Title

Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat

Journal

Plant Molecular Biology, 54(1)

ISSN

0167-4412

Authors

Kong, Xiu-Ying
Gu, Yong Qiang
You, Frank M
et al.

Publication Date

2004

DOI

10.1023/b:plan.0000028768.21587.dc

Peer reviewed

Plant Molecular Biology **54:** 55–69, 2004.

Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat

Xiu-Ying Kong[1,2,+], Yong Qiang Gu[3,+,*], Frank M. You[4], Jorge Dubcovsky[4] and Olin D. Anderson[3]

[1]*Genetic Resources Conservation Program, University of California, Davis, CA 95616, USA;* [2]*Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences, Beijing 100081, China;* [3]*Western Regional Research Center, Agricultural Research Service, U.S. Dept. of Agriculture, 800 Buchanan Street, Albany, CA 94710, USA (*author for correspondence; e-mail ygu@pw.usda.gov);* [4]*Department of Agronomy and Range Sciences, University of California, Davis, CA 95616, USA;* [+]*these authors contributed equally to this work*

Received 27 August 2003; accepted in revised form 23 December 2003

Key words: bacterial artificial chromosome, evolution, glutenin, retrotransposons, sequence, wheat

Abstract

Two overlapping bacterial artificial chromosome (BAC) clones from the B genome of the tetraploid wheat *Triticum turgidum* were identified, each of which contains one of the two high-molecular-weight (HMW) glutenin genes, comprising the complex *Glu-B1* locus. The complete sequence (285 506 bp of DNA) of this chromosomal region was determined. The two paralogous x-type (*Glu-1-1*) and y-type (*Glu-1-2*) HMW-glutenin genes of the complex *Glu-B1* locus were found to be separated by ca. 168 000 bp instead of the 51 000 bp separation previously reported for the orthologous *Glu-D1* locus of *Aegilops tauschii*, the D-genome donor of hexaploid wheat. This difference in intergene spacing is due almost entirely to be the insertion of clusters of nested retrotransposons. Otherwise, the orientation and order of the HMW glutenins and adjacent genes were identical in the two genomes. A comparison of these orthologous regions indicates modes and patterns of sequence divergence, with implications for the overall Triticeae genome structure and evolution. A duplicate globulin gene, found 5′ of each HMW-glutenin gene, assists to tentatively define the original duplication event leading to the paralogous x- and y-type HMW-glutenin genes. The intergenic regions of the two loci are composed of different patterns and classes of retrotransposons, indicating that insertion times of these retroelements were after the divergence of the two wheat genomes. In addition, a putative receptor kinase gene near the y-type HMW-glutenin gene at the *Glu-B1* locus is likely active as it matches recently reported ESTs from germinating barley endosperm. The presence of four genes represented only in the Triticeae endosperm ESTs suggests an endosperm-specific chromosome domain.

Abbreviations: BAC, bacterial artificial chromosome; HMW, high-molecular-weight

Introduction

Bread wheats are hexaploid with the three homoeologous genomes designated A, B, and D. The A genome was contributed by *Triticum urartu* and the B genome by an unknown close relative of *Aegilops speltoides* (Dvorak, 1998). About 8000 years ago, the D genome was added from *A. tauschii* (syn. *T. tauschii*) to the AB genome tetraploid *T. turgidum* (Feldman *et al.*, 1995). Genetic and marker mapping have confirmed the high degree of conservation of gene content and order among the homoeologous chromosomes (Devos *et al.*, 1995; Nelson *et al.*, 1995). The lack of homoeologous pairing during meiosis is due to the suppression of such pairings by the *Ph1* and *Ph2* genes (McFadden and Sears, 1946). The mode of action of these genes is not known, but must involve discrimination among homoeologues. Knowledge of the DNA structure of such homoeologues would contribute both an understanding of the mechanisms of homologue and

non-homoeologue pairing, and give insights into the general evolution of the Triticeae genomes since their divergences. In addition, the duplication of the gene complement within such recent allopolyploid plants allows studies of the diversity with orthologous loci and insights into the evolutionary fate of duplicated loci.

Although colinearity of genes is touted as sufficient to consider the grasses a 'single genetic system' (Devos and Gale, 2000), and although a few studies have shown regions of high micro-colinearity between wheat and rice (Roberts *et al*., 1999; Yan *et al*., 2003), numerous exceptions have been recently discussed (Bennetzen, 2000; Gaut, 2002; Feuillet and Keller, 2002). Estimates suggest that the grass family originated roughly 77 million years ago. It has been well documented that grass genomes are evolutionary labile in many respects, including genome size and chromosome numbers (Arumuganathan and Earle, 1991; Gaut, 2002). Comparative sequence analyses of several orthologous regions of distantly related grass genomes, mainly from rice, maize, sorghum, barley, and wheat, has begun to reveal the potential complications of using rice as a model for map-based isolation of genes from other grass genomes (for review, see Feuillet and Keller, 2002). First, the sizes of orthologous intergenic regions vary significantly, up to several hundred kilobases, and the sequences in these regions are not conserved. Second, the occurrence of a diverse array of sequence rearrangements including gene inversions, duplications, deletions, and translocations often disturb the gene colinearity among grass genomes. Despite these limitations, these studies offer us a first view of the structure and evolution of grass genomes at the molecular level.

Comparisons of orthologous and paralogous regions of the wheat genomes will provide insights into the early stages of plant chromosome differentiation, and particularly into polyploid genome evolution. A recent comparison of orthologous regions of the A genome of tetraploid wheat and the A^m genome of diploid *T. monococcum* that diverged roughly a million years ago has shown a rapid genome divergence in the intergenic regions (Wicker *et al*., 2003a). A similar comparison between wheat and barley orthologous regions (10–14 million years divergence time) has shown no similarity between the intergenic regions of these two genera (Ramakrishna *et al*., 2002; Gu *et al*., 2003). No study has been reported on the comparison of the orthologous regions between the different genomes of polyploid wheat that diverged 2.5–4.5 million years ago (Huang *et al*., 2002) and that, therefore, represents an intermediate degree of divergence relative to the two comparisons described above. We present here a comparison between orthologous regions including the high-molecular-weight (HMW) glutenins (Shewry *et al*., 1992) of the wheat B and D genomes.

The *Glu-1* loci encoding the HMW-glutenin protein subunits are essential for wheat flour quality and the subject of genetic engineering for improving quality (Blechl and Anderson, 1996; Rooke *et al*., 1999; Vasil and Anderson, 1997). The complex *Glu-1* locus is composed of two paralogous x-type and y-type HMW-glutenin genes on each of the homoeologous group 1 chromosomes of wheat and near relatives. Thus, in the allopolyploid wheats, there are four and six HMW-glutenin genes in tetraploid (*T. turgidum*; durum) and hexaploid (*T. aestivum*; bread) wheats. Previously, the DNA sequence of a 102 kb region surrounding the *Glu-D1* locus was reported from *A. tauschii* (Anderson *et al*., 2003). The present report compares the *Glu-B1* locus from *T. turgidum* to the *Glu-D1* locus and to the partial *Glu-B1* locus sequences from hexaploid wheat (Anderson *et al*., 2002), and discusses the dynamics of locus evolution.

Materials and methods

BAC selection

We recently constructed and characterized a large insert BAC library of a line of durum wheat, *T. turgidum ssp. durum*, cv. Langdon, a tetraploid species consisting of A and B genomes (Cenci *et al*., 2003), where a segment of chromosome 6BS from the wild tetraploid *T. turgidum* ssp. *dicoccoides* has been substituted into the selected Langdon line. A set of 28 high-density filters printed with the full complement of the BAC library clones was screened with a HMW glutenin probe consisting of a 2.2 kb *Hin*dIII coding fragment from HMW-glutenin clone Dx5B (Anderson *et al*., 2002). A total of 23 positive BAC clones were identified from the screening. Purified BAC DNAs were digested with restriction enzymes *Hin*dIII and *Eco*RI, transferred onto nylon membranes, and hybridized with the HMW-glutenin probe. The Southern hybridization data indicated that there was no single BAC harboring both the x-type and y-type HMW-glutenin genes (see later discussion). To determine if sequences of BAC clones containing x-type HMW glutenin overlapped with those of y-type HMW-glutenin BAC

clones, *Hin*dIII-restricted fragment bands for each BAC clones were called and edited manually. Bands were collected and used to perform contig assembly in the program FPC 4.7 (Soderlund *et al*., 2000). Two overlapping BACs (107M7 and 1126E20) were selected to sequence the HMW-glutenin region spanning the *Glu-B1* locus.

