 and TD130. Both accessions gave amplified products of 2500 bp, consistent with the size of the $1Dx5$ subunit gene from common wheat (ANDERSON *et al.* 1989). After purifying, cloning, and sequencing of the expected segments, two complete coding sequences with typical characters of HMW-GS genes were obtained. Both genes ended at a double stop codon (TGA, TAG) and introns were not present as for most other HMW-GS genes characterized so far. The $1Dx5^{5't}$ gene consisted of 2487 bp encoding a mature protein of 827 residues whereas $1Dx5.1^{1't}$ had 2532 bp encoding 842 amino acid residues.

Comparison of the deduced amino acid sequences (Figure 3) indicated that the $1Dx5^{5't}$ and $1Dx5.1^{1't}$ subunits shared a primary structure identical to those of $1Dx5$ and other subunits from common wheat. Three structural domains were present in all subunits: a nonrepetitive

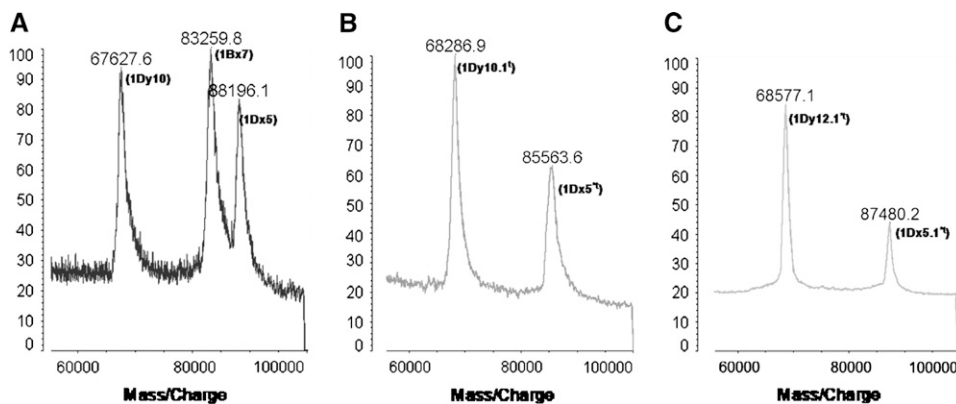


FIGURE 2.—The accurate molecular mass of HMW glutenin subunits from Dunav (A), TD81 (B), and TD130 (C) as determined by MALDI-TOF-MS. The subunits $1Dx5^{5't}$ and $1Dx5.1^{1't}$ with 85,563.6 and 87,480.2 Da, respectively, are shown.

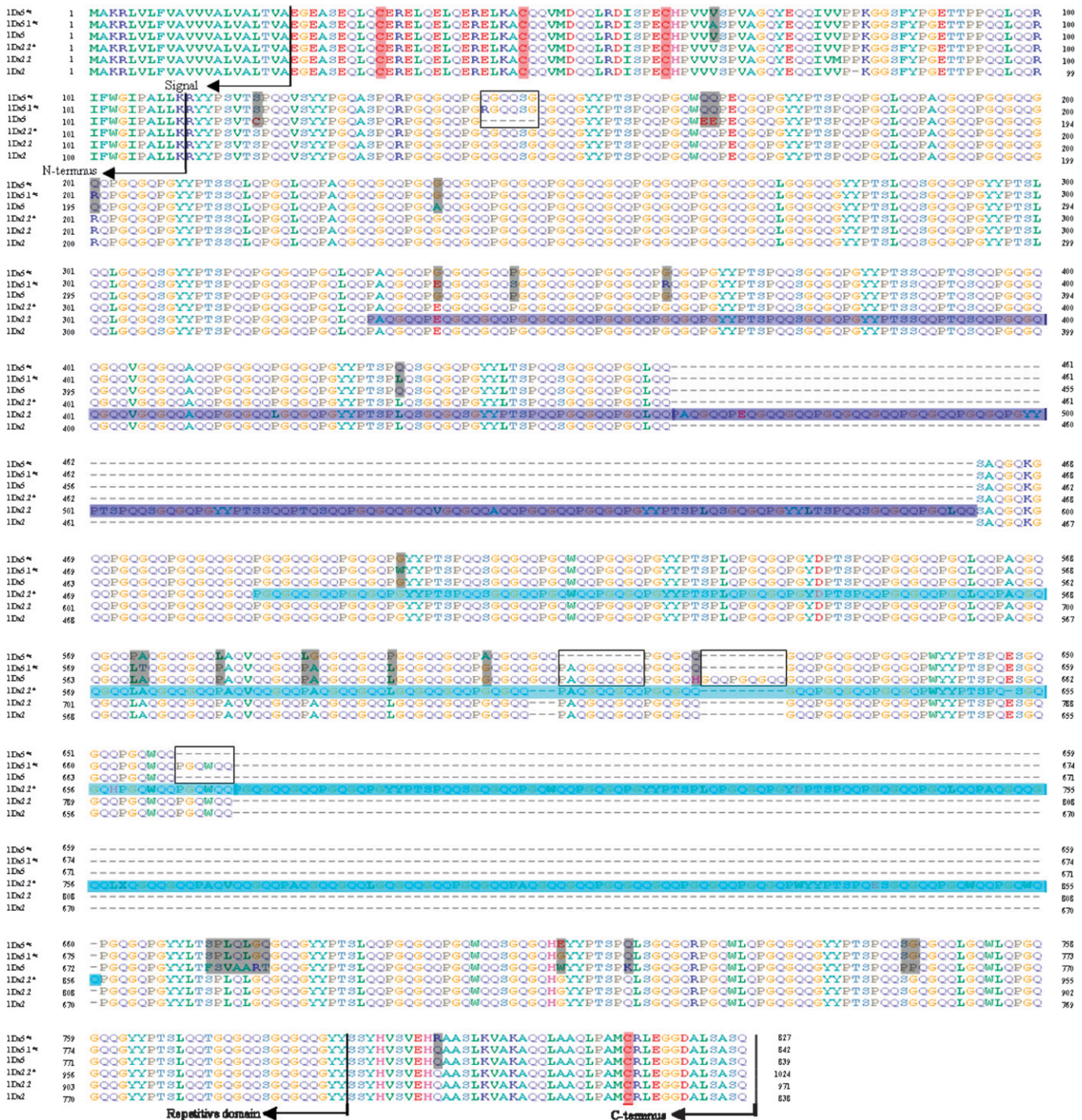


FIGURE 3.—Multiple sequence alignment of the deduced amino acid sequences of the *IDx5^t*, *IDx5.1^t*, *IDx5 IDx2.2*, *IDx2.2*, and *IDx2* subunits. Alignments were assembled by eye to demonstrate the repeat structure of the central domains. Cysteine residues are highlighted by red shading. Deletions are indicated by dashes. The gray shading and box indicate differences and indels among *IDx5^t*, *IDx5.1^t*, and *IDx5*, respectively. The regions repeated directly in *IDx2.2* and *IDx2.2* are dark blue and light blue, respectively.

