

Types and Rates of Sequence Evolution at the High-Molecular-Weight Glutenin Locus in Hexaploid Wheat and Its Ancestral Genomes

Yong Qiang Gu,^{*,1} Jérôme Salse,^{†,‡} Devin Coleman-Derr,^{*} Adeline Dupin,[†] Curt Crossman,^{*} Gerard R. Lazo,^{*} Naxin Huo,^{*} Harry Belcram,[†] Catherine Ravel,[‡] Gilles Charmet,[‡] Mathieu Charles,[†] Olin D. Anderson^{*} and Boulos Chalhou[†]

^{*}United States Department of Agriculture-Agricultural Research Service, Western Regional Research Center, Albany, California 94710,

[†]Laboratory of Genome Organization, Unité de Recherches en Génomique Végétale (URGV-INRA), 91057 Evry Cedex, France and

[‡]UMR INRA-UBP ASP Amélioration et Santé des Plantes, 63039 Clermont Ferrand, France

Manuscript received May 15, 2006

Accepted for publication August 29, 2006

ABSTRACT

The *Glu-1* locus, encoding the high-molecular-weight glutenin protein subunits, controls bread-making quality in hexaploid wheat (*Triticum aestivum*) and represents a recently evolved region unique to Triticeae genomes. To understand the molecular evolution of this locus region, three orthologous *Glu-1* regions from the three subgenomes of a single hexaploid wheat species were sequenced, totaling 729 kb of sequence. Comparing each *Glu-1* region with its corresponding homologous region from the D genome of diploid wheat, *Aegilops tauschii*, and the A and B genomes of tetraploid wheat, *Triticum turgidum*, revealed that, in addition to the conservation of microsynteny in the genic regions, sequences in the intergenic regions, composed of blocks of nested retroelements, are also generally conserved, although a few nonshared retroelements that differentiate the homologous *Glu-1* regions were detected in each pair of the A and D genomes. Analysis of the indel frequency and the rate of nucleotide substitution, which represent the most frequent types of sequence changes in the *Glu-1* regions, demonstrated that the two A genomes are significantly more divergent than the two B genomes, further supporting the hypothesis that hexaploid wheat may have more than one tetraploid ancestor.

POLYPLOIDIZATION, an evolutionary process resulting in more than one genome per cell, has played a significant role in the evolutionary history of plants, particularly in agriculturally important crops (MASTERSON 1994; SOLTIS and SOLTIS 1999; WENDEL 2000). Bread wheat is an allohexaploid species (*Triticum aestivum* L. $2n = 6x = 42$), consisting of three sets of highly related genomes (A, B, and D). Hexaploid wheat originated from two independent polyploidization events. The first event involved the hybridization of two diploid progenitors: an ancestor of *Triticum urartu* ($2n = 2x = 14$, genome AA) and an unconfirmed species (BB genome) related to *Aegilops speltoides* ($2n = 2x = 14$, genome SS), which resulted in cultivated allotetraploid emmer wheat (*T. turgidum* ssp. *dicoccum*, $2n = 4x = 28$, genomes AABB) (DVORAK *et al.* 1992; BLAKE *et al.* 1999). In the second event, which occurred 8000–10,000 years ago, an ancestor of the diploid *Aegilops tauschii* (DD genome) hybridized with the allotetraploid to form a hexaploid wheat ($2n = 6x = 42$) (FELDMAN *et al.* 1995).

Because of its relatively recent speciation, wheat represents an excellent system for studying evolutionary events that occur in genomes shortly after polyploidization. Different types of genetic and epigenetic changes, including DNA removal, changes in gene expression, reactivation of transposable elements, and functional diversification of duplicate genes, are all known to be significant to the evolutionary process in polyploid species (OZKAN *et al.* 2001; SHAKE *et al.* 2001; COMAI *et al.* 2002; HE *et al.* 2003; KASHKUSH *et al.* 2003; BLANC and WOLFE 2004). However, the particular molecular mechanisms underlying the rapid genome evolution observed in polyploid genomes are not well understood. Sequence comparisons of polyploid genomes with their ancestral genomes may represent the strategy of choice to identify sequence changes caused by recent genome evolution. Recently, this strategy has been successfully used to elucidate the molecular basis of the polyploidy-related deletion of the *Hardness* locus from the tetraploid wheat (*T. turgidum*) (CHANTRET *et al.* 2005)

One of the great challenges for wheat genome research is the size of Triticeae genomes. Compared with model plant species, such as *Arabidopsis* (130 Mb) and rice (430 Mb), the bread wheat genome (~16,000 Mb) is extremely large. We now know that although the sudden increase in chromosome numbers caused by polyploidization contributed to the overall increase in the wheat

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. DQ537335, DQ537336, and DQ537337.

¹Corresponding author: USDA-ARS, Western Regional Research Center, 800 Buchanan St., Albany, CA 94710. E-mail: ygu@pw.usda.gov

genome size, the replication and insertion of repetitive DNA, particularly long terminal repeat (LTR) retrotransposons (SANMIGUEL *et al.* 1996), is another major cause for genome expansion (BENNETZEN *et al.* 2005). In wheat, repetitive DNA accounts for ~90% of the genome, of which retrotransposons constitute 60–80% (SANMIGUEL *et al.* 2002; WICKER *et al.* 2003; GU *et al.* 2004).

In addition to being responsible for genome size variation, transposable elements often stimulate other types of genomic rearrangements, including unequal homologous and illegitimate recombination to remove nuclear DNA (DEVOS *et al.* 2002; MA *et al.* 2004; CHANTRET *et al.* 2005). Their transpositions can cause gene inactivation or changing expression of adjacent genes (KASHKUSH *et al.* 2003; GU *et al.* 2004). Repetitive DNA elements also make useful markers for revealing recent genomic changes because they are abundant and believed to be neutral under natural selection. Repetitive DNA, including retrotransposons, has a relative short turnover period (MA *et al.* 2004) so that colinear retrotransposons are usually not found in distantly related genomes. Our previous studies of comparative analyses of orthologous high-molecular-weight (HMW) glutenin regions indicated that only a few remnants of elements are colinear in the closely related wheat A, B, and D genomes that diverged from each other within the last 4–5 million years (MY) (HUANG *et al.* 2002; GU *et al.* 2004; KONG *et al.* 2004). Because of this, the intergenic regions are largely not conserved between the A, B, and D genomes of wheat, although gene colinearity is retained (GU *et al.* 2004). Furthermore, several recent studies suggested that retrotransposable elements may be one of the major causes for intraspecific sequence variations in the intergenic regions (BRUNNER *et al.* 2005; SCHERRER *et al.* 2005). A comparison of homologous regions from different inbred lines in maize indicated that >59% of the compared sequence is noncolinear largely due to the insertion of novel or allele-specific LTR retrotransposons (BRUNNER *et al.* 2005).

In addition to retroelements, other genomic rearrangements, particularly sequence insertions and deletions (indels), play significant roles in genome evolution (PETROV *et al.* 2000; GREGORY 2004). In rice, an estimated minimum of 119 Mb of sequence has been removed from the genome in the last 5 MY (MA and BENNETZEN 2004). A comparison of the *Hardness* locus in diploid and polyploid wheat species has indicated that multiple genomic deletions occurred independently in different genomes and that indels are one of the primary evolutionary mechanisms involved in reconstructing this domestication gene region (CHANTRET *et al.* 2005). Gene disruption caused by deletions and large inversions has resulted in various haplotypes in the wheat A genome (ISIDORE *et al.* 2005). Despite rapid DNA rearrangements and intraspecific violation of genetic colinearity observed in

recent studies (BRUNNER *et al.* 2005; CHANTRET *et al.* 2005; ISIDORE *et al.* 2005; SCHERRER *et al.* 2005), it remains to be shown whether such rapid sequence diversification has occurred throughout the whole genome or only in certain genetic loci. Future sequence comparisons of regions from homologous genomes, such as wheat genomes at different ploidy levels, will help answer this question.