BAC sequencing

The sequencing of the two *Triticum turgidum* BAC clones was carried out as described previously (Anderson *et al*., 2003). In brief, shotgun sequencing libraries for selected BAC clones were first constructed with randomly sheared BAC DNA isolated with a Large Construct Kit (Qiagen), in which an exonuclease treatment was included to remove *Escherichia coli* genomic DNA contamination. The sheared fragments were blunt-ended with a mung bean exonuclease treatment (BioLab) and dephosphorylated with a shrimp alkaline phosphatase treatment (USB). Single 'A' tails were added by incubating with *Taq* polymerase in the presence of dNTPs. The resulting DNA, with fragment sizes ranging from 2 to 5 kb, was gel-purified and then ligated into pCR4TOPO vectors with a TA cloning kit (Invitrogen). The ligation mixture was directly used to transform DH10B electroMAX cells (Invitrogen). The plasmid DNA from single colonies was purified using the PerfectPrep Direct Bind Kits (Eppendorf) for sequencing reactions. Inserts were sequenced from both directions with T7 and T3 primers by BigDye terminator chemistry (Applied Biosystems) on an ABI3700 capillary sequencer.

Sequence analysis

Sequence data were quality-analyzed with the PHRED software package (version 0.000925.c) with a cutoff of phred 15, below which all sequences were removed from further analysis. Vector sequence was removed using the CROSS_MATCH software (version 0.990329) with minmatch 5 and minscore 15. Sequences less than 300 bp in length, identified chimeric clones, and short segments of vector sequence frequently found on the ends of sequence files were all removed. The sequence data generated from each BAC clone was used to assemble continuous contigs with the Lasergene SeqMan module (DNAStar) (www.DNAStar.com). In this module, a high-stringency parameter for base calling and quality assessment was selected to generate the most accurate consensus sequence reads possible. For BAC clone 107M12, quality reads were generated by sequencing about 3000 inserts and then used for sequence assembling with a criteria of a 40 bp window size and a 96% match requirement. The initial assembling of the sequences resulted in 14 contigs with 13 gaps. Close inspection of the sequences at these gaps indicated that the failure to join the neighboring contigs together was often caused by the presence of G/C-rich areas in these regions, which lead to sequencing reaction termination. For gap filling, a primer-walking procedure was employed to sequence clones whose sequences are located at the ends of different contigs, and the dGTP BigDye terminator chemistry (Applied Biosystems) was used in the sequencing reaction to resolve the problem of potential secondary structures caused by high G/C content in the gap regions. For BAC clone 1126E20, sequences from 2000 shotgun library clones were used in the assembly, which gave rise to five contigs with four gaps. The gaps were filled by the same method described above.

For annotation, the assembled sequence of the two conjoined B genome BACs was compared with the previous annotations for the orthologous regions from the barley BAC (AY268139, Gu *et al*., 2003) and D genome BAC (AF497474, Anderson *et al*., 2003). In addition, a homology search was performed against NCBI nonredundant and dbEST databases with BLASTN, BLASTX, and TBLASTX algorithms. FGENESH (http://www.softberry.com/berry.phtml) and GENESCAN (http://genemark.mit.edu/GENESCAN.html) were used for gene prediction. Mobile elements were identified with NCBI BLAST searches, with DNAStar MegAlign dotplot analysis, and by comparison with the Triticeae Repeat Sequence Database (TREP) at the GrainGenes web site at (http://wheat.pw.usda.gov/ITMI/Repeats/). The definition and naming of new retrotransposons followed the system described by SanMiguel *et al*. (2002).

Results and discussion

Glu-B1 *BAC clones*

From the complete BAC library screening, 23 clones were identified with the HMW-glutenin probe. Southern analyses revealed that these clones could be divided into four groups based on their *Eco*RI and *Hin*dIII digestion patterns (Figure 1). As indicated in Figure 1, almost all the BAC clones contained one

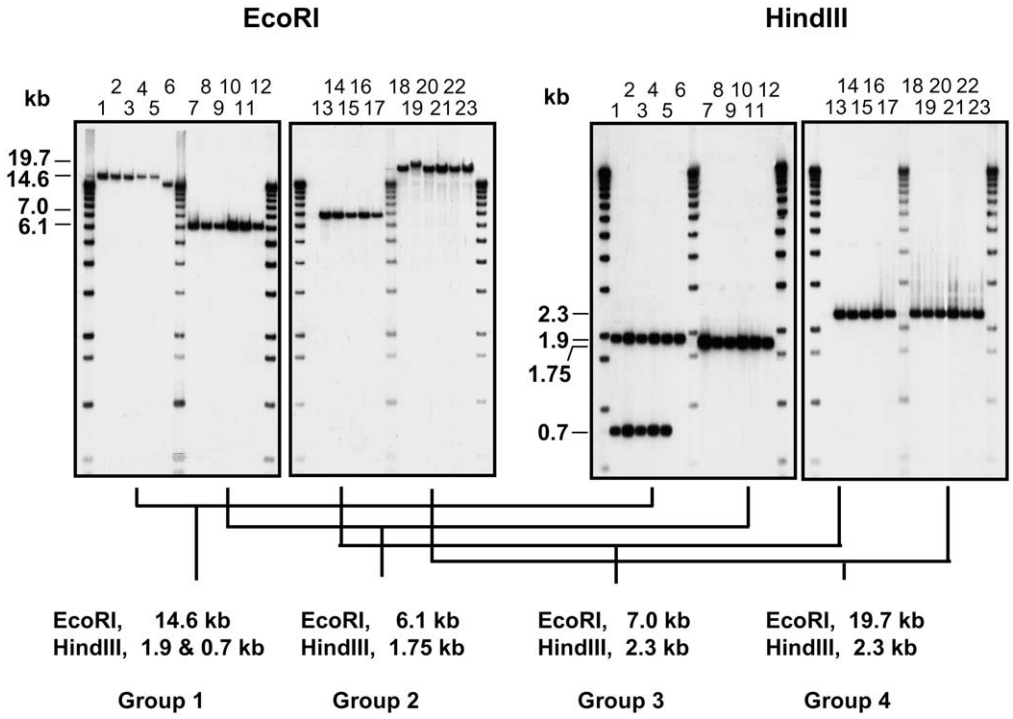

Figure 1. Southern analysis to size fragments containing HMW-glutenin sequences. A wheat tetraploid BAC library was screened for members containing HMW-glutenin sequences. Positive BACs were restricted separately with *Eco*RI and *Hin*dIII, the fragments separated by gel electrophoresis, and probed with a HMW-glutenin sequence. BAC clones are numerically named in the corresponding lanes. Group 1 contains BAC clones 1–6, Group 2 contains BAC clones 7–12, Group 3 contains BAC clones 13–17, and Group 4 contains BAC clones 18–23. Molecular markers were loaded in every seventh lane. The sizes of the hybridized HMW-glutenin bands are given.

of four different sizes of *Eco*RI fragments containing HMW-glutenin sequences, with the exception of two BAC clones, one (Clone 6) with smaller and the other (Clone 19) with larger than expected sizes. The smaller *Eco*RI band is likely due to only part of HMW-glutenin gene at one end of the BAC, as evidenced by the missing of the 0.7 kb *Hin*dIII fragment in the same clone (Figure 1). The larger *Eco*RI band might be caused by the loss of *Eco*RI site during the process of cloning. Since each *Glu-1* locus contains one x-type and one y-type HMW-glutenin gene (Shewry *et al*., 1992), and since the tetraploid library should include both the *Glu-A1* and *Glu-B1* loci, it was assumed that the four *Eco*RI fragments or five *Hin*dIII fragments in Figure 1 represent the two orthologous *Glu-1* loci. Since no BAC contained two fragments with both enzymes, and thus two HMW-glutenin genes, the spacing of the paralogous HMW-glutenin genes in both loci must be generally longer than possible to find within a single BAC. To determine if these BACs were of overlapping sequences, contig analysis was performed using agarose gel-separated restriction fragments and analysis with the FPC software for contig assembly (Soderlund *et al*., 2000). The analysis reported two contigs, containing 12 and 11 BACs, respectively. Contig 1 contained BACs of Group 1 and Group 4, while Contig 2 was composed of BACs of the other two groups (data not shown). Previously, it has been reported that the allelic genes from the B-genome of the hexaploid bread wheat cultivar Cheyenne are 15.8 kb (Halford *et al*., 1987) and 20.4 kb (Anderson *et al*., 2002), and the *Glu-A1* genes of cv Cheyenne are within 6.2 and 6.8 kb *Eco*RI fragments (Anderson *et al*., 2002). Based on the size of the four *Eco*RI HMW-glutenin fragments (6.1 kb, 7.0 kb, 14.6 kb, and 19.7 kb, Figure 1), and along with our previous results based on D genome substitution lines in Langdon (Cenci *et al*., 2003), Contig 1 with the 14.6 kb and 19.7 kb fragments represents the *Glu-B1* locus, and therefore Contig 2 represents the *Glu-A1* locus. Thus, Contig 1 was chosen for sequencing, and clone 14 (BAC library address 107M7) was selected for initial sequence since it overlaps with all the BAC clones containing the 14.6 kb *Eco*RI fragment and the 1.9