N-terminal domain of 89 amino acid residues, repetitive domains of 109 and 112 residues, respectively, and a nonrepetitive C-terminal domain of 42 residues. Both subunits contained the expected four cysteine residues at conserved positions (Table 1): three in the N-terminal domain (at positions 31, 46, and 61) and one in the

C-terminal domain (at positions 815 and 830). Like their orthologous subunits, the repetitive domains consisted of tandem and interspersed repeats based on tripeptide (consensus GQQ), hexapeptide (PGQGQQ), and nonapeptide (GYYPTSLQQ) motifs, indicating that they were typical x-type subunits. The deduced molec-

ular masses were 85,782 Da ($IDx5^{*t}$) and 87,663 Da ($IDx5.I^{*t}$), which were consistent with those determined by MALDI-TOF-MS, and the differences were within the limits of experimental error in the mass range of HMW glutenin (HICKMAN *et al.* 1995). This suggested that both subunits lacked extensive post-translational modifications, such as glycosylation and phosphorylation, as did other HMW-GSs analyzed by different proteome approaches (COZZOLINO *et al.* 2001; CUNSOLO *et al.* 2004; ALBERGHINA *et al.* 2005). The two novel subunit genes were deposited in the GenBank under accession nos. DQ681076 ($IDx5^{*t}$) and DQ681077 ($IDx5.I^{*t}$).

$IDx5^{*t}$ and $IDx5.I^{*t}$ had 12 residues fewer and three residues more, respectively, than $IDx5$. There were 19 amino acid substitutions, a hexapeptide insertion, and a hexapeptide and two nonapeptide deletions in $IDx5^{*t}$ relative to $IDx5$, and 24 amino acid substitutions, two hexapeptide insertions, and a nonapeptide deletion in $IDx5.I^{*t}$ relative to $IDx5$ (Figure 3). The characteristics of the deduced mature proteins of $IDx5^{*t}$ and $IDx5.I^{*t}$ and previously reported HMW-GSs from *Ae. tauschii* and common wheat are summarized in Table 1. In particular, the frequencies of amino acid variations from the tripeptide, hexapeptide, and nonapeptide consensus at each of the 18 positions are indicated. Obviously, more variations occurred at positions 1 and 6, and at positions 4 and 2, among hexapeptides of x-type subunits, whereas the variant positions in y-type subunits were 1 and 4. There were more variations at positions 7, 8, and 9 of the nonapeptides in x-type subunits whereas the variant positions in y-type subunits were mostly 2, 7, and 5. The average numbers of variants per repeat unit and the mean frequency of variant units of subunit $IDx5^{*t}$ were relatively lower, indicating that its repeat units were closer to the consensus.

SNP and indel analyses: The coding sequences of the $IDx5^{*t}$ and $IDx5.I^{*t}$ genes were aligned with the other 11 *Glu-IDx* subunit genes and the SNPs and indels were identified (Table 2). A total of 11 SNPs were detected at different positions and the numbers of SNPs in $IDx5^{*t}$ and $IDx5.I^{*t}$ were 3 (1/827 bp) and 8 (1/316 bp), respectively. Of the 11 SNPs, 8 were due to A-G or C-T transitions (average 73%), approximating the results for *Glu-IDy10.I'* (ZHANG *et al.* 2006). In nine cases, SNPs at positions 45 and 711 in $IDx5.I^{*t}$ were found to produce amino acid residue substitutions (nonsynonymous SNP), namely 1718 T → C at position 573 (leucine → proline), 1745 C → T at position 582 (proline → leucine), 2384 A → G at position 795 (glutamine → arginine) in $IDx5^{*t}$, and 197 T → C at position 66 (valine → alanine), 425 A → G at position 142 (glutamine → arginine), 1033 C → T at position 345 (proline → serine), 1081 G → A at position 361 (glycine → arginine), 1501 G → T at position 501 (glycine → tryptophan), and 1720 G → A at position 574 (alanine → threonine) in $IDx5.I^{*t}$. In addition, 15 and 3 residue deletions, located at positions 1824–1838 and 1848–1850, respectively, were also found in $IDx5^{*t}$.

***E. coli* expression analysis:** The coding regions of the two new genes were expressed in *E. coli* using the pET-30α expression vector after removal of the sequences encoding signal peptide (Figure 4). As shown in Figure 5, A-N5^{*t} and B-N5.1^{*t} showed the expressed proteins of $IDx5^{*t}$ and $IDx5.I^{*t}$ ORFs in *E. coli* comigrating with the protein subunits present in TD81 and TD130 seeds, respectively. This confirmed that the cloned sequences were an accurate representation of each ORF and ruled out the possibility of glycosylation and other post-translational modifications of both subunits, consistent with the results of mass spectrometry.