In wheat, the *Glu-1* locus encodes HMW glutenin protein subunits, which are the major determinants of bread-making quality for wheat flour, making *Glu-1* one of the most important genetic loci in wheat and frequently the target of genetic engineering efforts for the improvement of grain quality (BLECHL and ANDERSON 1996; ROOKE *et al.* 1999). The wheat HMW glutenin locus and the orthologous loci of barley and rye are unique to Triticeae species (SHEWRY and TATHAM 1990), which suggests that these loci evolved relatively recently within the Triticeae tribe. Studies focusing on these specific orthologous HMW glutenin regions from diploid (ANDERSON *et al.* 2003) and tetraploid wheat genomes (KONG *et al.* 2004), as well as from barley (GU *et al.* 2003), have provided the first view of genome evolution in three homeologous regions from the wheat A, B, and D genomes (GU *et al.* 2004). In this article, we report sequencing the three HMW glutenin locus regions located on the long arms of homeologous group 1 chromosomes from a single hexaploid wheat species, which has allowed us to compare the homologous *Glu-1* regions from the hexaploid wheat with its diploid and tetraploid ancestors. The sequence conservation and divergence detected could shed light on the evolutionary history of the wheat genomes.

MATERIALS AND METHODS

Isolation and sequencing of hexaploid wheat BACs: Hexaploid wheat BAC clones were obtained from the *T. aestivum* cv. Renan BAC library by screening with PCR primers specific to HMW glutenin genes. Assignment of BAC clones to the A, B, and D genomes was based on their characterization by restriction fragment length polymorphisms and contig assembly by BAC fingerprinting on agarose gels as described previously (KONG *et al.* 2004; CHANTRET *et al.* 2005). Selection of BACs for sequencing was based on contig maps constructed for each orthologous *Glu-1* locus and Southern hybridization data using various probes flanking the x-type and y-type HMW glutenin genes. BAC clones that covered the largest regions of each *Glu-1* locus were selected. The shotgun-sequencing libraries for hexaploid wheat BAC clones were constructed by the method described by either GU *et al.* (2003) or CHANTRET *et al.* (2005). Plasmid DNAs from single colonies were purified and inserts were sequenced from both directions with T7 and T3 primers using BigDye terminator chemistry (Applied Biosystems, Foster City, CA) on ABI3730 capillary sequencers. Gaps between sequence contigs were filled and sequenced by primer-walking and transposition reactions (Finnzymes, Espoo, Finland). Gaps caused by GC-rich regions were usually filled by resequencing using the dGTP BigDye terminator chemistry (Applied Biosystems).

Sequence analysis: For sequence assembly, a target of 10-fold coverage was chosen. Base calling and quality of the shotgun sequences were processed using Phred (EWING and GREEN 1998). The sequence data generated for each BAC clone was used to assemble continuous contigs using both the Lasergene SeqMan module (DNASTar) (<http://www.DNASTar.com>) and Phrap assembly engine (<http://www.phrap.org>). In some cases, assemblies with two different programs helped resolve gap regions and the order of contigs. The consensus sequences were obtained by analyzing at least three sequence reads (on both strands) or using sequencing methods based on two different labeling procedures applied on one strand. To validate the accuracy of the sequence assembly, digestion patterns of BAC DNAs with *Hind*III, *Eco*RI, and *Not*I were compared with the predicted restriction patterns of the computer-assembled sequences.

For annotation, the assembled sequences of the A, B, and D genome BACs from the hexaploid wheat were compared with the previous annotations for the *Glu-1* regions from the barley BAC (AY268139), diploid D-genome BAC (AF497474), tetraploid A-genome BAC (AY494981), and B genome BAC (AY368673). In addition, a homology search was performed against NCBI nonredundant and dbEST databases using BLASTN, BLASTX, and TBLASTX algorithms. FGENESH (<http://www.softberry.com/nucleo.html>) and GENESCAN (<http://genemark.mit.edu/GENESCAN.htm>) were used for gene prediction. DNA repetitive elements were identified with NCBI BLAST searches, with DNASTar MegAlign dot-plot analysis, and by comparison with the Triticeae Repeat Sequence Database at the GrainGenes website at <http://wheat.pw.usda.gov/ITMI/Repeats/>. The definition and naming of new retrotransposons were done according to the method described by SANMIGUEL *et al.* (1998).

The rate of nonsynonymous (K_a) vs. synonymous (K_s) substitutions were calculated for four genes (x-type, y-type HMW glutenin genes, globulin, and protein kinase genes) with MEGA3 (KUMAR *et al.* 2004). Dating of retrotransposon insertions was performed on the basis of the method described in SANMIGUEL *et al.* (1998). The number of transition and transversion mutations was calculated using the MEGA3 software (KUMAR *et al.* 2004). The average substitution rates of colinear retrotransposons in the two homologous genomes were also calculated in the same way and used to estimate the divergence times on the basis of the method described by WICKER *et al.* (2003).

The statistical significance regarding differences in rates of nucleotide substitution and indel frequency among each pair of the homologous A, B, and D genomes was measured by randomly sampling ~10-kb intervals in the *Glu-1* regions. The Student's *t*-test was employed for statistical analyses.

RESULTS

Isolation and sequencing of *Glu-1* regions from hexaploid wheat: Previously, we have reported sequences of the orthologous HMW glutenin regions from the D genome of the diploid *A. tauschii* and the A and B genomes of the tetraploid *T. turgidum*, representing the ancestral genomes of hexaploid wheat. To further examine the evolution of these genomic regions in hexaploid wheat, we screened a large insert BAC library constructed from the hexaploid wheat species *cv* Renan with PCR primers specific to the different copies of HMW glutenin genes. The screening of this BAC library identified 12 HMW glutenin BAC clones from the A

genome, 10 from the B genome, and 18 from the D genome. Further characterization of these BAC clones indicated that none of the BAC clones belonging to the A and B genomes contains both the y-type and x-type HMW glutenin genes (data not shown). BAC DNA was fingerprinted using the *Hind*III restriction enzyme and, using FPC software, contigs were built that span the *Glu-1* loci from each wheat genome (SODERLUND *et al.* 2000). The assembled contig maps and the BAC Southern data were used to guide the selection of BAC clones for sequencing. BAC clone 706G08 represents the *Glu-1* region from the D genome. The *Glu-1* region from the B genome was covered by the sequences from two BAC clones, 1289J04 and 2001P20. Three BACs were sequenced for the A-genome HMW glutenin region (BAC clones 1344C16, 754K10, and 1031P08). A 152,010-bp region derived from the D-genome BAC, a 285,506-bp region derived from the combined sequences of two B-genome BAC clones, and a 292,044-bp region from three A-genome BAC clones were annotated using a combination of bioinformatics tools and BLAST searches against publically available databases.

Gene colinearity between HMW glutenin genomic sequences across the three wheat ploidy levels: Sequence comparison of orthologous *Glu-1* regions between three homeologous wheat genomes revealed microcolinearity in the genic regions, but large-scale sequence divergence in the intergenic regions (GU *et al.* 2004). Sequencing the corresponding *Glu-1* regions from the three subgenomes of hexaploid wheat (A, B, and D) enables us to analyze sequence changes compared to homologous genomes from the diploid and tetraploid wheats. In all the sequenced *Glu-1* regions, there are a total of six genes (GU *et al.* 2004). The order of these genes is as follows: a leucine-rich-repeat receptor-like protein kinase, a globulin, a y-type HMW glutenin, a duplicate globulin, an x-type HMW glutenin, and a serine/threonine protein kinase. In the three genomes of hexaploid wheat, these six genes exist in the same order and orientation as those in the ancestor's genomes (Figure 1). Previously, a tandem duplication of an ancestral region containing the globulin and HMW glutenin genes was characterized at the *Glu-1* locus regions (KONG *et al.* 2004). This duplication, which led to the presence of the x-type and y-type copies of HMW glutenin genes in the wheat genome, occurred after the divergence of wheat and barley (GU *et al.* 2003; Figure 1). Extension of sequences in the 5' HMW glutenin regions identified a second tandem duplication, two copies of the receptor kinase genes present near the *Glu-1* locus. This duplication exists in at least the A and D genomes of the hexaploid wheat, suggesting the occurrence of the event prior to the divergence of the A and D genomes. The paralogous receptor kinase 1a and 1b in the hexaploid D genome share 90% sequence identity, whereas the orthologous receptor kinase 1a from the hexaploid D genome and the hexaploid A