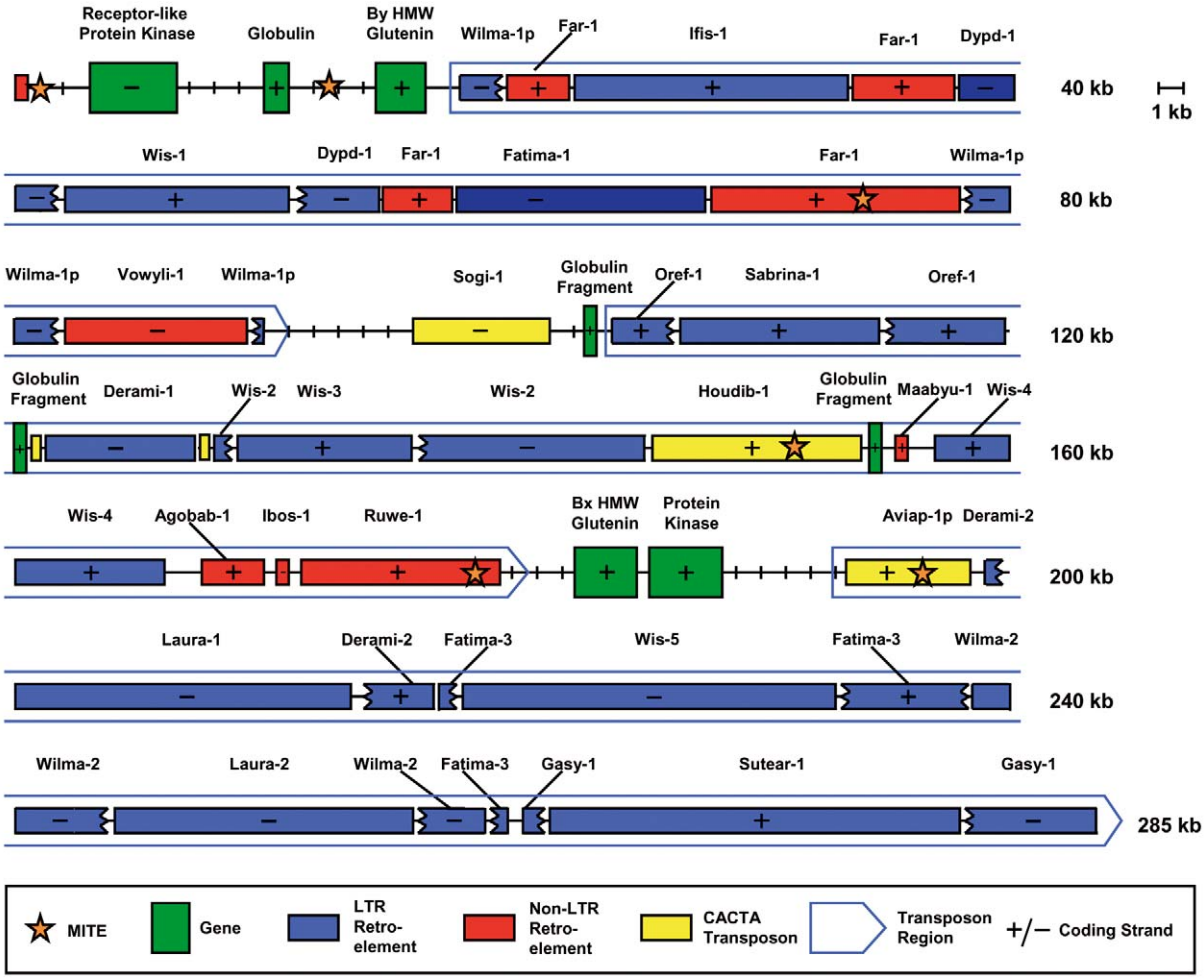

Figure 2. Organization of the HMW-glutenin *Glu-B1* locus. The 285 506 bp region of the *Glu-B1* locus in tetraploid wheat cv. Langdon is arrayed linearly and the major features indicated. Transposon clusters are indicated by blue lines enclosing those regions. The nomenclature for new transposable elements was according to the method described by SanMiguel *et al.* (2002). In the names of retroelements, P at the end means a partial retroelement with deletions of portions of LTR and/or coding region. Regions given the same name suggest that a retroelement sequence was interrupted, likely by the insertion of other retroelements.

and 0.7 kb *Hin*dIII fragments as indicated by the contig assembly. This BAC contained a large insert size (186 kb) as revealed by CHEF gel (data not shown).

The final assembled BAC 107M7 sequence covered 186 803 bp with 16× coverage. Once clone 107M7 was sequenced, the potentially overlapping BACs were sequenced at both ends and those sequences compared to the sequence of clone 14. One end of clones 2, 4, 5, and 6 were found to overlap with the sequence of clone 14. The end of clone 4 (BAC library address 1126E20) was at position 151 021 bp of the sequenced clone 14 and represented the smallest overlap (35 782 bp), and was chosen for sequencing to yield a contiguous DNA sequence between the two HMW-glutenin genes. The two overlapping BAC sequences were combined into one contiguous sequence of 285 506 bp and is deposited in GenBank as accession number AY368673.

Analysis of the Glu-B1 *region*

The general structure of the contiguous 285 kb *Glu-B1* region is shown in Figure 2. Six genes are identified as located in three gene-containing regions; each has three, two, or one gene. In the first region at the 5′ end, the three genes are clustered (gene cluster region), so are the two genes in the third region in the middle of the contiguous sequence (as shown). The globulin gene in the second region, originally belonging to the

same gene cluster with the x-type HMW-glutenin gene (see discussion later), was interrupted by multiple insertions of retrotransposons at two sites in the coding region, forming clusters of retrotransposons within the gene sequence (Figure 2 and see later discussion). In addition, five retrotransposons were identified between this globulin and the x-type HMW-glutenin genes. Consequently, the 80 kb space from this disrupted globulin gene to the x-type HMW glutenin is more like an intergenic region since retrotransposons account for 92% of the sequence. Sequences separating these gene-containing regions (intergenic region) are mainly composed of blocks of retrotransposons. We identified a total of 24 intact, solo, or partial retrotransposon elements (see later discussion), which comprise ca. 70% of the sequenced region. The level of retrotransposons in the *Glu-B1* is similar to those reported for the *Vrn* and *Lr10* regions of the diploid wheat *T. monococcum* (70%; Wicker *et al*., 2001; San-Miguel *et al*., 2002), but is much higher than that in the 106 kb *Lr21* region in the diploid wheat genome, *A. tauschii*, in which a 46 kb region containing seven coding sequences is retroelement-free (Brooks *et al*., 2002). Nevertheless, the gene organization in the *Glu-B* region generally supports the notion that genes are clustered in islands rather than distributed randomly in the genome and that retrotransposable elements are the dominant sequence type in the intergenic regions (Bennetzen, 2000).

Storage protein genes

Analysis of the contiguous sequence from the two BACs confirmed the presence of two HMW-glutenin genes. The start codons for these two genes were found at 14 638 bp and 182 711 bp of the 285 506 bp of contiguous sequence from two overlapping BACs for y- and x-type genes, respectively. The two genes are thus ca. 168 073 bp apart, significantly more than the 51 893 bp between the Dy-type and Dx-type HMW-glutenin genes of the D genome (Anderson *et al*., 2003). In both cases, most of the intervening sequences are composed of retrotransposon blocks, with the B-genome block about 3 times the size as in the D genome (see below). When the translated By and Bx peptide sequences were compared to its corresponding orthologue of the Dy and Dx in *A. tauschii*, high sequence conservation was found at the 5′ and 3′ non-repetitive regions, and the divergence observed in the middle repetitive regions were mostly caused by the difference in the number of the repetitive motifs (data not shown). Similar results have been reported when orthologous HMW glutenins were compared (Anderson *et al*., 2002).