Figure 5 shows that several other hybrid clones expressed a smaller protein (Figure 5A, S5^{*t}) relative to the authentic subunits present in seeds. The results of N-terminal sequencing, however, confirmed that the proteins of N5^{*t}, N5.1^{*t}, and S5^{*t} from *E. coli* expression and $IDx5^{*t}$ from seed extracts had the same amino acid residues. As shown in Figure 4A, the size of PCR products amplified by the Pbd-1 and Pbd-2 primers was exactly as expected, but after cloning with the expression vector pET-30a, some target fragments became shorter (Figure 4B, D1-1), including expression of a smaller peptide (Figure 5A, S5^{*t}). After sequencing the hybrid expressing plasmid pET30a- $IDx5^{*t}$, it became clear that a long deletion of 180 bp (from 1211 to 1032) and a key mutation (C → T) at position 1840 forming a stop codon TAG were present (Figure 6A). The derived molecular weight of the small protein was 65,220.33 Da, again consistent with the result of SDS-PAGE. The sequence analysis showed that there were three direct repeats (DR) in the repetitive domain as indicated by the boxes designated DRI, DRII, and DRIII in Figure 6. We believe that this was due to illegitimate recombination between repeats DRI and DRII. A proposed mechanism for the 180-bp deletion resulting from illegitimate recombination in $IDx5^{*t}$ is shown in Figure 6B. A similar phenomenon was reported in *Arabidopsis* (DEVOS *et al.* 2002). We also considered the possibility of illegitimate recombination between the other two pairs of repeats (DRI and DRIII or DRII and DRIII), but the frequencies would be much lower because of the much longer distances separating them. It is also possible that illegitimate recombination occurred in the $IDx2.2^*$ gene as shown in Figure 6C, resulting in a 558-bp deletion (from 2568 to 2011) and the generation of $IDx2$ in common wheat.

Similar phenomena were observed for LMW-GS gene recombination clones when using different *E. coli* strains (ALTENBACH 1998; MASCI *et al.* 1998). There, single deletions of 50–200 bp occurred within the repetitive domains of different recombinant clones. Differences among HMW glutenin subunits are mainly due to size and variation in the central repeat regions. Our investigations present direct evidence for recombinant clones and provide useful information for further studies on glutenin gene variation and its evolutionary direction.

TABLE 1
Characteristics of the deduced mature proteins of the *IDx5^{5t}* and *IDx5I^{5t}* subunits compared to previously reported HMW-GSs from *Ae. tauschii* and common wheat

Subunits	No. of amino acid residues and repeat unit variations																								MW (Da)					
	RD																													
	Tripeptide variants			Hexapeptide variants						Nanopeptide variants																No. of Cys residues				
	1	2	3	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	ND	RD	CD	Total								
ND	G	Q	Q	V/T ^a	P	G	Q	G	Q	Q	V/T ^a	G	Y	P	T	S	L	Q	Q	V/T ^a	V/T ^a	V/T ^b	CD	M						
<i>Dx5^{5t}</i>	89	1	1	1	2/20	20	10	0	13	0	16	46/70	2	0	1	2	0	0	12	2	3	14/19	67.41	42	0.91	3	0	1	4	85,782
<i>Dx5.1^{5t}</i>	89	1	2	1	3/21	23	13	1	14	2	16	52/72	1	0	1	2	0	0	15	3	3	16/19	75.82	42	1.03	3	0	1	4	87,663
<i>Dx5</i>	89	1	1	1	2/22	20	13	1	14	3	18	50/71	2	0	1	2	0	1	15	3	4	15/19	72.22	42	1.08	3	1	1	5	87,189
<i>Dx2</i>	88	1	2	1	3/20	21	11	0	14	2	15	49/72	1	0	1	2	0	0	15	3	3	15/19	70.32	42	0.97	3	0	1	4	87,023
<i>Dx2.1^t</i>	89	2	1	1	3/21	20	11	0	13	2	16	47/71	2	0	1	2	0	0	15	2	3	15/19	68.89	42	0.97	3	0	1	4	88,914
<i>Dx2.2^t</i>	89	1	2	1	3/23	20	15	0	15	1	20	54/81	1	0	1	2	0	0	18	4	3	18/21	70.59	42	0.98	3	0	1	4	103,048
<i>Dx2^t</i>	89	1	1	1	2/21	20	9	0	12	2	16	45/71	1	0	1	2	0	0	15	2	3	15/19	66.67	42	0.92	3	0	1	4	88,700
<i>Dx3^t</i>	89	1	1	1	2/21	20	9	1	13	2	16	45/71	1	0	1	2	1	0	15	2	3	15/19	66.67	42	0.96	3	0	1	4	88,829
<i>Dx5.2^t</i>	89	1	1	1	2/21	21	11	0	13	3	11	45/71	1	1	1	2	0	0	14	2	4	15/19	66.67	42	0.93	3	1	1	5	86,779
<i>Dy12</i>	104	—	—	—	—	18	6	3	13	6	5	30/50	1	10	2	1	8	0	10	0	2	15/19	65.22	42	1.21	5	1	1	7	68,711
<i>Dy10</i>	104	—	—	—	—	15	6	3	12	4	5	25/48	1	10	2	1	7	0	10	0	2	15/19	59.70	42	1.15	5	1	1	7	67,495

M, average number of variants per unit, including hexapeptides and nonapeptides. ND, N-terminal domain. RD, repetitive domain. CD, C-terminal domain. MW, molecular weights calculated from the derived amino acid sequences.

^a Variant motif number/total motif number.

^b Percentage of variant motifs relative to total motifs.

Network and neighbor-joining analysis: To investigate the phylogenetic relationships of x-type HMW-GSs encoded by the *Glu-D1-1* locus, we constructed a network with the signal peptides and N-terminal coding sequences and neighbor-joining trees with complete coding sequences (Figure 7). Thirteen x-type genes were analyzed, namely: *IDx5^{5t}* and *IDx5.I^{5t}* (this study, DQ681076 and DQ681077), *IDx5* (ANDERSON *et al.* 1989, X12928), *IDx2* (SUGIYAMA *et al.* 1985, X03346), *IDx2.I^t* (AF480486), *IDx2^t* (AF480485), *IDx1.5^t* (AY594355), *IDx1.6^t* (DQ857243), *IDx2.1* (AY517724), *IDx2.2* (WAN *et al.* 2005, AY159367), *IDx2.2^{*}* (D'OVIDIO *et al.* 1996; WAN *et al.* 2005, AJ893508), *IDx3^t* (DQ307383), and *IDx5.2^t* (DQ307384).