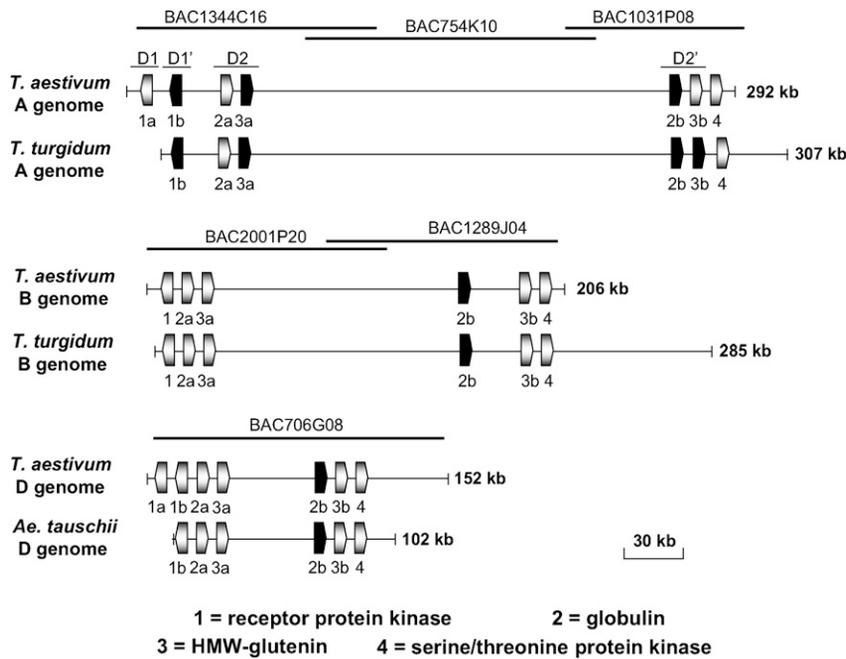


FIGURE 1.—Gene organization and comparison of orthologous and homologous *Glu-1* regions from ancestral wheat and hexaploid wheat genomes. Genes are represented by numbered pentagons, with their corresponding names listed below. The letters “a” and “b” denote distinct copies of the duplicated genes. The arrows of the pentagons indicate the direction of the potential transcription of these genes. Solid pentagons represent inactive genes or pseudogenes caused by sequence rearrangements identified in the analysis. Positions and addresses of BAC clones selected for complete sequencing of the regions are indicated. Regions involved in sequence duplications resulting in two copies of receptor protein kinase (D1, D1’) and two copies of HMW glutenin (D2, D2’) are labeled only in the A genome. A 30-kb segment is provided as a scale reference.

genome share 93% sequence identity. The receptor kinase 1b in both the hexaploid A genome and the tetraploid A genome is disrupted by the same transposable elements with identical insertion patterns (data not shown). Whether or not the duplication of the receptor kinase gene is present in the B genome is not determinable from the current data; the 5’-end of the B-genome BAC is downstream of where the receptor kinase 1a would be putatively located (Figure 1).

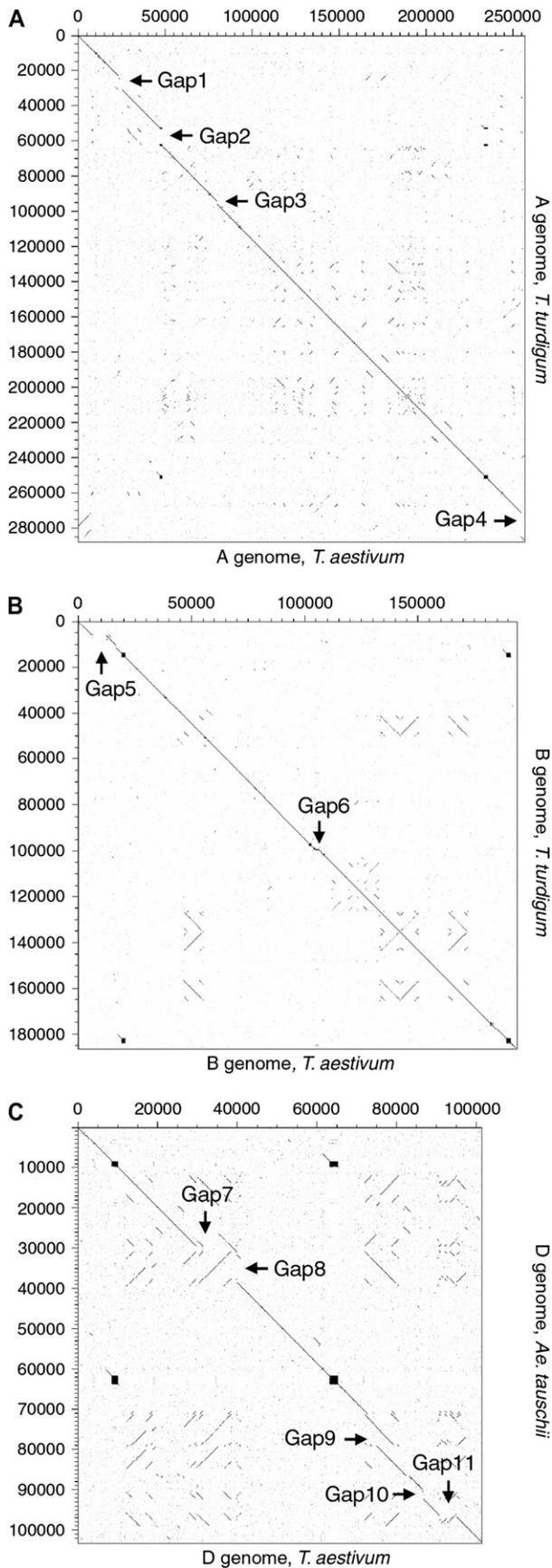
In an earlier study, we reported that although gene colinearity was not violated between homeologous wheat genomes, several genes were subjected to differential disruptions by various mechanisms, including repetitive DNA insertion, sequence deletion, and nucleotide substitutions causing an in-frame stop codon (Gu *et al.* 2004). Five genes found to be disrupted at the *Glu-1* locus regions in the hexaploid wheat appear to have the

same patterns of gene disruption as those identified in their ancestor genomes (Table 1). For example, the second globulin gene in the hexaploid B genome (2b) has the same four retrotransposon insertion events as those identified in the tetraploid B genome (KONG *et al.* 2004). Previously, we reported that the second globulin genes in the diploid D and tetraploid A genomes were disrupted by an identical deletion event, suggesting that this gene disruption occurred before the divergence of the wheat A and D genomes (ANDERSON *et al.* 2003; GU *et al.* 2004). In the hexaploid wheat, the second globulin gene (2b) in both the A and D genomes share this same deletion. The receptor kinase 1b in the hexaploid A genome contains the same multiple insertions of retroelements and miniature inverted repeat sequences (MITES) as the receptor kinase gene in the durum A genome (Gu *et al.* 2004). On the basis of this evidence, it

TABLE 1

Comparison of gene disruptions observed in hexaploid wheat and its ancestral genomes

	Receptor kinase 1b (A genome)	γ -type HMW-glutenin 3a (A genome)	Globulin 2b (A genome)	Globulin 2b (B genome)	Globulin 2b (D genome)
<i>T. turgidum</i>	Insertions of multiple retrotransposons	Point mutation resulting in a premature stop codon	Deletion of a coding sequence region	Insertions of multiple retrotransposons	
<i>T. aestivum</i>	Insertions of multiple retrotransposons	Point mutation resulting in a premature stop codon	Deletion of a coding sequence region	Insertions of multiple retrotransposons	Deletion of a coding sequence region
<i>Ae. tauschii</i>					Deletion of a coding sequence region



appears that the disruption of these genes in the hexaploid wheat had already occurred in its diploid or tetraploid progenitors.