The third seed-specific storage protein gene (*Glo-B2*) starts at position 9958 bp and is the orthologous globulin gene described in the wheat D genome (Figure 3; Anderson *et al*., 2003) and the barley genome (Gu *et al*., 2003). Outside of the Triticeae, this gene has closest matches to a single-copy rice 19 kDa α-globulin gene (Shorrosh *et al*., 1992; X63990; e^{-26}), and a rice 26 kDa seed globulin (Nakase *et al*., 1996; JC4784; e^{-26}). A blast search of ESTs in GenBank finds matches to a maize globulin-like EST (AF371278; $4e^{-27}$; Woo *et al*., 2001) and large numbers of exact to near exact Triticeae ESTs only from seed and endosperm libraries.

In the D genome, a region just 5′ to the Dx gene includes a 401 bp portion similar to the *Glo-D2* promoter, the start codon, and 15 bp into the signal peptide region, but the remainder of the sequence is suggested to be a series of rearrangements and short transpositions 5′ and 3′ to the remainder of the recognizable gene sequence (Anderson *et al*., 2003). In the B genome, remnants of an ancient globulin gene (ψ*Glo-B2*) are more intact, and Figure 3 shows the derived amino acid sequence of this gene after splicing out transposon insertions. As expected, the orthologous *Glo-B2* and *Glo-D2* sequences are more similar than to the paralogous inactive gene. There are 32 amino acid residue substitutions in comparison of the *Glo-B2* gene product with the paralogous amino acid sequence of the ψ*Glo-B2* gene, while a total of thirteen substitutions were identified in the orthologous *Glo-D2* gene product. In the ψ*Glo-B2* inactive globulin gene, the translation initiation codon (ATG) was mutated into a threonine codon (ACA) and two point mutations resulted in two in-frame stop codons in the coding regions. In addition, two insertion/deletion events (indels) occurred only in the inactive globulin gene (Figure 3). These sequence changes in the inactive globulin gene could be associated with the multiple insertions of retrotransposons, by which sequence rearrangements can occur (Feschotte *et al.*, 2002). Hence, we performed a detailed examination on the regions flanking the insertion sites. Sequence rearrangements were not found in these regions, suggesting that these sequence changes are independent of the retroelement insertions.

The inactive globulin gene is still detectable in the Triticeae genomes examined as yet, although in the case of the D genome the obliteration was near com-

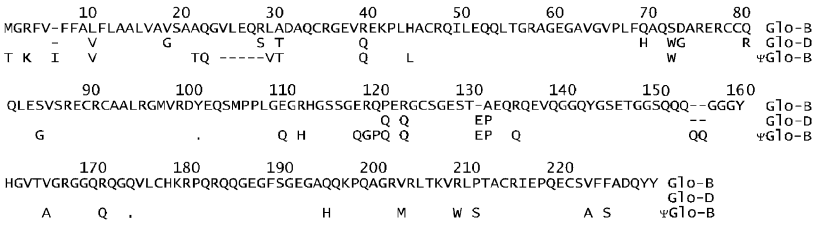

Figure 3. Comparison of globulin amino acid sequences in the B and D genomes. The B-genome globulin is compared to the D-genome globulin and a theoretical ancestral globulin of the B-genome globulin pseudogene. Differences are indicated by the different amino acid residues at those positions and deletions are indicated by a dash (-).

plete (Anderson *et al*., 2003). If the gene was further along in its disappearance, evidence for a multigene duplication would have been lost. Duplicate genes arise at a high rate, ca. 0.01 per gene per million years (Lynch and Conery, 2000). Most are silenced within a few million years with a calculated average half-life of 4 million years. The duplication of this globulin, along with the HMW glutenin, likely occurred after the divergence of the wheat and barley genomes, since the barley orthologous region contains only one copy each of the globulin and D-hordein (HMW-glutenin orthologue) gene (Gu *et al*., 2003). Given the recent report that rye possesses two copies of the HMW-glutenin genes (De Bustos *et al*., 2001), the duplication event might happen before the separation of the wheat and rye from their common ancestor, which is estimated to be about 7 million years (Huang *et al*., 2002).

Non-storage protein genes

From position 2989 bp to 6654 bp, the two gene-finding programs (GENESCAN and FGENESH) predicted the same number of exons that encode a leucine-rich repeat receptor kinase protein. This putative leucine-rich repeat receptor kinase (Figure 2) is similar to the A2 class (Wang *et al*., 1998) of the *Xa21* gene family of rice (Song *et al*., 1997) that confers resistance to the bacterial pathogen *Xanthomonas oryzae* pv. *oryzae*. In the available D-genome sequence, the leucine-rich repeat receptor kinase gene is at the end of the clone and it is not known if in addition to the leucine-rich repeat segment, the gene also contained a kinase domain (Anderson *et al*., 2003). A kinase domain was found to be present in the leucine-rich repeat receptor kinase gene identified in the orthologous D-hordein region in barley (Gu *et al*., 2003). Until recently, no evidence of the activity of this gene was found within Triticeae ESTs. A recent submission of barley ESTs from germinating seed endosperm has given two matches to two ESTs at significant confidence (BQ466381, CA015492; e^0), suggesting the leucine-rich repeat receptor kinase gene in barley is transcriptionally active. Whether or not the orthologous genes in any of the wheat genomes represent functional components remains to be determined.

From 185 490 bp, BLASTX reveals a sequence matching a rice serine/threonine protein kinase-like sequence (BAB89770; $6e^{-95}$) and several *Arabidopsis* serine/threonine protein kinases. A BLAST search against the Triticeace EST database failed to recover any sequences that showed best matches to the kinase gene sequence. The most significant match is to a barley EST from a developing spike library (BJ251833; e^{-12}). A BLASTX search with this EST, in turn, showed the best match to the kinase-like protein in rice (BAB89770; $2e^{-63}$). It is possible that current EST sequencing efforts have limitations in recovering transcripts that are rare or only expressed under certain conditions. The orthologous sequences of this putative protein kinase gene are present in the *A. tauschii* D genome (Anderson *et al*., 2003) and the barley D-hordein region (Gu *et al*., 2003). Sequence comparison revealed that the sequence in the B genome is 98% and 92% identical to those in the D and barley genomes, respectively. The high sequence conservation among different genomes is suggestive of the potential, but as yet unidentified, role of this protein kinase in plants.

In the 285 kb *Glu-B1* region, we identified a total of five genes that appear to be intact. Although the expression of the serine/threonine protein kinase gene has yet to be determined, the clustering of the other four consecutive genes from three different gene classes (receptor kinase, globulin, and HMW glutenin), all matched to ESTs from endosperm/seed cDNA libraries, suggesting an endosperm-specific chromosome domain. There are several examples of clusters of gene with related function in eukaryotes. For example, defense-related genes in barley and other

species (Hulbert *et al*., 2001; Wei *et al*., 2002). Other known examples include the multigene storage protein loci, such as the loci for the wheat LMW glutenins, γ-gliadins, and ω-gliadins on the group 1 chromosomes, and α-gliadins on the group 6 chromosomes. The apparent clustering of three different classes of endosperm active genes may have its roots in the potential coordinate control of multiple genes, possibly through the creation of chromosome domains competent for specific stimuli and/or developmental activation.

Repetitive DNAs

As shown in Figure 2, retrotransposable elements are the most dominant components in the sequenced region. Among the 24 retroelements identified, the WIS long-terminal-repeat (LTR) retrotransposons are the most abundant, accounting for about 20% of the contig sequence. WIS belongs to the *Ty1*-copia type of retrotransposable elements and is different from, but closely related to ANGELA, and BARE-1 retroelements (SanMiguel *et al*., 2002). Interestingly, the most abundant element in the orthologous HMW-glutenin region from the *A. tauschii* D genome is ANGELA, which makes up 25% of the 102 kb sequence. No WIS element was identified in the *Glu-D1* region of the D genome (Anderson *et al*., 2003). Much higher proportion of ANGELA elements were also found in the *Vrn* and *Lr10* regions in the *T. monococcum* A^m genome (Wicker *et al*., 2001; SanMiguel *et al*., 2002). It is possible that the wheat B genome contains more of the WIS-type subgroup, whereas the D and A^m genomes possess a majority of ANGELA-type elements. Alternatively, the high percentage of specific types of retroelements in these regions may represent a local activity of these retrotransposons, although this cannot be fully examined until large genomic segments of various regions from different wheat genomes have been sequenced and compared.

