Title

Genomic organization of the complex alpha-gliadin gene loci in wheat.

Journal

TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 109(3)

ISSN

0040-5752

Authors

Gu, Yong Qiang
Crossman, Curt
Kong, Xiuying
et al.

Publication Date

2004-08-01

DOI

10.1007/s00122-004-1672-2

Peer reviewed

Theor Appl Genet (2004) 109: 648–657
DOI 10.1007/s00122-004-1672-2

ORIGINAL PAPER

Yong Qiang Gu · Curt Crossman · Xiuying Kong · Mingcheng Luo · Frank M. You · Devin Coleman-Derr · Jorge Dubcovsky · Olin D. Anderson

Genomic organization of the complex α-gliadin gene loci in wheat

Received: 4 November 2003 / Accepted: 19 March 2004 / Published online: 22 April 2004

Abstract To better understand the molecular evolution of the large α-gliadin gene family, a half-million bacterial artificial chromosome (BAC) library clones from tetraploid durum wheat, *Triticum turgidum* ssp. *durum* (2*n*=4*x*=28, genome AB), were screened for large genomic segments carrying the α-gliadin genes of the *Gli-2* loci on the group 6 homoeologous chromosomes. The resulting 220 positive BAC clones—each containing between one and four copies of α-gliadin sequences—were fingerprinted for contig assembly to produce contiguous chromosomal regions covering the *Gli-2* loci. While contigs consisting of as many as 21 BAC clones and containing up to 17 α-gliadin genes were formed, many BAC clones remained as singletons. The accuracy of the order of BAC clones in the contigs was verified by Southern hybridization analysis of the BAC fingerprints using an α-gliadin probe. These results indicate that α-gliadin genes are not evenly dispersed in the *Gli-2* locus regions. Hybridization of these BACs with probes for long terminal repeat retrotransposons was used to determine the abundance and distribution of repetitive DNA in this region. Sequencing of BAC ends indicated that 70% of the sequences were significantly similar to different classes of retrotransposons, suggesting that these elements are abundant in this region. Several mechanisms underlying the dynamic evolution of the *Gli-2* loci are discussed.

Communicated by P. Langridge

Y. Q. Gu (✉) · C. Crossman · D. Coleman-Derr · O. D. Anderson
United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center,
800 Buchanan Street,
Albany, CA 94710, USA
e-mail: ygu@pw.usda.gov
Tel.: +1-510-5595732
Fax: +1-510-5595818

X. Kong
Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences,
Beijing, 100081, China

M. Luo · F. M. You · J. Dubcovsky
Department of Agronomy and Range Science, University of California,
Davis, CA 95616, USA

Introduction

Wheat grain, of which approximately 9–15% (dry weight) is protein, is a major source of energy and nutrition in the human diet. The majority of the seed storage proteins are prolamins, which are stored in the starchy endosperm and account for more than half of the proteins present in the mature grain. In addition, prolamins are the major components of gluten, the properties of which determine the bread-making quality of wheat flour. Wheat gluten is composed of two major groups fo proteins, gliadins and glutenins. The glutenins, which include the high molecular weight (HMW) glutenins and low molecular weight (LMW) glutenins, form large protein polymers connected by intermolecular disulphide bonds, whereas the gliadins are mainly monomeric, having only intramolecular disulphide bonds. Gliadins can be subdivided into several classes (α-, γ-, and ω-gliadins) based on their mobility in SDS-PAGE gels. Among them, the α-gliadins are the most abundant, comprising 15–30% of the wheat seed proteins. Therefore, the α-gliadins are among the most consumed proteins by humans. Unfortunately, they also seem to be the major initiators of celiac disease, which effects as many as 1 in 300 people (Shewry et al. 1992).

Like other wheat prolamins, the α-gliadin proteins are encoded by members of a multigene family. Examination of the number of α-gliadin proteins synthesized revealed at least 16 major α-gliadin spots by 2-D PAGE of protein extracts from cv. Cheyenne seed (Lafiandra et al. 1984). However, this number is considerably less than the gene copy number (about 150) estimated on the basis of genomic Southern analyses (Anderson et al. 1997). This can be partially explained by the assumption that some peptides encoded by the different gene family members comigrate during electrophoresis, resulting in single-protein bands or spots and by the fact that many of the

family members are pseudogenes. Anderson and Greene (1997) analyzed 27 α-gliadin gene sequences at both DNA and protein sequence levels. It was found that the high base substitution rate, especially C→T substitution, contributed to the generation of stop codons, causing inactivation of 50% of the α-gliadin genes.