Genes disrupted in only one of the homologous genomes were also identified. The Ax HMW glutenin gene falls within this category. A point mutation in the tetraploid Ax HMW glutenin gene has resulted in a premature stop codon, while the Ax HMW glutenin gene in the hexaploid wheat is intact.

Sequence comparison of the *Glu-1* regions across the three wheat ploidy levels: To further analyze the sequence variation present in homologous wheat genomes, we performed dot matrix analyses between pairs of corresponding genomes from different ploidy wheats. Sequence divergences are seen as disruptions in the main matrix diagonal line. Figure 2 suggests that the sequences between the homologous *Glu-1* regions are generally conserved. The gaps along the diagonal lines represent types of sequence rearrangements, such as deletions/insertions, duplications, and inversions, that differentiate the two homologous genomes. LTR retrotransposable elements are usually 6–8 kb and insertions of such elements that occur in only one of two compared regions will result in large gaps in the dot matrix analysis. The HMW glutenin region of the A genome of *T. turgidum* contains three nonshared retroelement insertions when compared with the corresponding region from the A genome of *T. aestivum*. In the region between the receptor kinase 1b and globulin 2a, there is a WIS-type element (*Wis-3*) insertion present in the coding region of a gypsy-class element *Boba-1* (Gap1). Gap2 was caused by the insertion of *Wis-4* in the coding region of the Ay HMW glutenin gene in tetraploid wheat. In the region between the Ax HMW glutenin and the serine/threonine protein kinase genes, there is a gypsy-class element, *Erika-2*, that is not present in the corresponding area of our *T. aestivum* sequence (Gap4). These results provide evidence that these elements have inserted since the two genomes last shared a common ancestor. In addition to these retroelement insertions, there are also two significant indel events in the *T. aestivum* sequence. The first of these removed the entire coding region and a fragment of each LTR from the element *Madil-1*, located in the intergenic region between the receptor kinase and globulin genes, leaving behind a sequence similar to a solo LTR in size and structure (Gap3). The second indel,

FIGURE 2.—Pairwise comparisons of the *Glu-1* regions between the two homologous genomes. The dot plot was performed using the DOTTER program (SONNHAMMER and DURBIN 1995) with default parameters between two A-genome sequences from *T. turgidum* and *T. aestivum* (A), two B-genome sequences from *T. turgidum* and *T. aestivum* (B), and two D-genome sequences from *Ae. tauschii* and *T. aestivum* (C). Each distinct gap in the dot plot is assigned a number, and sequence rearrangements causing the gaps are described in the text.

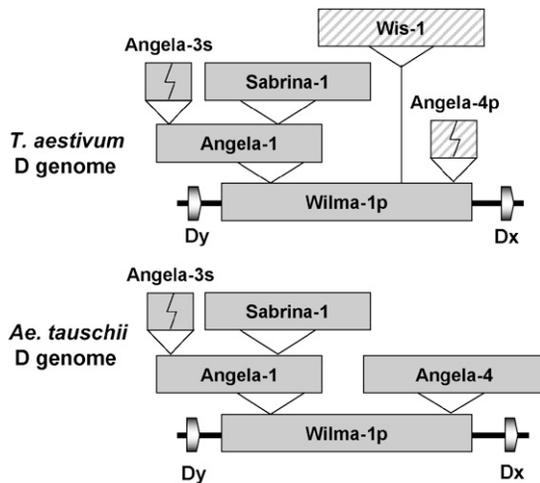


FIGURE 3.—Insertion and deletion of retrotransposons that differentiate two homologous D genomes. Retrotransposons that are colinear between the hexaploid D and diploid D genomes are boxed in gray. Retrotransposons that have inserted or rearranged by deletion in only one of the genomes are represented by hatched boxes. Genes are represented by pentagons. Dy and Dx are short for the y-type and x-type HMW-glutenin genes from the D genome. The letters “s” and “p” following the names of the retroelements indicate solo or partial, respectively.

which occurred in a relatively old CACTA element, *Jorge-1*, removed a portion of the coding sequence and possibly a MITE, *Eos-1*, as well.

By contrast, the HMW glutenin regions of the two B genomes revealed only a few major sequence differences. A 6-kb indel (Gap5) is present in the intergenic region between the receptor protein kinase 1b and globulin 2a coding regions. While the sequence of this 6-kb indel showed no sequence similarity to known transposable element sequences or structures, it is flanked by an 8-bp (TAAGAATT) perfect repeat, suggesting that illegitimate recombination resulted in this indel (WICKER *et al.* 2003; CHANTRET *et al.* 2005). In addition to the 6-kb indel, there is a smaller indel, ~1.7 kb (Gap6), present within the class I transposable element *Sogi-1*. Sequence comparisons revealed no novel retroelement insertions that could serve to differentiate the two B genomes.

The HMW glutenin regions of the two D genomes are the most divergent at the large-scale sequence rearrangement and insertion/deletion level. Since the time when the D-genome donors of sequenced accessions of *Ae. tauschii* and *T. aestivum* shared a common ancestor, three novel retroelement insertions have occurred in this area. In *Ae. tauschii*, two copia class retroelements, *Angela-5* (Gap9) and *Wis-1s* (Gap10), have inserted into a large block of retroelements in the region downstream of the serine/threonine protein kinase. In *T. aestivum*, a complete copy of the element *Wis-1* (Gap8) is present in the nested retroelement structure between the two HMW glutenin genes (Figure 3). In addition to these differential insertions, there is also a deletion of the

coding region and portions of both LTRs from *Angela-4* (Gap7), resulting in a partial LTR retroelement, *Angla-4p* in *Ae. tauschii* (Figure 3). A series of sequence rearrangements/insertions in the 5' LTR sequence of *Angela-7p* in *T. aestivum* generated Gap11 (Figure 2).

Shared and nonshared transposable elements between *Glu-1* homologous regions: Despite indel events shown in Figure 2, the dot matrix analyses revealed general sequence conservation of the *Glu-1* regions between two homologous genomes, suggesting that many transposable elements in the intergenic regions are shared or colinear. To test this, we calculated the number of shared retroelements *vs.* nonshared retroelements in the *Glu-1* regions from pairs of homologous genomes. On the basis of the activity of retroelement insertions, the two B genomes are the most conserved, with all 20 of the retroelements identified present in both genomes (Figure 4A). In the ~300-kb *Glu-1* regions from the two A genomes, we identified 35 retroelements that are present in both genomes. The two nonshared retroelements are both present in the A genome of durum wheat. In the ~100-kb homologous *Glu-1* regions of the two D genomes, 14 retroelements are found to be shared, while 3 are nonshared. Among these 3 new retroelements, there is 1 intact retroelement, one solo LTR, and 1 partial retroelement missing one LTR and a portion of the coding region. The solo LTR and the partial LTR retroelement that differentiate the two D genomes are somewhat surprising since newly inserted retroelements would be expected to be intact (MA *et al.* 2004; BRUNNER *et al.* 2005). However, this may suggest that insertions of new transposable elements can be followed by DNA recombination.

Although nonshared retroelements were observed in the homologous *Glu-1* regions between two A and D genomes, the frequency of nonshared retroelements is much lower than those found in homologous regions from different maize inbred lines, the *Rph7* locus regions from two different barley accessions, the *Lr10* locus from homologous A genomes in different ploidy levels, and the *Ha* locus from homologous A, B, and D genomes in different ploidy levels (BRUNNER *et al.* 2005; CHANTRET *et al.* 2005; ISIDORE *et al.* 2005; SCHERRER *et al.* 2005). In this study, only 5.7% of the retroelements are nonshared in the *Glu-1* regions from the two A genomes, 0% from the two B genomes, and 21% from the two D genomes. The nonshared retroelements account for 18.1% of the sequence divergence between the two D genomes and 8.2% between the two A genomes.