In addition to the six copies of WIS elements, there are twelve other LTR retroelements identified in the *Glu-B1* region, including three copies of FATIMA, two copies of each LAURA and WILMA. All 12 of these elements belong to the Ty3-gypsy retrotransposons. The LAURA retrotransposon, with 5 kb LTRs, was recently identified in the wheat A^m genome (Yan *et al*., 2003) and is related to the barley BAGY-1 element (Panstruga *et al*., 1998). Aside from these LTR retrotransposons, there are also identified seven non-LTR retrotransposons (*Far-1*, *Vowyli-1*, *Maabyu-1*, *Maaby-1*, *Agobab-1*, *Ibos-1*, and *Ruwe-1*). Non-

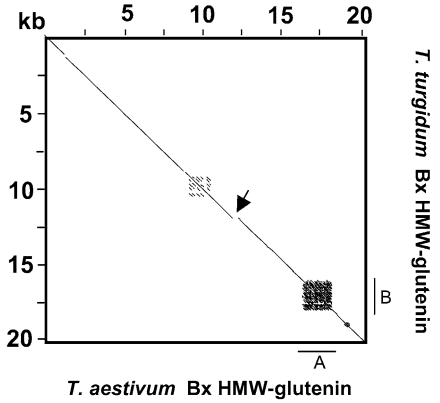

Figure 4. Comparison of two orthologous 20 kb regions including the Bx gene. A dot plot compares the tetraploid Bx region to the same region from hexaploid wheat. A comparison window of 30 bp and 70% identity was used. The light block in the middle of the diagonal is composed of uncharacterized repeat sequences, and the heavy block structure at the right end is caused by the repeat domain in the HMW-glutenin genes. The Bx HMW-glutenin gene regions of the hexaploid and tetraploid wheat are indicated by line A and B, respectively.

LTR retrotransposons are ubiquitous in plant genomes, however, they are not easy to identify due to the lack of distinct LTRs. The borders of the *Far-1* and *Vowyli-1* were identified because of their independent insertions into the *Wilma-1p* retroelement at different sites. All the other four non-LTR elements were inserted into the region between the inactive globulin and the x-type HMW-glutenin genes. Their insertion sites were delineated by comparing the orthologous regions of the B and D genomes.

Both LTR and non-LTR retrotransposons are classified as Class I repetitive elements, which replicate via an mRNA intermediate that is reverse transcribed into extrachromosomal DNA and integrated somewhere else in the genome. By contrast, Class II repetitive elements are DNA-mediated elements, whose movements involve excision from one site and integration into the different site of the genome. One Class II repetitive element is the CACTA transposon (Pereira *et al*., 1986). Recently, it has been shown that CACTA transposons belong to a diverse family of high-copy elements in Triticeae (Wicker *et al*., 2003b), implying that their roles in wheat genome evolution could be significant. In the *Glu-B1* region, we identified three CACTA elements (*Sogi-1*, *Houdib-1*, and *Aviap-1*), which constitute ca. 2.5% of the contig sequence.

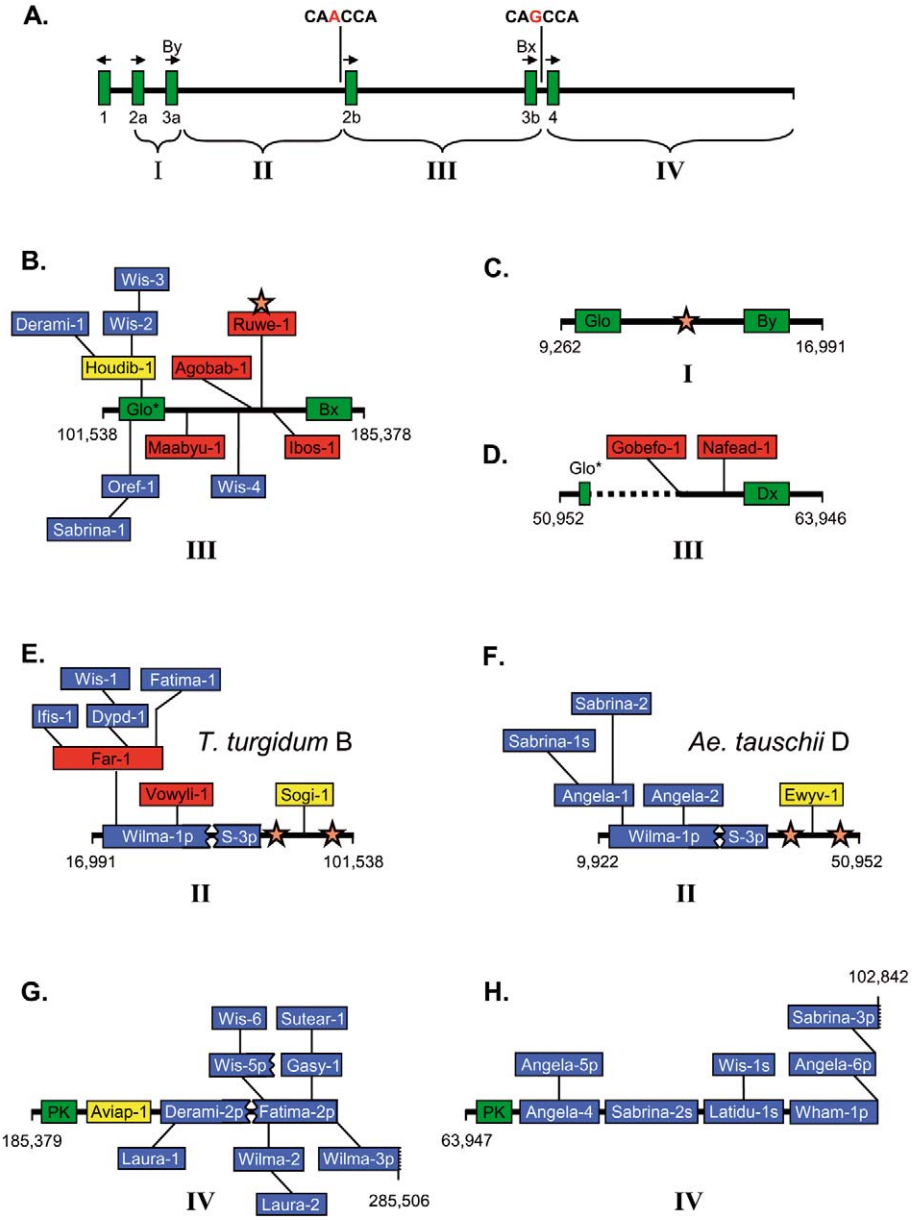

Figure 5. Conservation and divergences of intergenic regions flanking HMW-glutenin genes. The linear DNA sequence of the *Glu-1* region is diagrammed along with segments of the *Glu-B1* and *Glu-D1* regions with the proposed nesting organization of transposons. A. Linear order of *Glu-B1* region. Sections to be further discussed are indicated by Roman numerals (I–IV). Two hexanucleotide sequences of putative duplication borders at an ancestral *Glu-1* locus is given above. Arrows indicate directions of transcription. Genes are numbered as follows: 1, the receptor kinase gene; 2a and 2b, the two duplicated copies of globulin genes, 3a and 3b; the y-type and x-type copies of the HMW glutenin genes; and 4, the serine/threonine protein kinase gene. B and C. Paralogous regions containing globulin and HMW-glutenin genes are shown. D. Region containing the disrupted globulin gene and Dx (*Glu-D1-1*) genes are shown. Dash line stands for a deletion even that occurred in the globulin gene and the sequence before the Dy genes. E and F. Remnant of ancient retrotransposons 3′ to the y-type HMW-glutenin genes. The position of the WILMA (*Wilma-1p*) and SABRINA (*S-3p*, short for *Sabrina-3p*) retrotransposons existing prior to *Glu-B1* and *Glu-D1* divergence is shown along with retrotransposon insertions subsequent to divergence. Boxes above the *Wilma-1p* and *Sabrina-3p* elements indicated retrotransposon insertions at the points indicated by lines. Base positions at fragments ends are indicated. G and H. Divergence of sequences 3′ to the protein kinase gene. The approximate structure of DNA sequence from the protein kinase gene to the end of known sequence is indicated.

Comparing sequences from different germplasm sources of the B genome

Comparing closely related regions gives a picture of the earliest steps in sequence divergence. A 20.4 kb region of the B-genome was previously reported for the Bx7 gene from the bread wheat cultivar Cheyenne (Anderson *et al*., 2002), including 16 035 bp 5′ to the Bx-coding sequence. A pair-wise comparison of two available Bx region sequences is shown in Figure 4. The only major break in the diagonal is a 404 bp region probably caused by a transposon-related insertion into the hexaploid B genome since the region is bordered by an 8 bp duplication signature.