The number of α-gliadin genes in different wheat cultivars and ancestral species varies significantly, ranging from 25 to 150 copies (Herberd et al. 1985; Reeves and Okita 1987; Anderson et al. 1997). These differences are probably caused by duplication and deletion of chromosome segments, such as a wheat line lacking an entire cluster of the α-gliadin genes (D'Ovidio et al. 1991). Genetically, the α-gliadin genes are located at the *Gli-2* loci on the group 6 homoeologous chromosomes, whereas all the other wheat prolamin genes are found on the group 1 chromosomes (Payne 1987; D'Ovidio et al. 1992). These genetic studies suggest that the gene family members of wheat prolamins are clustered into chromosome segments. Our previous search for α-gliadin genes by screening several λ-phage libraries of wheat genomic DNA recovered 54 distinct clones, but none of these clones contained more than one copy of the α-gliadin genes (Anderson et al. 1997). One possible explanation could be that the distance separating α-gliadin genes is larger than the maximum insert size a λ-phage clone can accommodate. Therefore, the structural organization of these gene families with respect to distribution of the genes, the sizes of gliadin-containing chromosome segments, and their sequence composition remains unclear.

Bacterial artificial chromosome (BAC) libraries of large genomic DNA fragments are valuable tools for genomic research, including the development of physical maps of entire genomes, positional cloning of candidate genes with important biological or agronomic traits, and genomic structural analysis and comparison of specific regions of related species. Several wheat BAC libraries have been recently constructed (Lijavetzky et al. 1999; Moullet et al. 1999; Ma et al. 2000; Cenci et al. 2003). These libraries are also important resources for studying genome complexity and the evolution of wheat storage-protein genes. In this report, we screened a BAC library of the tetraploid wheat, *Triticum turgidum*, for large chromosome segments containing the α-gliadin gene(s). Contigs composed of overlapping BAC clones were assembled to study the structural organization of the α-gliadin gene family. Based on the contig result, the distribution of the α-gliadin genes in the *Gli-2* regions was examined, and gene spacing was estimated. In addition, the sequence composition of these contigs was also investigated to understand the molecular evolution of the prolamin gene loci in wheat.

Materials and methods

BAC library screening

A large insert BAC library was previously constructed and characterized for a line of durum wheat, *T. turgidum* ssp. *durum*, cv. Langdon, a tetraploid species consisting of A and B genomes (Cenci et al. 2003). The selected line contains a 30-cM segment of chromosome 6BS from the wild tetraploid *T. turgidum* ssp. *dicoccoides* carrying a gene for high grain-protein content (Joppa et al. 1997; Khan et al. 2000). This library contains half a million BAC clones with an average insert of 130 kb, providing a coverage of 5.1× genome equivalents for each of the two genomes. A ^{32}P-labeled probe, Y16 K, containing the complete α-gliadin coding sequence (Anderson et al. 1997) was used to screen a set of 28 high-density filters printed with the entire BAC library clones. This hybridization screening was performed according to methods previously described (Gu et al. 2003).

BAC fingerprinting and contig assembly

Individual α-gliadin positive BAC clones were inoculated into 48-deep well plates containing 2 ml LB medium plus 12.5 μg/ml chloramphenicol and grown in a HIGro shaker (Genemachines, Menlo Park, Calif., USA) at 37°C and 450 rpm for 16 h. BAC DNA was isolated and purified using a Qiagen 48-well BAC purification kit (Valencia, Calif., USA). For agarose gel-based DNA fingerprinting, the method described by Marra et al (1997) was employed with some modification. In brief, approximately 100 ng DNA in a 5-μl aliquot was digested with the restriction enzyme *Hin*dIII in a 10-μl reaction volume. After 5 h incubation at 37°C, 2 μl 6× loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 15% Ficol) was added to each reaction, and 2 μl of this mixture of each reaction was loaded into a 1% agarose gel with 61 wells, each having dimensions of 2 mm wide × 1 mm thick × 3 mm deep. A 2-μl mixture of marker DNA as described by Marra et al. (1997) was loaded into every seventh lane. DNA fragments were separated by electrophoresis at 60 V in circulating 1× TAE buffer at 14°C for 16 h. The gel was stained in 500 ml of a 1:10,000 dilution of Vistra Green (Molecular Probe, Eugene, Ore., USA) in 1× TAE for 40 min. Gel images were captured using a VersoDoc imaging system (BioRad, Hercules, Calif., USA) and exported as tagged image files. Fingerprints were edited using the Image 3.8 module of the finger printed contig (FPC) package (Soderlund et al. 2000). BAC contigs were assembled using the FPC, version 4.7 program.

Southern hybridization

For Southern hybridization of BAC DNA, the *Hin*dIII-digested fingerprint gels were blotted onto Hybond-N+ membranes (Amersham, Piscataway, N.J., USA) after the gel images were documented. Probes were either labeled with ^{32}P-isotope using the DECAprimeII DNA labeling Kit (Ambion, Austin, Tex., USA) or DIG labeled using the PCR DIG probe synthesis Kit (Roche, Indianapolis, Ind., USA). After hybridization, the images were detected with autoradiograph or the DIG Luminescent detection kit for nucleic acids, respectively.