As shown in Figure 7A, the network placed the alleles *IDx5^{5t}*, *IDx2^t*, *IDx1.6^t*, and *IDx2.2^{*}* at a principal node. Among them, *IDx2.2^{*}* is the largest subunit detected so far in hexaploid wheat and related species. The other alleles radiated in a star-like phylogeny from the principal node. Allele *IDx5.I^{5t}*, as well as *IDx5*, was directly linked to the principal node and had no direct link with any other allele. Moreover, *IDx5* has an extra cysteine in

the repetitive region, which is believed to have a positive effect on quality properties. This implied that, although deriving from the ilk principal node, one relative independent mutation case might have occurred in the *IDx5* during the evolutionary process. Apparently, a point mutation at position 290 (C → G) in *IDx5* produced an amino acid residue substitution at position 97 (serine → cysteine) and therefore resulted in generating the extra cysteine. This extra cysteine remains in *IDx5.I^{5t}* and *IDx5.I^{5t}*, indicating that these two proteins can be used as new source for wheat quality improvement.

The *IDx2.I^t* and *IDx2* genes to the principal node still linked directly to *IDx1.5^t* and *IDx2.2*, respectively, implying that *IDx1.5^t* and *IDx2.2* might be more recent genes deriving from *IDx2.I^t* and *IDx2*, respectively. The network analysis demonstrated that *IDx5^{5t}*, *IDx2^t*, *IDx1.6^t*, and *IDx2.2^{*}* might represent ancestral sequences and that recombination with other star-like phylogenetic *Glu-IDx* alleles had not occurred. This suggested that there were at least four *Ae. tauschii* sources that contributed germplasm to the D genome of hexaploid wheats.

TABLE 2
Positions of SNPs and indels identified in the *IDx5^{5t}* and *IDx5I^{5t}* ORFs

HMW-GS gene	45	197	425	711	1033	1081	1501	1718	1720	1745	1824–1838	1848–1850	2384
<i>IDx5^{5t}</i>	C	T	A	G	C	G	G	C	G	T	-----	---	G
<i>IDx5.I^{5t}</i>	A	C	G	A	T	A	T	T	A	C	G C C A G G A C A A G G G C A	G C A	A
11 other <i>Glu-IDx</i> genes	C	T	A	G	C	G	G	T	G	C	G C C A G G A C A A G G G C A	G C A	A

SNPs are indicated by italics and indels by a horizontal dash.

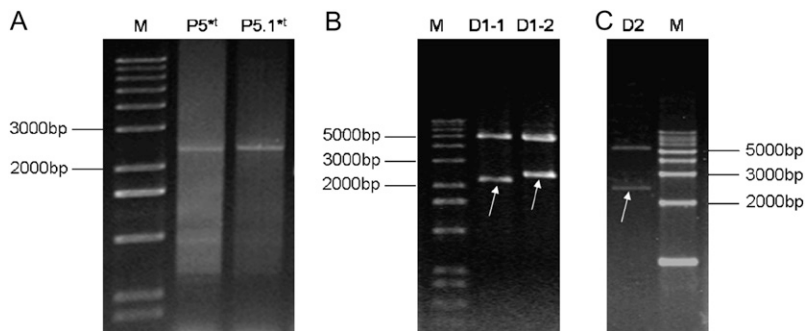


FIGURE 4.—Construction and identification of expression vectors pET30a-*IDx5[†]* and pET30a-*IDx5.1[†]*. (A) Amplified products using the primer Pbd-1+Pbd-2 from the hybrid pGEM-T vector containing the identified genes. (B) Digestion of the hybrid pET30a-*IDx5[†]* with *Nde*I and *Eco*RI restriction enzymes in tracks D1-1 and D1-2. (C) Digestion of the hybrid pET30a-*IDx5.1[†]* with *Nde*I and *Eco*RI restriction enzymes in track D3. M, 1-kb ladder markers. Inserted fragments are indicated by arrows.

The neighbor-joining tree (Figure 7B) revealed that *Glu-IDx* subunit genes were apparently divided into two clades; the greater group consisted of 12 genes at the top and the other group had only 2 genes at the bottom. Within the upper clade, a bootstrapping value of 100% gave high support for the split between *IDx1.6[†]* and the other alleles. The *IDx5[†]* and *IDx5.1[†]* were clustered into a separate subgroup. In the other clade, it was interesting that *IDx2.2^{*}* was clustered with *ISx2.5* from the S^b genome of *Ae. bicornis*, which was used as the outlier. This is consistent with the network analysis indicating that *IDx2.2^{*}* represents an ancestral gene although it was detected only in hexaploid wheats.

DISCUSSION

Network analysis for revealing relationships between glutenin alleles: Network analysis has been applied to study relationships between 5S rDNA sequences (ALLABY and BROWN 2001) and *Glu-Dy* (GILES and BROWN 2006) alleles. Such studies provide insights into the origin and evolution of hexaploid wheats and related species. Since network analysis makes few assumptions about the direction of evolution, the extent of sexual isolation, and the pattern of ancestry and descent, it describes relationships between sequences in a more realistic fashion than conventional tree building (ALLABY

and BROWN 2001). Previous research showed that it was valid to study glutenin phylogenetics by comparing partial glutenin sequences. Furthermore, the majority of partial alignments gave the same topology as the complete sequence alignments except for several of the partial repetitive regions (ALLABY *et al.* 1999). Because the software for construction of networks has not been developed and the difficulties in dealing with the complications involved in complete ORFs of 2.5–3.1 kb in length, it appears to be feasible and reasonable to perform network analysis with partial glutenin sequences. GILES and BROWN (2006) made a network analysis by examining a 284-bp sequence from the promoter region of the *Glu-Dy* locus and revealed aspects of the evolution and geographical origins of hexaploid wheat. In this work, we used the signal peptides and N-terminal domains of *Glu-D1-1* alleles to construct a network because the signal peptides and N-terminal sequences are relatively conserved among different HMW glutenin subunit genes. It is probably owing to their important roles, such as targeting the newly synthesized subunits into protein bodies and maintaining the high-order structure of the subunits. This suggests that these regions are subject to progressive changes during the evolution of HMW glutenin subunit genes and therefore are phylogenetically informative (LI *et al.* 2004).