Sequence insertions and deletions in the homologous *Glu-1* regions: Indels play a fundamental role in genome evolution through the loss or gain of DNA in specific regions (GREGORY 2004; TAYLOR *et al.* 2004). There is still little known about the distribution and frequency of indels in plant genomes. In addition to the large indels that cause visible gaps in the dot matrix analysis (Figure 2), we also analyzed smaller indels

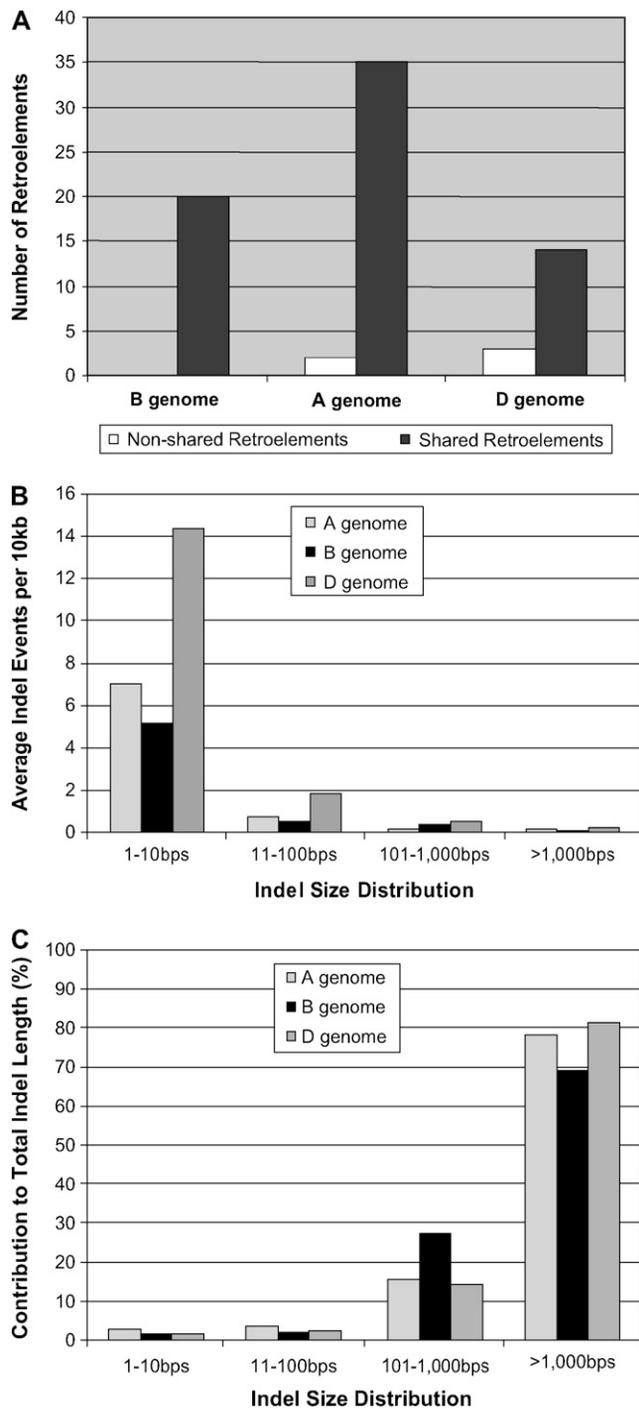


FIGURE 4.—A. Number of shared *vs.* nonshared retroelements between the two homologous wheat genomes. Shared and non-shared retroelements between the homologous genomes are represented by the black and gray bars, respectively. B. Indel frequency and size distribution in three wheat genomes. Indels identified in each pair of homologous genomes were categorized into four groups. The indel frequency for each group was normalized to the number of indels per 10 kb in the graph. C. Percentage of total indel length contributed by indels of different sizes. Indel length represented by different indel sizes were calculated and then divided by the total indel length for each of the A, B and D genomes. The total indel lengths are 10,738, 11,382, and 13,612 bp for the A, B, and D genomes, respectively.

present in the *Glu-1* regions of each pair of the homologous wheat genomes (Figure 4B). Clearly, smaller indels (≤ 10 bp) greatly outnumber large indels in terms of frequency in all three genomes. They account for 88.2% of the indel frequency for the A genome, 83.3% for the B genome, and 85.0% for the D genome.

When the indel frequencies were compared in the A, B, and D genomes, the frequency of indels differed significantly between any two genomes in the *Glu-1* regions ($P < 0.05$, Student's *t*-test). The D genome had the highest frequency of indels (16.9 indels/10 kb). The indel frequency for the A and B genomes was 8.0 and 6.2, respectively. Indels are a type of sequence rearrangement event and the frequency reflects the evolutionary distance of two related genomes in comparison (OGURTSOV *et al.* 2004). Therefore, our data suggest that the two D genomes are more distantly related as compared to the two A and B genomes.

Indels directly contribute to the sequence divergence between two compared sequences. We calculated the total length of sequences represented by the indels. The total length of the indel sequence is 10,738 bp for the A genome, 11,382 bp for the B genome, and 13,612 bp for the D genome. They account for 4.0, 6.2, and 13.3% of sequence divergence in the compared regions for the A, B, and D genomes, respectively. When the length of indels represented by different indel sizes was calculated, indels >100 bp contributed much more to the total indel length (Figure 4C). They account for 93.5, 96.5, and 95.8% of the total indel length for the A, B, and D genomes, respectively. These results suggest that large indels are primarily responsible for the indel-based sequence divergence between the two homologous wheat genomes.

Nucleotide substitutions in the *Glu-1* region of the homologous wheat genomes: Another type of common sequence variation, occurring in even the most conserved regions, is the single nucleotide polymorphism, caused by nucleotide substitutions (BRUMFIELD *et al.* 2003). The high level of sequence conservation observed in the *Glu-1* regions of the homologous genomes allows us to examine nucleotide substitutions in large contiguous segments spanning both the genic and intergenic regions. We calculated that the average number of nucleotide substitutions *per site* is 0.0093, 0.0056, and 0.0137 for the A, B, and D genomes, respectively. The variations in average nucleotide substitutions are all statistically significant ($P < 0.001$). The data indicate that the nucleotide substitution rates for the A and B genomes are different despite the fact that the two genomes have been co-evolving in the nuclei of the same species.

The rate of nucleotide substitutions in intergenic regions is most often considerably higher than those found in genic regions and protein-coding sequences (MA and BENNETZEN 2004). We also compared the

TABLE 2

 K_s and K_a values and standard errors for pairwise comparisons of four genes in the *Glu-1* region

	A genome		B genome		D genome	
	K_s	K_a	K_s	K_a	K_s	K_a
<i>Globulin 2a</i>	0	0.0019 (± 0.0019)	0.0059 (± 0.0064)	0.0020 (± 0.0020)	0.0059 (± 0.0058)	0.0020 (± 0.0019)
<i>HMW-glutenin 3a</i>	0.0162 (± 0.0068)	0.0078 (± 0.0026)	0.0055 (± 0.0031)	0.0055 (± 0.0017)	0.0395 (± 0.0101)	0.0108 (± 0.0029)
<i>HMW-glutenin 3b</i>	0.0081 (± 0.0035)	0.0069 (± 0.0019)	0.0142 (± 0.0048)	0.0165 (± 0.0030)	0.0061 (± 0.0031)	0.0050 (± 0.0017)
<i>Protein kinase 4</i>	0	0.0014 (± 0.0010)	0	0.0012 (± 0.0011)	0.0049 (± 0.0034)	0.0020 (± 0.0012)

number of nucleotide substitutions *per site* between the HMW glutenin genes with those of the intact globulin and protein kinase genes that flank the *Glu-1* locus in all three wheat genomes. The number of synonymous substitutions (K_s) and nonsynonymous substitutions (K_a) *per site* between genes from the homologous genomes is given in Table 2. Considerable variations in nucleotide substitutions are observed among these genes, suggesting that different genes evolve at different rates. The globulin 2a in the two A genomes and protein kinase genes in both the A and B genomes appear to be more conserved. The number of nucleotide substitutions *per site* in the HMW glutenin regions is at least similar to or greater than that found in the neighboring intergenic regions. The high nucleotide substitution rate coincides with the previous observation that HMW glutenin genes are highly polymorphic in wild wheat and display significant variation in modern cultivars (ALLABY *et al.* 1999; SHEWRY *et al.* 2003).