To detect the α-gliadin genes in each BAC clone, two probes were used—one derived from the Y16 K and the other containing the α-gliadin coding sequence from an EST clone (BE422727)—that showed sequence divergence with the Y16 K (see discussion later). The size of each α-gliadin fragment was determined using the marker DNA as a standard with the Quantity One program (BioRad).

To examine retrotransposable elements in the BAC clones, probes specific for the long terminal repeat (LTR) retrotransposons *Wis*, *Wham*, and *Sabrina* were generated by PCR amplification of their unique LTR sequences. Primer pairs for the *Wis* LTR were Wis-F AAGCCTCGTTTTCAGTAA3′ and Wis-R TTGTCTCTAATGCCTAACC; for the *Wham* LTR, Wham-F AGCAATAGCAACGGAAAAGTAAAT and Wham-R ATGGTTGATGATGACGATGGCGAC; and for the *Sabrina* LTR, Sabrina-F TGAGCACTGCGTTGGTTTTC and Sabrina-R CAATCTTGCGGTGTCCTTTTC. A probe for the integrase sequence from the integrase coding region of the *Wis* retrotransposon was amplified

using the primer pairs of integrase-F AAGCCTCGTTTTCAGTAA and integrase-R TTGTCTCTAATGCCTAACC. The PCR templates for each probe were the plasmid DNA containing the corresponding retrotransposon LTR sequence.

BAC end sequencing and sequence analysis

BAC end sequencing of the α-gliadin positive clones was performed using DNA purified with Eppenddorf's BAC Purification Kit (Boulder, Colo., USA). Inserts were sequenced from both directions with T7 and SP6 primers using BigDye termination chemistry (Applied Biosystems, Foster City, Calif., USA) and run on an ABI3100 capillary sequencer. The sequence data were quality analyzed using the PHRED software package (version 0.000925.c). Sequence data with a phred score below 20 were removed. Quality sequences were analyzed by BLAST searches against the National Center for Biological Information (NCBI) database using BLASTN, BLASTX, and TLABST algorithms (Altschul et al. 1997). Repetitive sequences were identified by blast searches against the Triticeae Repeat Sequence Database (TREP) at the GrainGenes Web site at http://wheat.pw.usda.gov/ITMI/Repeats/

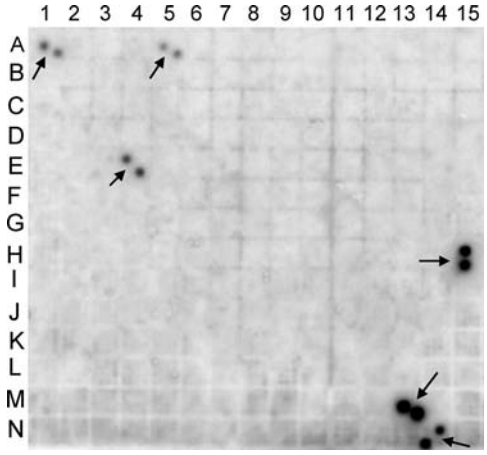

Fig. 1 Example of bacterial artificial chromosome (BAC) library hybridization. A portion of an autoradiograph of a BAC library filter is shown. The filter was hybridized with a ^{32}P-labeled Y16 K probe, which contains the coding region of the α-gliadin gene. Five positive clones—each double printed onto the filter—are indicated with *arrows*

Results

Durum wheat BAC library screening

The half-million clone BAC library of durum wheat provides an expected 5.1× genome equivalents for each of the two genomes, allowing a >99% probability of recovering any specific sequences from each of the two genomes (Cenci et al. 2003). The experimental validation of the genome coverage by hybridizations with probes of single copy sequences per genome yielded an average of 5.8 positive BAC clones (Cenci et al. 2003). The screening with the α-gliadin probe resulted in the selection of 220 BAC clones. Based on this result and the 5.1× genome coverage, it could be estimated that at least 43 α-gliadin genes are present in the tetraploid genome. This is probably an underestimate because single BAC clones might contain more than one α-gliadin gene (see results below).

Figure 1 shows that the positive BAC clones displayed differential hybridization signal intensities in the autoradiograph. Although a number of factors can cause variations in signal intensity—such as uneven hybridization across the BAC filter—the two most likely reasons are that these positive BAC clones contain various copies of the α-gliadin genes due to the clustering of the gene family members, and that the dynamic evolution of storage protein genes has resulted in extensive sequence variation among gene family members (Anderson et al. 1997).