Implications for the evolution of the *Glu-D1-1* locus and the origin of cultivated wheats: As shown in this work, the *IDx5[†]*, *IDx2^{*}*, and *IDx1.6[†]* genes from diploid *Ae. tauschii* and *IDx2.2^{*}* from common wheat were located at the shared principal node in the phylogenetic network, representing a root within a network and corresponding to the common ancestor of the genes in the associated star-like phylogeny (ALLABY and BROWN 2001). These nodal *Glu-D1-1* genes appear to be of considerable antiquity, suggesting that there were at least four independent origins of hexaploid wheat. Two very large HMW subunits, *IDx2.2^{*}* and *IDx2.2*, were detected only in hexaploid wheat, and similar large Dx subunits have not been found in *Ae. tauschii* accessions (LAGUDAH and HALLORAN 1988; YAN *et al.* 2003a). Molecular characterization showed that the large size of the *IDx2.2^{*}* gene is due to a single duplication of 561 bp within the repetitive domain, possibly arising from unequal crossing over during meiosis (D'OVIDIO

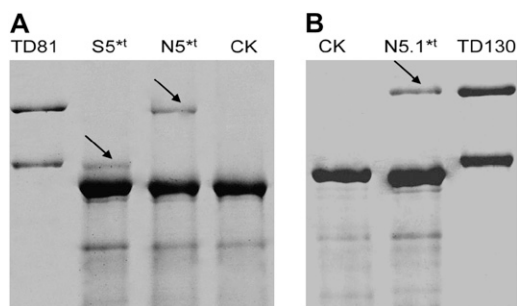


FIGURE 5.—Expression of the cloned *IDx5[†]* and *IDx5.1[†]* gene ORFs in *E. coli*. (A) Expressed proteins of different pET30a-*IDx5[†]* plasmid clones. The expressed normal and smaller proteins are indicated in tracks N5[†] and S5[†], respectively, by arrows. (B) Expressed protein of pET30a-*IDx5.1[†]* plasmid clones (indicated by the arrow).

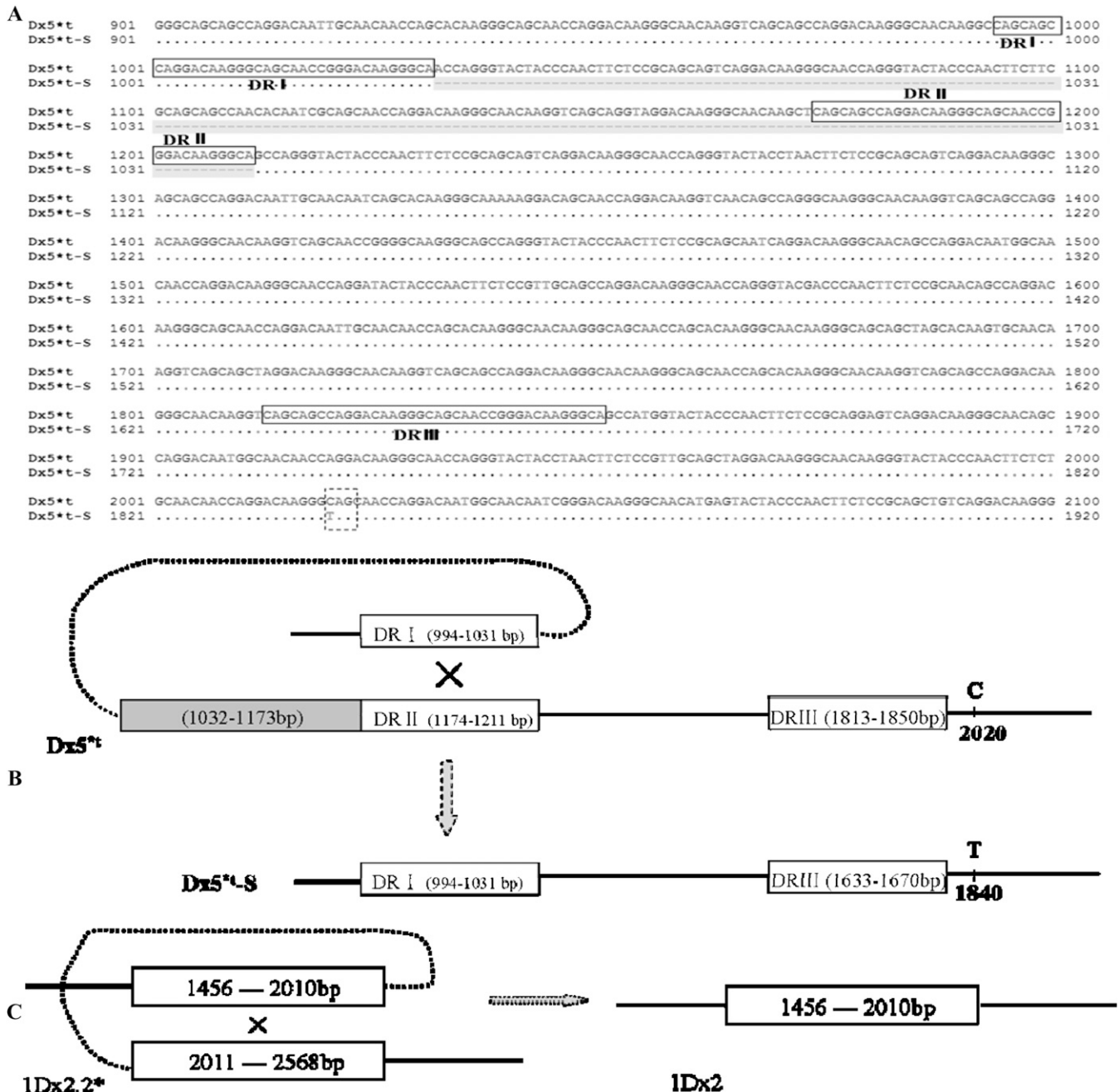


FIGURE 6.—Comparison of the coding sequences of *IDx5⁺t* and *IDx5⁺t-S* from the deletion hybrid $p^{ET30a-IDx5^+t}$ (A), the possible mechanisms generating *IDx5⁺t-S* (B) and *1Dx2* from *1Dx2.2** through illegitimate recombination (C). (A) The same sequences and deletions are indicated by a dot and by light gray shaded dashes, respectively. The direct repeats are boxed and designated as DR I, DR II, and DR III, respectively. The key mutation resulting in a stop codon is boxed by dashed line. (B) Illegitimate recombination between DR I and DR II (see A) resulting in the deletion of the domain (the shaded box) and repeat II. (C) The illegitimate recombination that probably occurred in *IDx2.2** that resulted in the *IDx2* gene. The dotted lines in B and C indicate the illegitimate recombination event.

et al. 1996). According to our network analysis, it is most likely that the *IDx2.2** gene could have been present in *Ae. tauschii* and then transferred by rare hybridization into common wheat as suggested by D'OVIDIO *et al.* (1996). The absence of large *IDx2.2**-like Dx genes in *Ae. tauschii* is probably due to the limited number of accessions surveyed so far or could have been lost from the population.