Time of divergence of the homologous genomes: In the *Glu-1* regions, the identification of a number of LTR retrotransposons that are intact and shared by two homologous genomes allows us to further examine the sequence changes and divergence times of the homologous genomes in wheat. Because of the particular mechanism of reverse transposition, the two LTR sequences from a single retrotransposon are identical at the time of its insertion (BOEKE and CORCES 1989). On the basis of this property, the nucleotide substitutions identified in two LTRs of a retrotransposon can be used to estimate the date of its insertion (SANMIGUEL *et al.* 1998). Previously, we estimated insertion times for 22 LTR retrotransposons in the wheat *Glu-1* region using a mutation rate of 6.5×10^{-9} substitutions/synonymous site/year (GUAT *et al.* 1996; SANMIGUEL *et al.* 1998). It appears that these datable LTR retrotransposons have all inserted into their current positions within the last 4–5 MY (GU *et al.* 2004). A similar methodology can be employed to estimate divergence times for two homologous sequences on the basis of the number of nucleotide substitutions accumulated after the ancestor genome started to split into two distinct descendants. However, considerable variation in estimated divergence time has been noted when single gene sequences are used (HUANG *et al.* 2002; DEVOS *et al.* 2005). One

possible explanation is that rates of nucleotide substitution are different for genes that are under different selective forces. A better representation of divergence time for two closely related sequences may be obtained by comparing multiple regions, ideally intergenic sequences that have few selective forces acting on them (WAKELEY and HEY 1997). In this study, we used the same colinear LTR retrotransposons previously identified and dated in the *Glu-1* regions to estimate the divergence time of the two homologous genomes. A total of 8–9 intact colinear LTR retrotransposons for each B and each A genome pair, representing ~40–50 kb of sequence, were used to estimate the divergence times. For the two homologous D genomes, only two full-length, datable LTR retrotransposons are colinear; we identified additional partial yet colinear retroelements to obtain ~40 kb of total sequence for comparison. Using the same molecular clock for dating the LTR retrotransposon insertions, we estimated the divergence time by calculating rates of nucleotide substitution between each pair of colinear retroelements to examine the variation in different sequences (Table 3). The divergence times estimated, using common retroelements, for the two homologous A genomes range from 0.38 to 1.2 MY, with an average of 0.81 MY. The two B genomes are estimated to have diverged in the last 0.24–0.78 MY, with an average of 0.48 million years ago (MYA). The two D genomes are the most divergent, ranging from 1.12 to 1.86 MYA with an average of 1.39 MYA. The longer divergence time between the two D genomes shows that the diploid D-genome sequence is divergent from the D-genome donor(s) of the hexaploid wheat by ~1.39 MY (Table 3). It has been noted that LTR retroelements evolve at least two times faster than genes and UTR regions (SANMIGUEL *et al.* 1998; WICKER *et al.* 2003; MA and BENNETZEN 2004). If this is the case, the divergence times estimated for each pair of the homologous genomes should be divided by a factor of 2.

DISCUSSION

In this study, a detailed sequence comparison was performed to study the organization and evolution of the hexaploid wheat genome with its ancestral diploid and tetraploid wheat genomes. On the basis of the

TABLE 3
Estimates of divergence time of the homologous wheat genomes

LTR retrotransposons	Divergence time (MYA)	SD
Shared in A genome		
<i>Pivu-1</i>	0.38	0.11
<i>Gujog-1</i>	0.80	0.35
<i>Nusif-1</i>	0.97	0.29
<i>Apüp-1</i>	1.20	0.27
<i>Sabrina-2</i>	0.60	0.17
<i>Fatimah-1</i>	1.13	0.42
<i>Sabrina-3</i>	0.71	0.19
<i>Ames-1</i>	0.75	0.12
<i>Fatimah-2</i>	0.74	0.35
Average	0.81	0.24
Shared in B genome		
<i>Ifis-1</i>	0.37	0.22
<i>Wis-1</i>	0.55	0.16
<i>Fatimah-1</i>	0.57	0.15
<i>Sabrina-1</i>	0.24	0.10
<i>Derami-1</i>	0.38	0.16
<i>Wis-2</i>	0.37	0.13
<i>Wis-3</i>	0.57	0.16
<i>Wis-4</i>	0.78	0.18
Average	0.48	0.16
Shared in D genome		
<i>Wilma-1p</i>	1.18	0.25
<i>Angela-2</i>	1.12	0.23
<i>Sabrina-2</i>	1.34	0.25
<i>Angela-3s</i>	1.59	0.27
<i>Angela-7p</i>	1.30	0.23
<i>Latitude-1s</i>	1.74	0.20
<i>Wham-1p</i>	1.86	0.32
<i>Sabrina-4p</i>	1.13	0.23
Average	1.39	0.28

Full-length LTR retrotransposons that are colinear in the *Glu-1* regions were used to calculate the divergence time of two homologous genomes. Divergence time was estimated as described in SANMIGUEL *et al.* (1998).

sequence analysis of *Glu-1* regions, our results reveal that although sequence rearrangements differentiating each pair of two homologous genomes are easily visible, a considerable portion of the sequences are highly conserved between the two homologous genomes, including large segments of intergenic regions composed of nested retroelements. Although retroelements provide the majority of sequence divergence between two homologous genomes, indels and nucleotide substitutions are the most frequent events in the compared *Glu-1* regions.

Sequence conservation in the homologous *Glu-1* region: Our previous study indicated that microcolinearity is maintained in the orthologous *Glu-1* regions from homeologous wheat genomes, but intergenic regions were not conserved due to rapid amplification/deletion of retroelements (GU *et al.* 2004). This study revealed

general sequence conservation between two homologous genomes from the different wheat ploidy levels. In addition to the conservation of the genic regions, sequences in the intergenic regions are also highly colinear (Figure 4A). This suggests that a vast majority of retroelements in the hexaploid wheat genome were inherited from their diploid and tetraploid ancestral genomes. In the *Glu-1* region of the tetraploid A genome, an intergenic region between the y-type and x-type HMW glutenin genes contains large blocks of nested retrotransposon insertions, with as many as 19 members spanning a region of 140 kb (GU *et al.* 2004). The same nested retroelement structure is also present in the hexaploid A genome, suggesting that the intergenic region has not been drastically changed due to major sequence rearrangements. The presence of a high number of colinear retroelements in the *Glu-1* regions from different ploidy wheat genomes is surprising, considering the results from similar comparative sequence studies on other homologous locus regions. In the genomes of different maize inbred lines, it was found that 70% of LTR retrotransposons are allele specific (BRUNNER *et al.* 2005). When homologous regions containing the barley *Rph7* locus from resistance and susceptible lines were compared, the number and type of repetitive elements were completely different in 65% of the sequenced regions (SCHERRER *et al.* 2005). In the wheat *Lr10* region, transposon insertions resulted in >70% sequence divergence among three A genomes from diploid, tetraploid, and hexaploid wheats (ISIDORE *et al.* 2005). In the *Glu-1* regions, retroelements resulted in only 8.2% of nonshared sequence between the two A genomes, 18.1% between the two D genomes, and 0% between the two B genomes.

In the *Glu-1* regions, violations of gene colinearity were not detected between two homologous genomes. This is also significantly different from the other genetic loci studied in maize (BRUNNER *et al.* 2005), barley (SCHERRER *et al.* 2005), and wheat (WICKER *et al.* 2003; CHANTRET *et al.* 2005; ISIDORE *et al.* 2005). One possible explanation for the sequence conservation observed in the *Glu-1* region could be that the hexaploid line (Renan) selected in this study has the same or a very similar haplotype as the tetraploid wheat *cv.* Langdon. However, our haplotype analyses suggest that the two homologous A and B genomes belong to different haplotypes (see supplemental data at <http://www.genetics.org/supplemental>). Therefore, our results suggest that the wheat *Glu-1* locus is a conserved genetic region, likely with a low recombination rate since recombination rates are positively correlated with the frequency of locus duplication and deletions that often cause synteny perturbation between wheat homeologous chromosomes (AKHUNOV *et al.* 2003). Furthermore, the conservation of the *Glu-1* locus regions is further supported by the lack of recombination events observed between x- and y-type HMW glutenin genes, despite the wide

interest in breaking the linkage for wheat quality improvement in breeding (SHEWRY *et al.* 2003). In addition, our data indicate that at the HMW glutenin loci there is no evidence for polyploidization-induced genomic changes that resulted in significant sequence rearrangements.