Southern hybridization of α-gliadin positive clones

To determine if the positive BAC clones contained multiple copies of tightly linked α-gliadin genes, BAC DNA was digested with the *Hin*dIII restriction enzyme and hybridized with the α-gliadin probe (Fig. 2). Previously, we used the *Hin*dIII enzyme to estimate the copy number of α-gliadin genes in different wheat genomes (Anderson et al. 1997). As shown in Fig. 2, the number of *Hin*dIII-digested fragments recognized by the probe in each clone range from one to four, with total of 6.3% of the clones containing four bands, 13.1% three bands, 41.7% two bands, and 38.9% one band. To further assess if the fragment number in each lane reflects the copy number of the α-gliadin genes in each respective clone, the BAC DNA was also digested with *Eco*RI and hybridized with the same probe, Y16 K. In most cases, the number of the hybridized *Hin*dIII fragments is the same as that of the *Eco*RI fragments (data not shown), suggesting that each fragment represents one copy of an α-gliadin gene. Therefore, the observation that BAC clones contain various copy numbers of α-gliadin genes provides an explanation for the differential hybridization intensities of positive BAC clones in our library screening. It was also noticed that the signal intensities of the hybridized bands in these BAC clones also showed considerable variation (Fig. 2). Such variation could be explained if sequences have diverged considerably among the members of the α-gliadin family and the probe hybridized better to its closely related members. To test this possibility, we searched the wheat EST database for α-gliadin sequences with considerable sequence variation from the Y16 K probe. The wheat EST clone BE422727 differs from the Y16 K probe by five deletions/insertions of 3–46 bp in size plus 86 single-nucleotide substitutions in the approximate 0.85-kb probe sequence regions (data not shown). When the blots were hybridized with the EST probe, it was found that several *Hin*dIII fragments that gave strong signal intensities with the Y16 K probe had much reduced intensities, while many of those having weak intensities with the Y16 K probe had considerably greater intensities with the EST probe (Fig. 2b). This result is consistent with the previous phylogenetic analyses based on the available α-gliadin sequence data (Anderson et al. 1997b). More-

over, comparison of the two hybridization data sets allows us to confirm that some of the weak bands detected with either probe are members of the α-gliadin gene family.

DNA fingerprinting and contig assembly of the α-gliadin BAC clones

Given the large number of α-gliadin positive BAC clones, along with the observation that 61% of the clones seem to contain more than one copy of the gene, it is obvious that the *Gli-2* loci are complex regions that harbor a cluster of closely related genes. In order to facilitate investigation of the physical structure of the gene clusters and to establish the order and spacing of the α-gliadin genes, we attempted to reconstruct the genomic context of this gene family by fingerprinting and contig assembly of the α-gliadin positive BAC clones. The *Hin*dIII digestion generated 20–40 restriction fragments for each BAC clone (data not shown). Contigs were first assembled using the Sulston score of 10^{-5} and fixed tolerance of 7. At this assembling stringency, 34 contigs (with more than one BAC clone) were formed from the α-gliadin clones. There were 12 contigs with at least four or more BAC clones, and the largest contig, Contig6, contained 21 BAC clones (Table 1). However, 69 BAC clones remained as singletons.

The contig results suggest that some α-gliadin genes are clustered into gliadin-rich regions, while others are more distant from one another in the *Gli-2* regions. To verify the accuracy of the contigs, we attempted to determine if the order of the clones in the contigs agreed with the BAC Southern hybridization data, assuming that overlapping regions of the BAC clones contained shared α-gliadin gene sequences. Figure 3 is an example of such analysis, in which we first determined the sizes of the α-gliadin fragments in each BAC and then aligned the shared fragments in the overlapping clones based on the order of the clones in the contig (Fig. 3b). Such alignment, in turn, enabled us to anchor each α-gliadin gene at an approximate position in the individual clones (Fig. 3a). Figure 3 shows that most of the overlapping regions contained one or two α-gliadin genes, and clones 1199L22 and 118I23 likely shared three α-gliadin genes. Meanwhile, it was expected that some of the overlapping regions would contain no α-gliadin genes. We found that the overlapping regions between 1274D05 and 713K03 and between 950P05 and 1058I11 contain sequences that were not recognized by the α-gliadin probe, suggesting that only the intergenic regions were shared by these clones. Using anchored α-gliadin genes, we determined that there were a total of 17 α-gliadin genes in Contig6. Using this approach, we examined the rest of the contigs by aligning the *Hin*dIII-digested α-gliadin fragments in the overlapping clones and then assessed the gene copy number in each contig. In all the cases, the alignment of the *Hin*dIII-digested α-gliadin fragments in the overlapping clones was consistent with the order of the clones in each contig.