In terms of evolutionary analysis, the *Glu-Dx* subunit alleles arose mainly from four principal ancestor genes. We propose that *IDx2.2**, deriving from a more primitive gene, was present in *Ae. tauschii* and was transferred by rare outcrossing into common wheats. A mutational event, probably by illegitimate recombination, resulted in loss of a large DNA fragment and generation of the *IDx2* gene as indicated in Figure 6C (the corresponding

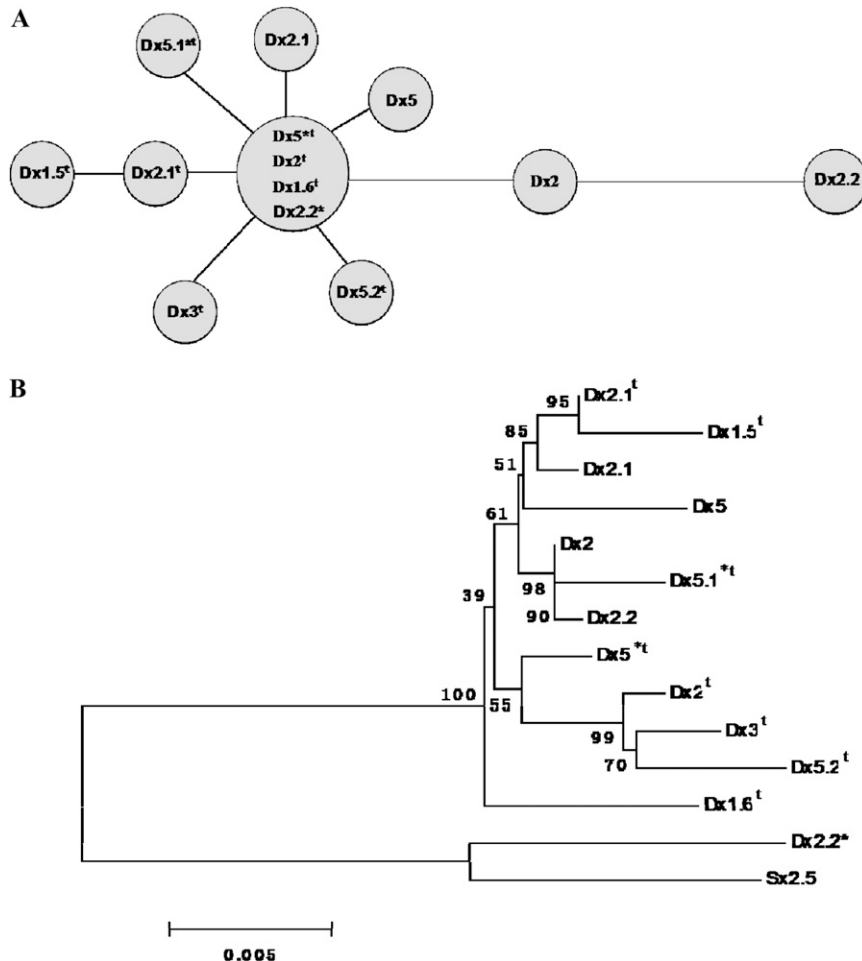


FIGURE 7.—Phylogenetic network based on nucleic acid sequence variations of signal and N-terminal domains (A) and the neighbor-joining tree based on multi-alignment of complete coding sequences of 13 *Glu-D1-1* alleles from *Ae. tauschii* and common wheat and of *1Sx2.5* from the S^b genome of *Ae. bicornis* (B).

amino acid sequences of the duplicated DNA sequences are light blue in Figure 3). In the recent evolutionary timescale of common wheats, one large fragment duplication in the repetitive domain of *IDx2* (shaded dark blue in Figure 3) presumably resulted from unequal crossing over as well as illegitimate recombination to generate the *IDx2.2* allele.

The hypotheses of multiple origins of the D genome in hexaploid wheat are supported by the studies of LAGUDAH and HALLORAN (1988) and YAN *et al.* (2003b) as well as by evidence from the A and B genomes (ALLABY *et al.* 1999; GU *et al.* 2004). More recently, GILES and BROWN (2006) investigated the evolution and geographical origins of hexaploid wheat by examining *Glu-Dy* alleles. The existence of two shared alleles suggested that there were at least two independent origins of hexaploid wheat. Other investigations demonstrated that the A and B genomes might have introgressed into hexaploid wheat at different times and rates (BLATTER *et al.* 2004; ZHANG *et al.* 2006) and that, on the basis of the fact that there are two distinct types of alleles at the *Glu-A1* locus, hexaploid wheat might have more than one tetraploid ancestor (GU *et al.* 2004). Furthermore, surveys of sequence tagged sites (TALBERT *et al.* 1998), RFLP markers (DVORAK *et al.* 1998), and microsatellites (LELLEY *et al.* 2000) and analyses of the

Xwy838 and *Gss* loci (CALDWELL *et al.* 2004) also indicated multiple origins for hexaploid wheat.

A possible mechanism for duplication and deletion in *Glu-1* genes through illegitimate recombination: Modern hexaploid wheat is an allohexaploid species with genomes A, B, and D and an extremely large and complex genome (16,000 Mb). Although hexaploid wheat is the product of the hybridization between tetraploid wheat (AABB) and diploid goat grass (DD) that happened ~10,000 years ago, its genomes have been evolving dynamically and rapidly (HUANG *et al.* 2002; ANDERSON *et al.* 2003; WICKER *et al.* 2003; GU *et al.* 2004). The evolution of these genomes was characterized as the balance between expansion and reduction through transposon and retrotransposon insertions, duplications, and deletions of large fragments. These rapid changes happened in overall gene organization not only in intergenic regions but also in coding regions as shown in this work on glutenin genes. In the *Glu-1* loci, the extensive allelic variations are mainly the result of SNPs and indels, probably resulting from unequal crossing over, slip-mismatching, and point mutations. As shown for *IDx5*1* and *IDx5.1*1* genes, many SNPs resulted from the C-T transitions because C is readily methylated and deaminated (RAZIN and RIGGS 1980).