Sequence rearrangements by indels in the wheat *Glu-1* regions: Despite the general conservation in the *Glu-1* regions, indels, particularly small indels, are frequently observed between the homologous genomes. In the *Glu-1* regions, small indels (<10 bp, ~85%) greatly outnumber larger indels (>10 bp, ~15%) in all three pairs of homologous genomes. Small indels often occur in coding regions, making them a major player in gene evolution (TAYLOR *et al.* 2004). In addition, small indel events might be associated with nucleotide substitutions, resulting in an increased rate of single nucleotide polymorphisms (MA and BENNETZEN 2004). Indel occurrences have been used to estimate evolutionary distance between related genomes (OGURTSOV *et al.* 2004). In this study, we also detect a positive correlation between the frequency of indels and the rate of nucleotide substitution. The two homologous D genomes that have the longest divergence times show the highest indel frequency and greatest rate of nucleotide substitution.

In the homologous *Glu-1* regions, although small indels are more ubiquitous, they contribute less to the genome size difference. However, large indels contribute more heavily to overall sequence divergence in closely related genomes (Figure 4C). In addition, it is often difficult to distinguish if an indel was caused by an insertion or a deletion. Deletions generally outnumber insertions (BLUMENSTIEL *et al.* 2002; GREGORY 2004; MA and BENNETZEN 2004). In the *Glu-1* regions, most large indels were identified to be deletions since they occur in the sequences of well-characterized retroelements. It is likely that nonintact or repetitive DNA fragments observed in the intergenic regions are caused mainly by deletion events.

In general, small indels are caused by replication slippages (GREGORY 2004), whereas large indels seems to involve different mechanisms. One mechanism is the unequal homologous recombination that often acts on LTR transposable elements, resulting in solo LTRs (DEVOS *et al.* 2002; MA *et al.* 2004). In contrast, indels caused by illegitimate recombination appear to occur in any region of nucleotide sequence with as little as a few base pairs of sequence identity (WICKER *et al.* 2003; CHANTRET *et al.* 2005). It has been reported that the deletion events attributable to illegitimate recombination are much more frequent than deletion events caused by unequal homologous recombination (WICKER *et al.* 2003). Illegitimate recombination is primarily responsible for the removal of nonessential DNA in Arabidopsis (DEVOS *et al.* 2002). In wheat, indels caused by illegitimate DNA recombination are one of the major molecular mechanisms responsible for reshaping the

Ha locus during wheat genome evolution (CHANTRET *et al.* 2005). Our results indicate that indels are dynamic evolutionary processes that contribute to sequence divergence between homologous wheat genomes.

Wheat genome evolution: Hexaploid wheat contains three homeologous genomes. The origin and evolution of these evolutionarily closely related genomes has been analyzed phylogenetically using sequences from HMW glutenin genes (ALLABY *et al.* 1999; BLATTER *et al.* 2004). The results indicated that wheat homeologous genomes diverged ~5.0–6.9 MYA (ALLABY *et al.* 1999).

In this study, we compared sequence changes among the homologous wheat genomes. Analysis of the *Glu-1* regions from tetraploid and hexaploid wheat revealed that the A genome exhibits higher sequence variation than the B genome. This is supported by at least four lines of evidence. First, the rate of nucleotide substitution in the A genome is significantly higher than that detected in the B genome ($P < 0.01$). Second, the B genomes are devoid of nonshared retroelements, whereas two nonshared retrotransposons were identified in the A genome (Figure 4A). Third, the indel frequency in the A genome is also significantly higher than that found in the B genome ($P < 0.01$) (Figure 4). Fourth, the divergence times estimated for A and B genomes are significantly different (Table 3). The greater sequence conservation was also reported for the same two B genomes at the *Ha* locus on chromosome 5, where there are no nonshared retroelements and the two sequences have 99% identity (CHANTRET *et al.* 2005), suggesting that such sequence conservation is present at multiple genetic regions and is unlikely caused by introgression recombination. Although the possibility of introgression recombination cannot be completely excluded, the variation of sequence divergence between the A and B homologous pairs could be explained by a hypothesis that the A and B genomes in the tetraploid and hexaploid wheats evolve at different rates. This explanation would have to be based on the assumption that significant sequence changes have occurred since the tetraploidization event, which was estimated to have occurred in the last 0.36–0.5 MY (HUANG *et al.* 2002; DVORAK and AKHUNOV 2005). Different rates of genome evolution have been noted for *japonica* and *indica* rice, which have experienced independent variation for ~0.44 MY (MA and BENNETZEN 2004). However, in this study, both the A and B genomes are the subgenomes in a single polyploid wheat. As far as we know, no reports have shown that different subgenomes in a single species are subject to different evolutionary rates.

Another likely explanation for the observed difference in sequence variation is that hexaploid wheats have more than one tetraploid ancestor, resulting from independent tetraploidization events, in which two divergent A genomes hybridized with two less divergent B genomes, resulting in the different rates of sequence variation observed in our study. Our estimates of the

divergence times for the two A genomes and the two B genomes, calculated by comparing colinear retroelements, is in accordance with this hypothesis. Our results indicated that the two B genomes have diverged for ~0.48 MY, while the two A genomes have been separated for the last ~0.81 MY. Another piece of supporting evidence includes the finding from another study that the A genome from the hexaploid wheat Renan has a different haplotype than the A genome from durum wheat at the *Lr10* locus and that this haplotype originated from ancient DNA rearrangements at the diploid level (ISIDORE *et al.* 2005). Furthermore, we previously reported that the sequence in the *Glu-1* locus from the hexaploid wheat *cv.* Chinese Spring is more closely related to that of the A genome from durum wheat, primarily because of shared mechanisms of gene disruption in the two orthologous Ay HMW glutenin genes and because of higher sequence identity compared to allelic HMW glutenin genes in other bread wheats, such as Cheyenne (GU *et al.* 2004). It is likely that the Chinese Spring A genome has the same lineage as the A genome from the durum wheat and that the A genome in Renan and Cheyenne represents a different A genome lineage. Wheat exists at three ploidy levels, and multiple origins of hexaploid wheat have been suggested (DVORAK *et al.* 1998; TALBERT *et al.* 1998; BLATTER *et al.* 2004). Such multiple independent polyploidizations would serve to greatly increase the genetic diversity in the wheat gene pool. Our results provide further evidence based on the analysis of sequences of large wheat genomic regions spanning the *Glu-1* locus across three ploidy levels. This study allows us to examine the various events of sequence changes that have occurred in the evolutionary history of wheat genomes. A genomewide analysis using various tools such as haplotype genotyping in different ploidy wheats will promote us to elucidate the occurrence of wheat speciation and the molecular evolution of the wheat genome.

We thank Mingcheng Luo for providing wheat materials for the haplotype analysis, Frank You for assistance in statistical and bioinformatics analysis, and Roger Thilmony for critical reading of the manuscript. We sincerely thank the Genoplante consortium (<http://www.genoplate.com>) for making available the BAC library from the *T. aestivum* cultivar Renan and the *Glu-1* BAC clones. Sequencing of the BAC clones 1289J4 and 1001P20 covering the *Glu-1* regions from the B genome was supported by the Genoplante consortium. Sequencing of the other BAC clones was supported by U.S. Department of Agriculture-Agriculture Research Service grant CRIS 5325022100-011. This work is also supported in part by National Science Foundation Plant Genome grant no. DBI-0321757.