A total of 137 differences distinguish the two sequences. Of the 115 single-base changes, 76% are transitions, consistent with the 75% rate found in other organisms (Li, 1997). Of the 22 indels, eight occur in short homogenous runs of a single base, and five involve 3–9 bp duplications that could be remnants of insertions and subsequent movement of transposable elements.

Both the tetraploid and hexaploid wheats are relatively young polyploids: *T. turgidum* evolved roughly 400 000 years ago, whereas *T. aestivum* has only existed for 8000–10 000 years (Feldman *et al*., 1995). The sequence variations between the two B genomes in these different wheat species may reflect evolutionary events that have occurred since the formation of the hexaploid wheat. However, we have to consider that haplotypes are present in the B genomes of the tetraploid species prior to the polyploidization of the hexaploid wheat. These haplotypes will contribute to the variations observed in the different germplasm sources of the B genome.

Comparison of the paralogous regions in the Glu-B1 *sequence*

Genetic studies had shown that the genes encoding x-type and y-type HMW glutenins are tightly linked and indeed, there has never been a confirmed recombination event between them (Lafiandra *et al*., 1993). Sequence analysis of the *Glu-D1* region revealed that a region containing each of the globulin and HMW-glutenin gene was duplicated, giving rise to the presence of the two paralogous copies of the HMW-glutenin genes in the wheat genome. This duplication event could be responsible for the physical and biochemical properties of wheat flour that make it suitable for food processing. To further understand this important duplication event, we closely examine the

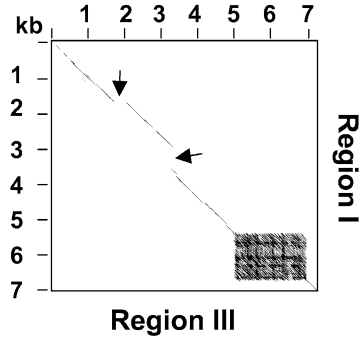

Figure 6. Comparison of reconstructed conserved sequences between the globulin and HMW-glutenin genes. The paralogous region between the ancestral globulin and HMW-glutenin genes is compared after splicing out recognizable transposon insertions. Arrows indicate unidentified events, like transposable element deletions or insertions, occurred in the compared regions. The dot plot used a window of 30 bp and a 70% minimum match.

paralogous sequences in the *Glu-B1* region and found that a 6 bp imperfect repeat defines the boundary of the paralogous sequence containing x-type HMW glutenin when compared to the region containing the y-type HMW-glutenin gene (Figure 5A). Although it is not established that the imperfect repeat is associated with the duplication event, this suggests that the region containing the x-type HMW glutenin was duplicated from that containing the y-type HMW glutenin. For the ease of comparisons between the paralogues in the same genome and among orthologues from different wheat genomes, several regions are defined in Figure 5A. Regions I and III are the paralogous sequences, Region II is the sequence separating I and III, and Region IV is the intergenic sequence after the serine/threonine protein kinase.

As indicated in Figure 5C, Region I is likely conserved since both globulin and HMW-glutenin genes are active and only one *Stowaway* MITE was inserted into this region. In contrast, the Region III contains 11 transposable element insertions, which added a total of 75 kb of sequence to the region compared to the paralogous region I (Figure 5B). This implies that these insertion events occurred after the duplication of the HMW-glutenin gene. To inspect other sequence rearrangements in these paralogous regions, we removed the sequences contributed by these repetitive DNAs and then conducted pair-wise dot-plot analysis as shown in Figure 6. After removal of the repetitive sequences, Regions I and III have a similar size of about 7.2 kb, suggesting its original duplicated size. The dot-plot analysis also revealed that there were two indels with sizes of over 400 bp in these

regions. Although indels are ubiquitous in plant genomes, the mechanism underlying these events has not been clearly understood. Recent studies suggested illegitimate recombination might be primarily responsible for the removal of nonessential DNA and is a major factor in the evolution of plant genomes (Devos *et al.*, 2002; Wicker *et al.*, 2003a). It has also been reported that many of the small indels between wheat and barley (60%) are flanked by direct repeats (Ramakrishna *et al.*, 2002).

As shown in Figure 6, the block structure is where the repetitive motif of the HMW-glutenin repeat domains find multiple matches. The coding regions of By (2163 bp) and Bx (2475 bp) HMW-glutenin genes vary in length by 312 bp. These differences are mainly caused by multiple indels that have occurred in the repetitive motifs of these genes. The indels in the repetitive motifs are usually small in size, ranging from 3 to 45 bp. Otherwise, the 5′ and 3′ non-repetitive regions are highly conserved at the nucleotide level (data not shown). It has been suggested that the repetitive motifs represent supervariable domains that lead to frequent sequence rearrangements among the HMW-glutenin genes (Anderson *et al.*, 2002). Irrespective of these changes, both genes are actively transcribed during endosperm development, suggesting a selection advantage to keep several copies of the gene expressed in polyploid wheats. It is also possible that the domestication of wheat continually selects for these gene variants controlling grain quality. In comparison with the HMW-glutenin genes, the two globulin genes contain fewer sequence changes at the DNA level: there are only three indels in the compared coding regions and the two sequences are about 88% identical when the retroelements inserted in the inactive globulin gene are removed and not taken into account.

Comparison to the orthologous Glu-D1 *region*

Several reports have shown that there is less conservation of intergenic DNA sequences than the gene-coding sequences when comparing several combinations of the major cereals – barley, maize, rice, sorghum, and diploid wheat (Chen *et al.*, 1997; Tikhonov *et al.*, 1999; Dubcovsky *et al.*, 2001; Ramakrishna *et al.*, 2002; SanMiguel *et al.*, 2002; Wicker *et al.*, 2003a; Gu *et al.*, 2003). Even among the Triticeae, intergenic sequences showed almost complete divergence when orthologous regions were compared: barley vs. diploid wheat, *T. monococcum* (SanMiguel *et al.*, 2002), and barley vs. diploid wheat, *A. tauschii* (Gu *et al.*, 2003). However, barley and wheat, which have diverged about 11 million years ago (Huang *et al.*, 2002), are among the most diverged members of the Triticeae. To examine more closely related Triticeae, we compared the complete sequence of the *Glu-B1* region with the previously reported *Glu-D1* region. Sequence conservation was observed in the gene-containing regions (see above discussion).

We then compared the intergenic regions, which were mainly composed of repetitive DNAs. The only repetitive sequences that are shared by both B and D genomes were found in Region II (Figure 5E and F). In this intergenic region of both genomes, a partial *Wilma* element (*Wilma-1p*) is immediately followed by a partial *Sabrina* element (*Sabrina-3p*). This structure was likely created by a deletion event that removed the sequences of the 3′ end of the *Wilma* and 5′ end of the *Sabrina* elements. The colinearity of these repetitive elements was confirmed by the exact sequences flanking each of the structures in the B and D genomes. In addition to these remnants of retrotransposons, a *Stowaway* MITE (Bureau and Wessler, 1994) was also found to be held in common, as identified by their same insertion sites. The presence of these elements in both the B and D genomes implies that their insertion times were before the divergence of the wheat ancestral genomes. Meanwhile, Figure 5E and F also indicate the insertion of several other repetitive elements into each Region II of the B and D genomes. Because of the lack of colinearity of these retroelements, acquisition of these elements likely occurred after the separation of the wheat genomes.

Region IV represents another intergenic region. There are ca. 40 and 100 kb sequences in this region from the *Glu-D1* and *Glu-B1* locus, respectively. No collinearity of the eight retrotransposons in the *Glu-D1* region was found with any of the eleven retroelements in the *Glu-B1* region (Figure 5G and H), again suggesting that the insertion of these elements into their present locations occurred after the separation of the wheat genomes. Taken together, repetitive DNAs, which comprise 80% of the total sequence, are primarily responsible for the significant sequence divergence of the orthologous HMW-glutenin regions from the wheat B and D genomes.