Illegitimate recombination was considered a major factor in the evolution of the wheat (WICKER *et al.* 2003) and Arabidopsis (DEVOS *et al.* 2002) genomes. This process generates duplications and deletions and therefore, like unequal homologous recombination, results in genome expansion and contraction. Because illegitimate recombination requires only a few base pairs of sequence identity (WICKER *et al.* 2003), it is likely to occur within *Glu-D1-1* alleles because of a few long direct repeats generally present in the repetitive domain of x-type HMW genes as shown in *IDx5'* and *IDx2.2'* (Figures 3 and 6). Although fragment deletions within the repetitive domains of glutenin genes in recombinant clones were found (ALTENBACH 1998; MASCI *et al.* 1998), their molecular origins are still unknown. Our results provide direct evidence for illegitimate recombination in *IDx5'* within the *E. coli* heterologous expression system and have led to a large fragment deletion and generation of a small protein. This suggests that duplications and deletions of large fragments within glutenin genes could be produced through illegitimate recombination and indicates a possible molecular mechanism for generating novel allelic variation at *Glu-1* loci.

We are grateful to Robert McIntosh for constructive suggestions and for reviewing the manuscript. This research was financially supported by grants from the National Natural Science Foundation of China (30571154, 30771334) and the Chinese Ministry of Science and Technology (2002CB111300, 2006AA10Z186).

LITERATURE CITED

- ALBERGHINA, G., R. COZZOLINO, S. FISICHELLA, D. GAROZZO and A. SAVARINO, 2005 Proteomics of gluten: mapping of the 1Bx7 glutenin subunit in Chinese Spring cultivar by matrix-assisted laser desorption/ionization. *Rapid Commun. Mass Spectrom.* **19**: 2069–2074.
- ALLABY, R. G., and T. A. BROWN, 2001 Network analysis provides insights into evolution of 5S rDNA arrays in *Triticum* and *Aegilops*. *Genetics* **157**: 1331–1341.
- ALLABY, R. G., M. BANERJEE and T. A. BROWN, 1999 Evolution of the high molecular weight glutenin loci of the A, B, D and G genomes of wheat. *Genome* **42**: 296–307.
- ALTENBACH, S. B., 1998 Quantification of individual low-molecular-weight glutenin subunit transcripts in developing wheat grains by competitive RT-PCR. *Theor. Appl. Genet.* **97**: 413–421.
- ANDERSON, O. D., and F. C. GREENE, 1989 The characterization and comparative analysis of HMW glutenin genes from genomes A and B of hexaploid bread wheat. *Theor. Appl. Genet.* **77**: 689–700.
- ANDERSON, O. D., F. C. GREENE, R. E. YIP, N. G. HALFORD, P. R. SHEWRY *et al.*, 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.
- ANDERSON, O. D., C. RAUSCH, O. MOULLET and W. S. LAGUDAH, 2003 The wheat D-genome HMW-glutenin locus: BAC sequencing, gene distribution, and retrotransposon clusters. *Funct. Integr. Genomics* **3**: 56–68.
- ANDREWS, J. L., R. L. HAY, J. H. SKERRITT and H. SUTTON, 1994 HPLC and immunoassay-based glutenin subunit analysis: screening for dough properties in wheats grown under different environmental conditions. *J. Cereal Sci.* **20**: 203–215.
- BLATTER, R. H. E., S. JACOMET and A. SCHLUMBAUM, 2004 About the origin of European spelt (*Triticum spelta* L.): allelic differentiation of the HMW glutenin *BI-1* and *AI-2* subunit genes. *Theor. Appl. Genet.* **108**: 360–367.
- CALDWELL, K. S., J. DVORAK, E. S. LAGUDAH, E. AKHUNOV, M. C. LUO *et al.*, 2004 Sequence polymorphism in polyploidy wheat and their D-genome ancestor. *Genetics* **167**: 941–947.
- COZZOLINO, R., G. S. DI, S. FISICHELLA, D. GAROZZO, D. LAFIANDRA *et al.*, 2001 Matrix-assisted laser desorption/ionization mass spectrometric peptide mapping of high molecular weight glutenin subunits 1Bx7 and 1Dy10 in Cheyenne cultivar. *Rapid Commun. Mass Spectrom.* **15**: 778–787.
- CUNSOLO, V., S. FOTI, R. SALETTI, S. GILBERT, A. S. TATHAM *et al.*, 2004 Structural studies of the allelic wheat glutenin subunits 1Bx7 and 1Bx20 by matrix-assisted laser desorption/ionization mass spectrometry and high-performance liquid chromatography/electrospray ionization mass spectrometry. *J. Mass Spectrom.* **39**: 66–78.
- DEVOS, K. M., J. K. M. BROWN and J. L. BENNETZEN, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12**: 1075–1079.
- D'OVIDIO, R., D. LAFIANDRA and E. PORCEDDU, 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.
- DVORAK, J., M. C. LUO, Z. L. YANG and H. B. ZHANG, 1998 The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. *Theor. Appl. Genet.* **97**: 657–670.
- GIANIBELLI, M. C., M. ECHAIDE, O. R. LARROQUE, J. M. CARILLO and J. DUBCOVSKY, 2002 Biochemical and molecular characterization of *Glu-1* loci in Argentinian wheat cultivars. *Euphytica* **128**: 61–73.
- GILES, R. J., and T. A. BROWN, 2006 GluDy allele variations in *Aegilops tauschii* and *Triticum aestivum*: implications for the origins of hexaploid wheat. *Theor. Appl. Genet.* **112**: 1563–1572.
- GOLDSBROUGH, A. P., N. J. BULLEID, R. B. FREEDMAN and R. B. FLAVELL, 1989 Conformational differences between two wheat (*Triticum aestivum*) 'high-molecular-weight' glutenin subunits are due to a short region containing six amino acid differences. *Biochem. J.* **263**: 837–842.
- GU, Y. Q., C. D. DEVIN, X. Y. KONG and O. D. ANDERSON, 2004 Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four *Triticeae* genomes. *Plant Physiol.* **135**: 459–470.
- HICKMAN, D. R., P. ROEPSTORFF, P. R. SHEWRY and A. S. TATHAM, 1995 Molecular weights of high molecular weight subunits of glutenin determined by mass spectrometry. *J. Cereal Sci.* **22**: 99–103.
- HUANG, S., A. SIRIKHACHORNKIT, X. J. SU, J. FARIS, B. GILL *et al.*, 2002 Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and evolutionary history of polyploid wheat. *Proc. Natl. Acad. Sci. USA* **99**: 8133–8138.
- LAGUDAH, E. S., and G. M. HALLORAN, 1988 Phylogenetic relationships of *Triticum tauschii* the D genome donor to hexaploid wheat. 1. Variation in HMW subunits of glutenin and gliadin. *Theor. Appl. Genet.* **75**: 592–598.
- LELLEY, T., M. STACHEL, H. GRAUSGRUBER and J. VOLLMANN, 2000 Analysis of relationships between *Ae. tauschii* and the D genome of wheat utilizing microsatellite. *Genome* **43**: 661–668.
- LI, W., Y. WAN, Z. LIU, K. LIU, B. LI *et al.*, 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.
- MA, W., R. APPELS, F. BEKES, O. LARROQUE, M. K. MORELL *et al.*, 2005 Genetic characterisation of dough rheological properties in a wheat doubled haploid population: additive genetic effects and epistatic interactions. *Theor. Appl. Genet.* **111**: 410–422.
- MASCI, S., R. D'OVIDIO, D. LAFIANDRA and D. D. KASARDA, 1998 Characterization of a low-molecular-weight glutenin subunit gene from bread wheat and the corresponding protein that represents a major subunit of the glutenin polymers. *Plant Physiol.* **118**: 1147–1158.
- McFADDEN, E. S., and E. R. SEARS, 1946a The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37**: 81–89.
- McFADDEN, E. S., and E. R. SEARS, 1946b The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37**: 107–116.