The Sulston score (or cutoff value) used in the contig assembly determines the maximum allowable probability of a match between any two clones. The smaller the Sulston score value, the lower the probability that the match has arisen by chance and the more extensive the overlap between any two clones. To determine the optimal Sulston score, we also performed the assembly at Sulston scores with higher (10^{-4}) and lower (10^{-6}) values than 10^{-5}. Assembly at 10^{-4} and 10^{-6} gave 36 and 30 contigs, respectively. The contigs consisting of at least four BAC clones are listed in Table 1. The number of these contigs is the same in each assembly. A comparison of contigs assembled at different scores revealed that while many contigs assembled at 10^{-5} were found to have the same or similar contigs assembled at different scores, some showed considerable difference. For example, at the lowest assembling stringency (10^{-4}), the largest contig contains 41 clones, while at the highest stringency (10^{-6}), the

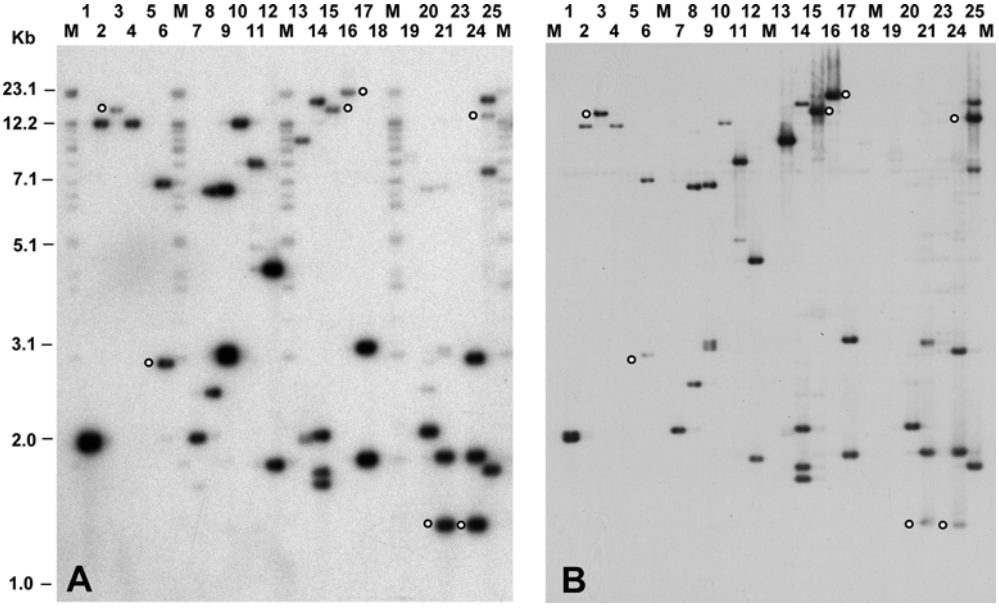

Fig. 2a, b Southern hybridization of digested BAC clones with α-gliadin probes. BAC DNA was digested with the *Hin*dIII restriction enzyme. The restriction fragments were separated by agarose gel electrophoresis. Duplicate gels were blotted onto nitrocellulose membranes and hybridized with either the Y16K probe (**a**) or the wheat α-gliadin EST probe (**b**). Lanes with BAC DNA are numbered *1–25*. Negative controls of BAC clones containing no α-gliadin genes are in *lanes 5*, *18*, and *19*. Lanes loaded with marker DNA are labeled *M*. The sizes of certain marker DNA are given at positions corresponding to where they migrated. Hybridized fragments showing differential intensities with these two probes are marked with *circles*.

Table 1 Contig assembly at different Sulston cutoff scores. Contigs consisting of four or more clones are listed for assemblies at different cutoff values. Contigs are numerically named in each assembly. Although, in some cases, contigs with the same name comprise the same or similar sets of clones; similar contigs may be given different names

Contig	No. of clones
Cutoff 10^{-4}	
Contig6	42
Contig4	11
Contig11	9
Contig7	8
Contig5	5
Contig13	5
Contig9	5
Contig1	5
Contig2	4
Contig15	4
Contig10	4
Contig19	4
Cutoff 10^{-5}	
Contig6	21
Contig4	11
Contig9	10
Contig12	8
Contig7	6
Contig1	5
Contig10	5
Contig5	5
Contig2	4
Contig16	4
Contig11	4
Contig14	4
Cutoff 10^{-6}	
Contig6	16
Contig5	10
Contig10	8
Contig7	6
Contig4	5
Contig1	5
Contig9	5
Contig19	4
Contig11	4
Contig12	4
Contig14	4
Contig2	4

largest contig contains 16 clones. These contigs were compared with the Contig6 assembled at 10^{-5}. It was found that several clones at the left end of Contig6 assembled at 10^{-5} (603I16, 118I23, 1199L22, and 223I19) were removed at the higher assembling stringency. At the lower assembling stringency, several singletons and contigs with two BAC clones were merged with the large contig. However, based on the clone order in the contig, these merges caused problems in the alignment of the α-gliadin fragments in overlapping clones (data not shown). It was decided that contigs assembled at 10^{-5} gave the best agreement with the BAC Southern hybridization data.