One of the striking findings from the comparison of the orthologous regions is the difference between the two inactive globulin genes in the B and D genomes. While a deletion event in the D genome obliterated most of the globulin coding sequence, insertions of five retrotransposons expanded the globulin gene in

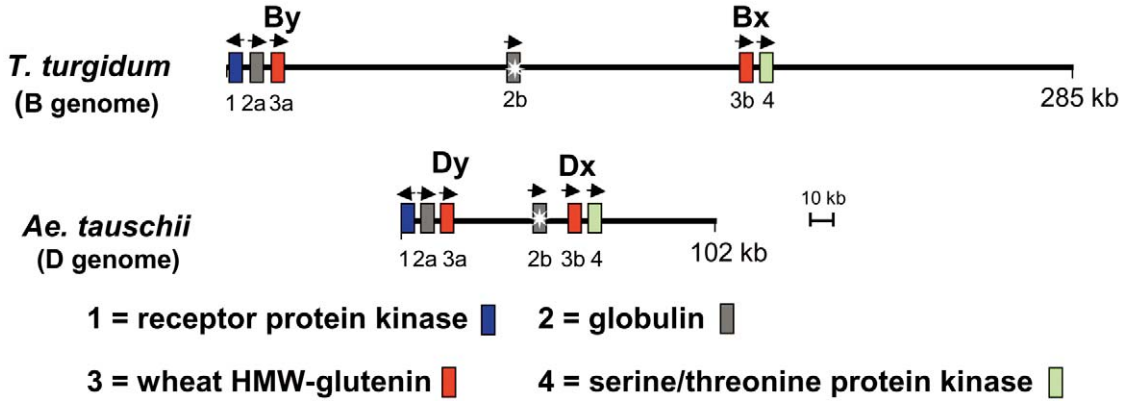

Figure 7. Comparison of gene order and separation of the HMW-glutenin genes within the *Glu-B1* (*T. turgidum*) and *Glu-D1* (*A. tauschii*) loci. Boxes indicate gene-coding regions. Genes are labeled with numerical numbers and the corresponding name of each gene is given as indicated. a and b means that the gene is duplicated in the sequenced region. Arrowheads indicate direction of potential transcription.

the B genome from 800 bp to over 40 000 bp (Figure 5B and D). The latter suggests that once a gene is tagged by a retrotransposable element, multiple insertions of retroelements will be possible in the gene region, which could lead to retrotransposon-nested structures or clustering as shown in Figure 5B. Consequently, the sequence originally belonging to a gene region now becomes more like an intergenic region. In addition, the observed difference in sequence rearrangement also implies that different mechanisms were involved in the inactivation of the globulin genes in the B and D genomes. However, it is also possible that the inactivation was initially caused by the same mechanism, probably because of the relaxation of selection on the second copy of the globulin gene, and that since then, evolution has reshaped these regions independently.

Gene spacing and synteny

An important aspect of comparisons among related genomes is to provide information on issues such as gene colinearity after significant evolutionary divergence. The order and spacing of genes within the *Glu-B1* and *Glu-D1* regions is diagrammed in Figure 7. The order of the genes is conserved, though the spacing is not, as also indicated in Figure 7. The two conserved islands with intact genes, the receptor kinase/globulin/y-type HMW glutenin and the x-type HMW glutenin/protein kinase islands, remain arranged closely. The main difference is the intergenic region that separates the two conserved gene islands. It is 51 kb in the D genome, while it is 168 kb in the B genome. Generally, retrotransposons appear to be relatively sequestered away from genes and mainly responsible for the difference in the intergenic regions (Kumar and Bennetzen, 1999). However, in the case of the inactive globulin gene region of the B genome as discussed above, the acquisition of retroelement sequences contributed significantly to the size difference of the intergenic regions between the two gene islands in the B and D genomes.

Gene density estimates for the Triticeae based on limited data from several reported BAC sequences vary greatly, on the order of 1 gene per 5–45 kb, values that are at least 12 times higher than would be expected for a random gene distribution. This could be partially explained by the fact that these sequenced regions were targeted towards those containing genes under study and thus may represent the gene-rich regions. However, high gene densities found in several sequenced regions appeared to be associated with large numbers of gene duplication events. The overall gene density in the barley 261 kb *Mla* locus region sequenced by Wei *et al*. (2002) is one gene per 8.1 kb, and 4.6 kb if only the gene-rich islands are considered. Similarly, Brooks *et al*. (2002) reported an overall gene density of one gene per 8.9 kb at the 106 kb *Lr21* locus in the *A. tauschii* D genome. Both the *Mla* and *Lr21* loci are located on the chromosome segments containing *R* gene clustering as a result of frequent *R* gene duplication (Ellis *et al*., 2000). In maize, 22 of the 23 members of the 22 kDa α-zein gene family are within a 168 kb segment of the maize genome and these family members are likely derived from several rounds of duplication events (Song *et al*., 2001). Gene duplication and subsequent divergence is probably one of the most important processes un-

derlying the evolution of gene function (Lynch and Conery, 2000; Bancroft, 2001). To date, the mechanism that determines regions of duplication is not well understood. Recently, Akhunov *et al*. (2003) reported that relative gene density and recombination rates increased with the relative distance away from the centromere. In addition, multi-gene loci tended to be more frequent in distal, high-recombination regions, suggesting a direct relationship between gene family multiplicity, recombination frequency, and gene density.

The presence of the conservation of gene order allows cross-species comparisons and would allow leveraging advances in one species to be applied more readily to other (Schmidt, 2000). Such microcolinearity is widespread in plants (Bancroft, 2001; Oh *et al*., 2002), although exceptions have often been noted in distantly related species (for a detailed review, Bennetzen, 2000; Gaut, 2002). Wheat genomes, which have radiated about 2.5–4.5 million years, are considered to be closely related (Huang *et al*., 2002). However, in the orthologous HMW-glutenin regions of the wheat D and B genomes, the intergenic regions are almost completely diverged, while the gene order and orientation are conserved. Of course, this comparison represents a small portion of the large wheat genome. Local variation along the chromosomes in the rate of genome evolution is possible. Recently, Wicker *et al*. (2003a) reported a comparison of part of the low-molecular-weight (LMW)-glutenin loci of the wheat tetraploid A and diploid A^m genomes, which diverged about 1 to 3 million years ago. It was found that differential amplification of the glutenin loci contributed to the rapid genome divergence observed in the compared regions. Another similar comparison including 340 kb of orthologous regions in the *Vrn2* region showed that 75% of the intergenic regions were non-orthologous retroelements between the tetraploid A and the diploid A^m genomes (J. Dubcovsky and P. SanMiguel, unpublished data). These results indicate a fast differentiation of the intergenic regions in the wheat species. It has been suggested that grass genomes including wheat evolved much more rapidly than we thought (Song *et al*., 2002; Wicker *et al*., 2003a). However, the evolution, structure, and function of wheat genome will be better understood when large numbers of orthologous regions from polyploid wheats and their diploid ancestral genomes are thoroughly compared.

Acknowledgements

We thank Roger Thilmony and Devin Coleman-Derr for critical reading of the manuscript. We also thank Dave Hane for assistance in the preparation of the figures. This research was partially supported by Research Grant US-3224-01R from BARD (United States-Israel Binational Agricultural Research and Development Fund).

References

Akhunov, E.D., Goodyear, A.W., Geng, S., Qi, L.L., Echalier, B., Gill, B.S., Miftahudin, Gustafson, J.P., Lazo, G., Chao, S., Anderson, O.D., Linkiewicz, A.M., Dubcovsky, J., Rota, M.L., Sorrells, M.E., Zhang, D., Nguyen, H.T., Kalavacharla, V., Hossain, K., Kianian, S.F., Peng, J., Lapitan, N.L., Gonzalez-Hernandez, J.L., Anderson, J.A., Choi, D.W., Close, T.J., Dilbirligi, M., Gill, K.S., Walker-Simmons, M.K., Steber, C., McGuire, P.E., Qualset, C.O. and Dvorak, J. 2003. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res. 13: 753–763.

Anderson, O.D. and Greene, F.C. 1989. The characterization and comparative analysis of high-molecular-weight-glutenin genes from genomes A and B of a hexaploid bread wheat. Theor. Appl. Genet. 77: 689–700.

Anderson, O.D., Larka, L., Christoffers, M.J., McCue, K.F. and Gustafson, J.P. 2002. Comparison of orthologous and paralogous DNA flanking the wheat high molecular weight glutenin genes: sequence conservation and divergence, transposon distribution, and matrix-attachment regions. Genome 45: 367–380.

Anderson, O.D., Rausch, C., Moullet, O. and Lagudah, E.S. 2003. The wheat D-genome HMW-glutenin locus: BAC sequencing, gene distribution, and retrotransposon clusters. Funct. Integr. Genomics 3: 56–68.

Arumuganathan, K. and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9: 208–218.