- PAYNE, P. I., 1987 Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Annu. Rev. Plant Physiol.* **38**: 141–153.
- RAZIN, A., and A. D. RIGGS, 1980 DNA methylation and gene function. *Science* **210**: 604–610.
- SHEWRY, P. R., and N. G. HALFORD, 2002 Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.* **53**: 947–958.
- SHEWRY, P. R., N. G. HALFORD and A. S. TATHAM, 1992 The high molecular weight subunits of wheat glutenin. *J. Cereal Sci.* **15**: 105–120.
- SHEWRY, P. R., Y. POPINEAU, D. LAFIANDRA and P. BELTON, 2001 Wheat glutenin subunits and dough elasticity findings of the EUROWHEAT project. *Trends Food Sci. Technol.* **11**: 433–441.
- SUGIYAMA, T., A. RAFALSKI, D. PETERSON and D. SOL, 1985 A wheat HMW glutenin subunit gene reveals a highly repeated structure. *Nucleic Acid Res.* **13**: 8729–8737.
- SUN, M. M., Y. M. YAN, Y. JIANG, Y. H. XIAO, Y. K. HU *et al.*, 2004 Molecular cloning and comparative analysis of a y-type inactive HMW glutenin subunit gene from cultivated emmer wheat (*Triticum dicoccum* L.). *Hereditas* **141**: 46–54.
- SUN, X., S. HU, X. LIU, W. QIAN, S. HAO *et al.*, 2006 Characterization of the HMW glutenin subunits from *Aegilops searsii* and identification of a novel variant HMW glutenin subunit. *Theor. Appl. Genet.* **113**: 631–641.
- TALBERT, L. E., L. Y. SMITH and N. K. BLAKE, 1998 More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome* **41**: 402–407.
- TATHAM, A. S., and P. R. SHEWRY, 1985 The conformation of wheat gluten proteins. The secondary structures and thermal stabilities of alpha-, beta-, gamma- and omega-gliadins. *J. Cereal Sci.* **3**: 103–113.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 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.
- WAN, Y., D. WANG, P. R. SHEWRY and N. G. HALFORD, 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., Z. YAN, K. LIU, Y. ZHENG, R. D'OVIDIO *et al.*, 2005 Comparative analysis of D genome-encoded high-molecular weight subunits of glutenin. *Theor. Appl. Genet.* **111**: 1183–1190.
- WICKER, T., N. YAHIAOUI, R. GUYOT, E. SCHLAGENHAUF, Z. D. LIU *et al.*, 2003 Rapid genome divergence at orthologous low molecular weight glutenin loci of A and A^m genomes of wheat. *Plant Cell* **15**: 1186–1197.
- WRIGLEY, C. W., 1996 Giant proteins with flour power. *Nature* **381**: 738–739.
- YAN, Y., S. L. K. HSAM, J. Z. YU, Y. JIANG and F. J. ZELLER, 2003a Allelic variation of the HMW glutenin subunits in *Aegilops tauschii* accessions detected by sodium dodecyl sulphate (SDS-PAGE), acid polyacrylamide gel (A-PAGE) and capillary electrophoresis. *Euphytica* **130**: 377–385.
- YAN, Y., S. L. K. HSAM, J. YU, Y. JIANG, I. OHTSUKA *et al.*, 2003b HMW and LMW glutenin alleles among putative tetraploid and hexaploid *T. spelta* progenitors. *Theor. Appl. Genet.* **107**: 1321–1330.
- YAN, Y., J. ZHENG, Y. XIAO, J. YU, Y. HU *et al.*, 2004 Identification and molecular characterization of a novel y-type *Glu-D'1* glutenin gene of *Aegilops tauschii*. *Theor. Appl. Genet.* **108**: 1349–1358.
- ZHANG, Y. Z., Q. Y. LI, Y. M. YAN, J. G. ZHENG, X. L. AN *et al.*, 2006 Molecular characterization and phylogenetic analysis a novel glutenin gene (*Dy10.1'*) from *Aegilops tauschii*. *Genome* **49**: 735–745.

Communicating editor: J. A. BIRCHLER