Determination of retrotransposons in BAC contigs

The average size of the BAC insert is around 130 kb (Cenci et al. 2003), whereas the α-gliadin genes are only about 3 kb in size (Anderson et al. 1997); yet 37% of BAC clones identified contain only one copy of the α-gliadin gene. This, taken with the contig results, implies that in many cases, the α-gliadin genes are not arranged in tandem repeats, and that most regions in the BAC clones are not related to the α-gliadin gene sequences. Recent studies of wheat genome structure and organization have revealed that more than 80% of wheat genomic sequences are repetitive DNA (SanMiguel et al. 2002; Anderson et al. 2003; Wicker et al. 2003a; Yan et al. 2003). Among these repetitive DNA, LTR retrotransposons are the most common components. To understand sequence composition in the *Gli-2* loci, we hybridized *Hin*dIII-digested BAC DNA with conserved LTR sequences for *Wis*, *Wham*, and *Sabrina* retrotransposable elements (Fig. 4a–c). *Wis* is a *Ty1-copia*-like retrotransposable element, which was first found in wheat and is one of the best characterized (Herberd et al. 1987; Muniz et al. 2001). In the 286-kb *Glu-B1* region of the wheat B genome, *Wis* retroelements account for about 25% of the sequence (X. Kong et al., unpublished data). In almost all of the BAC clones, *Wis* LTR-related fragments were recognized by the probe, suggesting that the *Wis* retroelement is a prevalent component in the *Gli-2* loci. The number of *Wis* LTR-related fragments ranges from one to five, with an average of 3.5 fragments per BAC clone. However, fragments with weaker hybridization signals might correspond to the closely related *Angela* and *BARE-1* retroelements (see results below; SanMiguel et al. 2002)

The *Sabrina* retrotransposable element was first described in a 66-kb barley DNA sequence (Shirasu et al. 2000). This element is probably an ancient retrotransposon, as only partial sequences of the element are often detected. Recently, it was classified into the *Athila* family based on sequence analysis of a full-length *Sabrina* element (Wei et al. 2002). Like the *Wis* LTR probe, the *Sabrina* LTR probe detected multiple fragments in the majority of BAC clones. Two BAC clones (lanes 5 and 11) did not contain any fragment recognized by the *Sabrina* LTR probe, but they contained sequences that hybridized to the *Wis* LTR probe (Fig. 4a). The average number of *Sabrina* LTR fragments was about 3.4 per BAC clone, comparable to those of the *Wis* LTR fragments.

In addition, several *Hin*dIII fragments were recognized by both the *Wis* and *Sabrina* probes, suggesting that the two retroelements were close to each other, or they were part of a nested retrotransposon cluster (SanMiguel et al. 1996). In cereals with large genomes, retrotransposable elements are often organized in a nested clustering pattern by preferential insertion into existing elements.

We also examined a gypsy-like retrotransposon, the *Wham* retroelement (SanMiguel et al. 2002), in the BAC clones. Apparently, the number of *Wham* LTR fragments in these BAC clones is fewer than the number of *Wis* and *Sabrina* LTR fragments. Most of the BAC clones

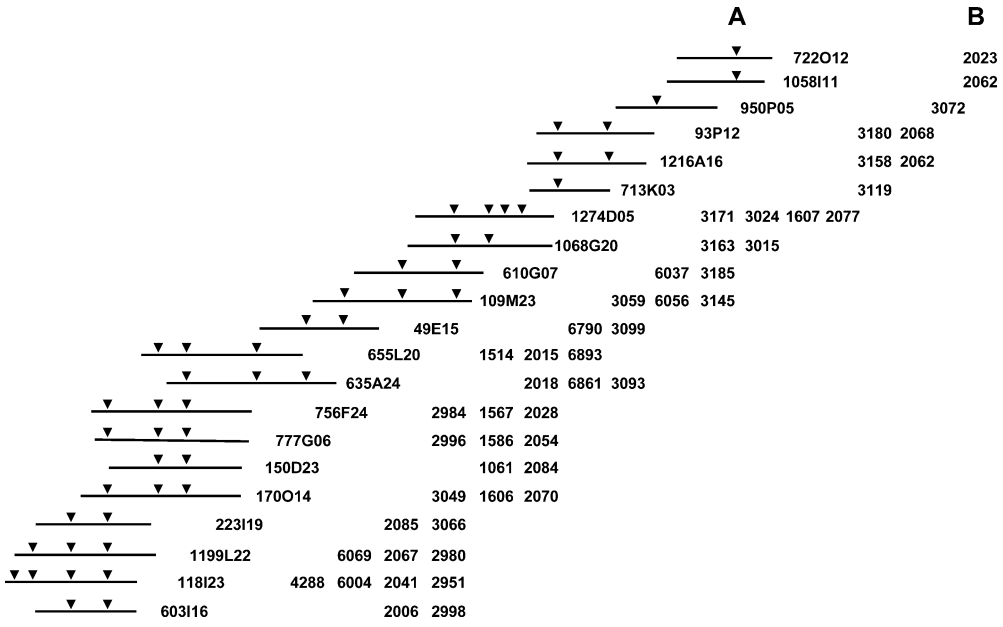

Fig. 3a, b Alignment of the BAC clones in Contig6 with the *Hin*dIII fragments containing α-gliadin genes in the overlapping BAC clones. **a** Order of the BAC clones in Contig6 of the assembly at Sulston score of 10^{-5}. **b** The calculated α-gliadin fragment sizes of the *Hin*dIII digests determined by Quantity One software (see Materials and methods) and the alignment of these fragments in the overlapping BAC clones based on their sizes. The tentative locations of these α-gliadin genes are then indicated with *solid triangles* in Contig6. The shared α-gliadin genes in overlapping clones are indicated by *aligned triangles* in the contig

contained only one or two bands, and 10% of the BAC clones had no bands. The fewer *Wham* LTR fragments imply that *Wham* elements may not be as abundant in the *Gli-2* regions as the *Wis* and *Sabrina* retroelements.