Bancroft, I. 2001. Duplicate and diverge: the evolution of plant genome microstructure. Trends Genet. 17: 89–93.

Bennetzen, J.L. 2000. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12: 1021–1029.

Blechl, A.E. and Anderson, O.D. 1996. Expression of a novel high-molecular-weight glutenin subunit gene in transgenic wheat. Nature Biotechnol. 14: 875–879.

Brooks, S.A., Huang L., Gill, B.S. and Fellers, J.P. 2002. Analysis of 106 kb of contiguous DNA sequence from the D genome of wheat reveals high gene density and a complex arrangement of genes related to disease resistance. Genome 45: 963–972.

Bureau, T.E. and Wessler, S.R. 1994. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6: 907–916.

Cenci, A., Chantret, N., Kong, X., Gu, Y.-Q., Anderson, O.D., Fahima, T., Distelfeld, A. and Dubcovsky, J. 2003. Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turdigum* ssp. *durum*). Theor. Appl. Genet. 107: 931–939.

Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.-S., Zhang, H., Wing, R.A. and Bennetzen, J.L. 1997. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc. Natl. Acad. Sci. USA 94: 3431–3435.

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

Devos, K.M, and Gale, M.D. 2000. Genome relationships: the grass model in current research. Plant Cell 12: 637–646.

Devos, K.M., Dubcovsky, J., Dvorak, J., Chinoy, C.N. and Gale, M.D. 1995 Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor. Appl. Genet. 91: 282–288.

Devos, K.M., Brown, J.K.M. and Bennetzen, J.L. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12: 1075–1079.

Dubcovsky, J., Ramakrishna, W., SanMiguel, P.J., Busso, C.S., Yan, L., Shiloff, B.A. and Bennetzen, J.L. 2001. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol. 125: 1342–1353.

Dvorak, J. 1998. Genome analysis in the *Triticum-Aegilops* alliance. In: A.E. Slinkard (Ed.) Proceedings of the 9th International Wheat Genetics Symposium (Saskatoon, Saskatchewan, Canada), University Extension Press, University of Saskatchewan, vol. 1, pp. 8–11.

Ellis, J., Dodds, P. and Pryor, T. 2000. Structure, function and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. 3: 278–284.

Feldman, M., Lupton, F.G.H. and Millers, T.E. 1995. Wheats. In: J. Smartt, N.W and Simmonds (Eds.) Evolution of Crops, Longman Scientific, London, pp. 184–192.

Feschotte, C., Jiang, N. and Wessler, S.R. 2002. Plant transposable elements: where genetics meets genomics. Nature Rev. Genet. 3: 329–341.

Feuillet, C. and Keller, B. 2002. Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann. Bot. 89: 3–10.

Gaut, B.S. 2002. Evolutionary dynamics of grass genomes. New Phytol. 154: 15–28.

Gu, Y.-Q., Anderson, O.D., Londeore, C., Kong, X., Chibbar, R.N. and Lazo, G.R. 2003. Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. Genome 46: 1087–1097.

Halford, N.G., Forde, J., Anderson, O.D., Greene, F.C. and Shewry, P.R. 1987. The nucleotide and deduced amino acid sequences of an HMW glutenin subunit gene from chromosome IB of bread wheat (*Triticum aestirum L*) and comparison with those of genes from chromosomes IA and ID. Theor. Appl. Genet. 75: 117–126.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R. and Gornicki, P. 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.

Hulbert, S., Webb, C.A., Smith, S.M. and Sun, Q. 2001. Resistance gene complexes: evolution and utilization. Annu. Rev. Phytopath. 39: 285–312.

Kumar, A. and Bennetzen, J.L. 1999. Plant retrotransposons. Annu. Rev. Genet. 33: 479–532.

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

Li, W.-H. 1997. Molecular Evolution. Sinauer, Sunderland, MA.

Lynch, M. and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.

MacFadden, E.S. and Sears, E.R. 1946. The artificial synthesis of *Triticum spelta*. Rec. Soc. Genet. Am. 13: 26–27.

Nakase, M., Hotta, H., Adachi, T., Aoki, N., Nakamura, R., Masumura, T., Tanaka, K. and Matsuda, T. 1996. Cloning of the rice seed alpha-globulin-encoding gene: sequence similarity of the 5′-flanking region to those of the genes encoding wheat high-molecular-weight glutenin and barley D hordein. Gene 170: 223–226.

Nelson, J.C., Sorrells, M.E., Van Denze, A.E., Lu, Y.H., Atkinson, M., Bernard, M., Leroy, P., Faris, J.D. and Anderson, J.A. 1995. Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 141: 721–731.

Oh, K.C., Hardeman, K., Ivanchenko, M.G., Ellard-Ivey, M., Nebenführ, A., White, T.J. and Lomax, T.L. 2002. Fine mapping in tomato using microsynteny with the *Arabidopsis* genome: the *Diageotropica* (*Dgt*) locus. Genome Biol. 3(9): research0049.1–0049.11.

Panstruga, R., Buschges, R., Piffanelli, P. and Schulze-Lefert, P. 1998. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucl. Acids Res. 26: 1056–1062.

Pereira, A., Cuypers, H., Gierl, A., Sommer, Z.S. and Saedler, H. 1986. Molecular analysis of the En/spm transposable element system of *Zea mays*. EMBO J. 5: 835–841.

Ramakrishna, W., Dubcovsky, J., Park, Y.-J., Busso, C., Emberton, J., SanMiguel, P. and Bennetzen, J.L. 2002. Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 162: 1389–1499.

Roberts, M.A., Reader, S.M., Dalgliesh, C., Miller, T.E., Foote, T.N., Fish, L.J., Snape, J.W. and Moore, G. 1999. Induction and characterization of *Ph1* wheat mutants. Genetics 153: 1909–1918.

Rooke, L., Bekes, F., Fido, R., Barro, F., Gras, P., Tatham, A.S., Barcelo, P.A., Lazzeri, P. and Shewry, P.R. 1999. Overexpresssion of a gluten protein in transgenic wheat results in greatly increased dough strength. J. Cereal Sci. 30: 115–120.

SanMiguel, P., Ramakrishna, W., Bennetzen, J.L., Busso, C.S. and Dubcovsky, J. 2002. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct. Integr. Genomics 2: 70–80.

Schmidt, R. 2000. Synteny: recent advances and future prospects. Curr. Opin. Plant Biol. 3: 97–102.

Shewry, P.R., Halford, N.G. and Tatham, A.S. 1992. High molecular weight subunits of wheat glutenin. J. Cereal Sci. 15: 105–120.

Shorrosh, B.S., Wen, L., Kuo-Chang, Z., Huang, J.-K., Pan, J.S., Hermodson, M.A., Tanaka, K., Muthukrishnan, S. and Reeck, G.R. 1992. A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice. Plant Mol. Biol. 18: 151–154.

Soderlund, C., Humphreys, S., Dunham, A. and French, I. 2000. Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10: 1772–1787.

Song, R., Llaca, V., Linton, E. and Messing, J. 2001. Sequence, regulation, and evolution of the maize 22-kD alpha zein gene family. Genome Res. 11: 1817–1825.

Song, R., Llaca, V. and Messing, J. 2002. Mosaic organization of orthologous sequences in grass genomes. Genome Res. 12: 1549–1555.

Song, W.Y., Pi, L.Y., Wang, G.L., Gardner, J., Holsten, T. and Ronald, P.C. 1997. Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9: 1279–1287.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.D., Bennetzen, J.L. and Avramova, Z. 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. USA 96: 7409–7414.

Vasil, I.K. and Anderson, O.D. 1997. Genetic engineering of wheat gluten. Trends Plant Sci. 8: 292–297.

Wang, G.-L., Ruan, D.-L., Song, W.-Y., Sideris, S., Chen, L., Pi, L.-Y., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B.S., Whalen, M.C. and Ronald, P.C. 1998. *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10: 765–779.

Wei, F., Wing, R.A. and Wise, R.P. 2002. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 14: 1903–1917.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E. and Keller, B. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J. 26: 307–316.

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.D., Dubcovsky, J. and Keller, B. 2003a. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15: 1186–1197.

Wicker, T., Guyot, R., Yahiaoui, N. and Keller, B. 2003b. CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol. 132: 52–63.

Woo, Y.M., Hu, D.W., Larkins, B.A. and Jung, R. 2001. Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. Plant Cell 13: 2297–2317.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T. and Dubcovsky, J. 2003. Positional cloning of the wheat vernalization gene *VRN1*. Proc. Natl. Acad. Sci. USA 100: 6263–6268.