LITERATURE CITED

- AKHUNOV, E. D., A. R. AKHUNOV, A. M. LINKIEWICZ, J. DUBCOVSKY, D. HUMMEL *et al.*, 2003 Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. *Proc. Natl. Acad. Sci. USA* **100**: 10836–10841.
- 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.
- ANDERSON, O. D., C. RAUSCH, O. MOULLET and E. S. LAGUDAH, 2003 The wheat D-genome HMW-glutenin locus: BAC sequencing, gene distribution, and retrotransposon cluster. *Funct. Integr. Genomics* **3**: 56–68.
- BENNETZEN, J. L., J. MA and K. M. DEVOS, 2005 Mechanism of recent genome size variation in flowering plants. *Ann. Bot.* **95**: 127–132.
- BLAKE, N. K., B. R. LEHFELDT, M. LAVIN and L. E. TALBERT, 1999 Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. *Genome* **42**: 351–360.
- BLANC, G., and K. H. WOLFE, 2004 Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* **16**: 1679–1691.
- 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 B1–1 and A1–2 subunit genes. *Theor. Appl. Genet.* **108**: 360–367.
- BLECHL, A. E., and O. D. ANDERSON, 1996 Expression of a novel high-molecular-weight glutenin subunit gene in transgenic wheat. *Nat. Biotechnol.* **14**: 875–879.
- BLUMENSTIEL, J. P., D. L. HARTL and E. R. LOZOVSKY, 2002 Pattern of insertion and deletion in contrasting chromatin domains. *Mol. Biol. Evol.* **19**: 2211–2225.
- BOEKE, J. D., and V. G. CORCES, 1989 Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* **43**: 403–434.
- BRUMFELD, R. T., P. BEERLI, D. A. NICKERSON and S. T. EDWARDS, 2003 The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.* **18**: 249–256.
- BRUNNER, S., K. FENGLER, M. MORGANTE, S. TINGEY and A. RAFALSKI, 2005 Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* **17**: 343–360.
- CHANTRET, N., J. SALSE, F. SABOT, S. RAHMAN, A. BELLAC *et al.*, 2005 Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploidy wheat species (*Triticum* and *Aegilops*). *Plant Cell* **17**: 1033–1045.
- COMAI, L., A. P. TYAGI, K. WINTER, R. HOLMES-DAVIS, S. H. REYNOLDS *et al.*, 2002 Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* **12**: 1551–1568.
- DEVOS, K. M., J. K. BROWN and J. L. BENNETZEN, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12**: 1075–1079.
- DEVOS, K. M., J. BEALES, Y. OGIHARA and A. N. DOUST, 2005 Comparative sequence analysis of the phytochrome C gene and its upstream region in allohexaploid wheat reveals new data on the evolution of its three constituent genomes. *Plant Mol. Biol.* **58**: 625–641.
- DVORAK, J., and E. D. AKHUNOV, 2005 Tempos of gene locus deletion and duplications and their relationship to recombination rate during diploid and polyploidy evolution in the Aegilops-Triticum alliance. *Genetics* **171**: 323–332.
- DVORAK, J., P. DI TERLIZZI, H-B. ZHANG and P. RESTA, 1992 The evolution of polyploid wheats: identification of the A genome donor species. *Genome* **36**: 21–31.
- 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.
- EWING, B., and P. GREEN, 1998 Base-calling of automated sequencer traces using *Phred II*. Error probabilities. *Genome Res.* **8**: 186–194.
- FELDMAN, M., F. G. H. LUPTON and T. E. MILLER, 1995 Wheats, pp. 184–192 in *Evolution of Crops*, Ed. 2, edited by J. SMARTT and N. W. SIMMONDS. Longman Scientific, London.
- GREGORY, T. R., 2004 Insertion-deletion biases and the evolution of genome size. *Gene* **423**: 15–34.
- GU, Y. Q., O. D. ANDERSON, C. LONDEORE, X. KONG, R. N. CHIBBAR *et al.*, 2003 Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. *Genome* **46**: 1084–1097.
- GU, Y. Q., D. COLEMAN-DERR, X. 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.

- GUAT, B. S., B. R. MORTON, B. C. MCCAIG and M. T. CLEGG, 1996 Substitution rate comparisons between grasses and palm: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcl*. *Proc. Natl. Acad. Sci. USA* **93**: 10274–10279.
- HE, P., B. R. FRIEBE, B. S. GILL and J. M. ZHOU, 2003 Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant Mol. Biol.* **52**: 401–414.
- HUANG, S., A. SIRIKHACHORNKIT, X. SU, J. FARIS, B. GILL *et al.*, 2002 Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl. Acad. Sci. USA* **99**: 8133–8138.
- ISIDORE, E., B. SCHERRER, B. CHALHOUB, C. FEUILLET and B. KELLER, 2005 Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. *Genome Res.* **15**: 526–536.
- KASHKUSH, K., M. FELDMAN and A. A. LEVY, 2003 Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* **33**: 102–106.
- KONG, X., Y. Q. GU, F. M. YOU, J. DUBCOVSKY and O. D. ANDERSON, 2004 Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. *Plant Mol. Biol.* **54**: 56–69.
- KUMAR, S., K. TAMURA and M. NEI, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinformatics* **5**: 150–163.
- MA, L., and J. L. BENNETZEN, 2004 Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. USA* **101**: 12404–12410.
- MA, J., K. M. DEVOS and J. L. BENNETZEN, 2004 Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14**: 860–869.
- MASTERSON, J., 1994 Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**: 421–423.
- OGURTSOV, A. Y., S. SUNYAEV and A. S. KONDRASHOV, 2004 Indel-based evolutionary distance and mouse-human divergence. *Genome Res.* **14**: 1610–1616.
- OZKAN, H., A. A. LEVY and M. FELDMAN, 2001 Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13**: 1735–1747.
- PETROV, D. A., T. A. SANGSTER, J. S. JOHNSTON, D. L. HARTL and K. L. SHAW, 2000 Evidence for DNA loss as a determinant of genome size. *Science* **287**: 1060–1062.
- ROOKE, L., F. BEKES, R. FIDO, F. BARRO, P. GRAS *et al.*, 1999 Overexpression of a gluten protein in transgenic wheat results in greatly increased dough strength. *J. Cereal Sci.* **30**: 115–120.
- SANMIGUEL, P., A. TIKONOV, Y. K. JIN, N. MOTCHOULSKAIA, D. ZAKHAROV *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- SANMIGUEL, P., B. S. GAU, A. TIKHONOV, Y. NAKAJIMA and J. L. BENNETZEN, 1998 The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**: 43–45.
- SANMIGUEL, P. J., W. RAMAKRISHNA, J. L. BENNETZEN, C. S. BUSO and J. DUBCOVSKY, 2002 Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). *Funct. Integr. Genomics* **2**: 70–80.
- SCHERRER, B., E. ISIDORE, P. KLEIN, J. S. KIM, A. BELLAC *et al.*, 2005 Large intraspecific haplotype variability at the *Rph7* locus results from rapid and recent divergence in the barley genome. *Plant Cell* **17**: 361–374.
- SHAKE, H., K. KASHKUSH, H. OZKAN, M. FELDMAN and A. A. LEVY, 2001 Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **13**: 1749–1759.
- SHEWRY, P. R., and A. S. TATHAM, 1990 The prolamin storage proteins of cereals: structure and evolution. *Biochem. J.* **267**: 1–12.
- SHEWRY, P. R., N. G. HALFFORD and A. S. TATHAM, 2003 The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. *Adv. Food Nutri. Res.* **45**: 221–303.
- SODERLUND, C., S. HUMPHRY, A. DUNHAM and L. FRENCH, 2000 Contigs built with fingerprints, markers, and PC v4.7. *Genome Res.* **10**: 1817–1825.
- SOLTIS, D. E., and P. S. SOLTIS, 1999 Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* **14**: 348–352.
- SONNHAMMER, E. L., and R. DURBIN, 1995 A dot matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GCl–10.
- 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.
- TAYLOR, M. S., C. P. PONTING and R. R. COPLEY, 2004 Occurrence and consequences of coding sequence insertions and deletions in mammalian genomes. *Genome Res.* **14**: 555–566.
- WAKELEY, J., and J. HEY, 1997 Estimating ancestral population parameters. *Genetics* **145**: 847–855.
- WENDEL, J. F., 2000 Genome evolution in polyploids. *Plant Mol. Biol.* **42**: 225–249.
- 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 the A and A^m genomes of wheat. *Plant Cell* **15**: 1187–1197.

Communicating editor: S. R. WESSLER