We hybridized the BAC DNA using a probe derived from the internal integrase domain of the *Wis* element. Previously, it was reported that the barley genome contained a large excess of LTRs relative to the internal regions of the *BARE-1*, a barley retroelement closely related to the *Wis* element in the wheat genome (SanMiguel et al. 2002). It was estimated that the barley genome contained an average of 16.5 times more LTRs than integrase copies. This differs significantly from the expected ratio of 2:1 (Vicient et al. 1999). Our results show that only a few of the α-gliadin BAC clones contained the integrase gene sequence (Fig. 4d). The average LTR:integrase ratio was about 16:1, similar to what was observed for the barley genome. The higher-than-expected ratio of LTR:integrase may be caused by unequal crossing-over and/or intrachromosomal recombination between LTRs. One product resulting from this type of recombination is the solo LTR (Shirasu et al. 2000; Devos et al. 2002).

Sequence composition as revealed by BAC end sequencing

To further determine the sequence composition of the *Gli-2* locus region, we sequenced the ends of each α-gliadin BAC clones using both T7 and SP6 primers. We generated a total of 400 quality reads with an average size of 450 bp per end, representing a 92% success rate of BAC end sequencing. This provided a total of approximately 180 kb of sequence for the *Gli-2* locus regions. Individual sequences were compared with the NCBI database for homology to characterized genes or proteins and with the TREP database for known repetitive DNA. The search results are summarized in Table 2. Among the 400 sequences, 16 of them matched α-gliadin gene coding sequences. However, the most frequent matches were to repetitive DNA. Two hundred eighty sequences (70%) showed significant similarities to different types of repetitive DNA (cutoff $<10^{-10}$). The most abundant type of repetitive DNA is the LTR retrotransposable element. We found that 260 sequences showed significant matches to known LTR retrotransposons in the database. Different types of LTR retroelements were present at different frequencies. Retrotransposons *Fatima* and *Sabrina* were matched 59 and 56 times, respectively, suggesting they are abundant in the *Gli-2* regions. Although Southern hybridization using retroelement LTR probes suggested that *Sabrina* and *Wis* elements were similarly abundant, there were only 11 end sequences matching the *Wis* retroelement. However, *Wis*, *BARE-1*, and *Angela* are closely related retrotransposons (SanMiguel et al. 2002), and the *Wis* LTR probe may recognize all of these elements. The total number of *Wis*, *BARE-1*, and *Angela* retroelements was similar to that of the *Sabrina* or *Fatima* elements. We identified eight BAC end sequences as *Wham* elements, which was consistent with the observation that *Wham* LTR fragments were less abundant than the *Wis* and *Sabrina* fragments (Fig. 4c).

In addition to the above class I retrotransposons, which replicate via an mRNA intermediate, we identified 20 sequences that belong to class II repetitive DNA elements, whose movements involve excision from one site and integration into a different site in the genome. One class II repetitive element is the CACTA transposon, which recently has been characterized in the wheat genome (Wicker et al. 2003b). All 20 class II repetitive DNA elements identified in the *Gli-2* regions showed significant matches to the CACTA transposons, supporting an active role of the CACTA element in wheat genome evolution.

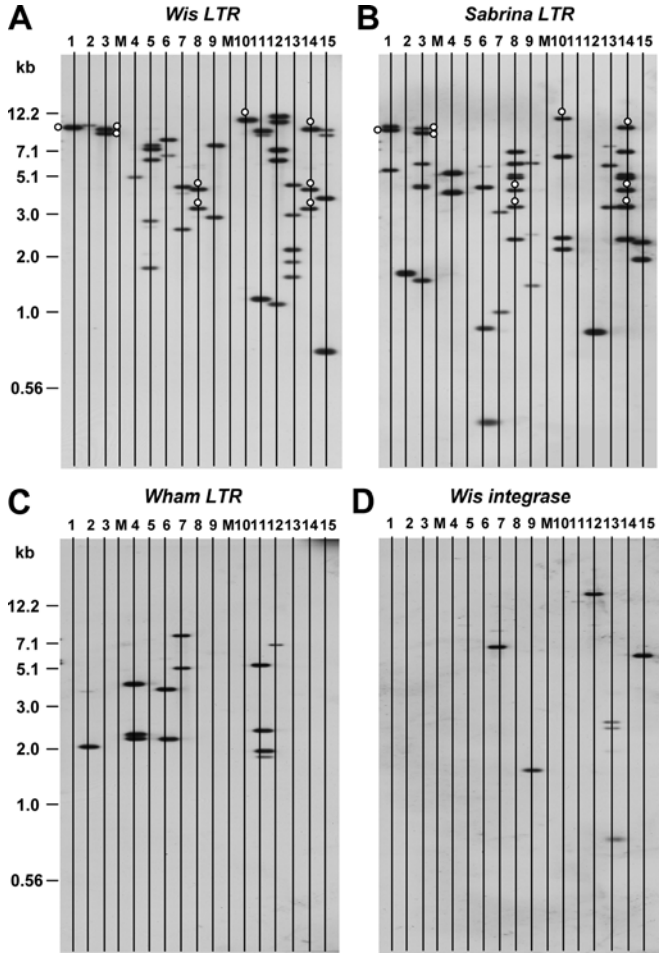

Fig. 4a–d Examples of Southern hybridization of BAC DNA using probes specific for different retrotransposons. BAC DNA was digested with *Hin*dIII, and the resulting fragments were separated by agarose gel electrophoresis. Duplicate gels were blotted onto nitrocellulose membranes and hybridized with specific probes derived from the LTR regions of *Sabrina* (**a**), *Wis* (**b**), and *Wham* (**c**), and from the integrase region of *Wis* (**d**). Lanes containing BAC DNA are numbered *1–15*. Lanes loaded with marker DNA are labeled *M*. The sizes of the marker DNA are indicated. In **a** and **b**, *Hin*dIII fragments hybridized by both *Wis* and *Sabrina* LTR probes are marked with *circles*

To search for sequences transcribed from the *Gli-2* regions, the BAC end sequences were also BLASTed against the EST database using BLASTN. In addition to the α-gliadin and transposon coding sequences, 17 other sequences also found significant matches (cutoff $<10^{-10}$) in the EST database. While 13 of them were unknown sequences, three sequences (857E16T3, 987H15T3, and 104L14T3) showed significant similarities to sequences in the protein database using BLASTX search. They were the zinc finger protein (AAN37404, 4×10^{-29}), protein F2D10.32 (E86340, 2×10^{-20}), and the NBS/LRR-like protein (AAM47598, 2×10^{-10}), respectively. The BAC clones carrying these sequences were all singletons, so it is likely that they are putative genes close to the gliadin regions, rather than the genes interspersed in the cluster of α-gliadin genes. However, we cannot exclude the possibility that these genes are located within the *Gli-2* regions as a result of translocation events. Furthermore, 57 BAC end sequences (14.3%) showed no significant similarity to any sequences in the current database; their identities have yet to be determined.

Table 2 Analysis of BAC end sequences

Homology	Number of sequences[a]
α-gliadin	16
Class I retrotransposon[b]	260
Sabrina	(56)
Wis	(11)
Angela	(23)
BARE-1	(11)
Wham	(8)
Fatima	(59)
Wilma	(15)
Barbara	(10)
Deniela	(12)
Class II transposon	20
EST and protein	4
EST only	13
No hits	57

[a]The number of sequences matching the corresponding element is provided the *parenthesis*
[b]Only abundant long terminal repeat retroelements of class I retrotransposons are listed

Discussion

Genomic organization of the *Gli-2* loci

The *Gli-2* loci represent complex genomic regions that have undergone dynamic evolution. Significant variation of α-gliadin genes has been detected among the genomes of different wheat cultivars and their related species (Reeves and Okita 1987; Anderson et al. 1997). The availability of BAC libraries has allowed a better characterization of the physical structures of large, contiguous chromosomal regions. Although α-gliadin genes are considered to cluster within the *Gli-2* loci, our characterization of α-gliadin BAC clones from tetraploid wheat revealed that distance between α-gliadin genes varied significantly. The α-gliadin gene density ranges from one to four genes per 130 kb based on the number of α-gliadin genes in a single BAC clone. Four contigs containing six or more genes were generated on the basis of overlapping sequences among α-gliadin BAC clones. Genomic Southern hybridization of wheat DNA indicated that the A genome carries multiple *Hin*dIII fragments containing α-gliadin sequences with sizes around 3 kb (Anderson et al. 1997). Contig6 contained 17 α-gliadin genes, with six in the 3-kb region when the BACs were digested with *Hin*dIII, suggesting that this contig belongs to the A genome of tetraploid wheat. Contig4 and Contig12, each having seven α-gliadin genes but no 3-

kb *Hin*dIII fragment, might be derived from the B genome (Anderson et al. 1997). The contigs represent contiguous regions rich in α-gliadins, and detailed sequence analyses of these genomic regions will provide insights into the chromosomal structure and evolution of the multigene α-gliadin family.

In the course of contig assembly, a large number of BAC clones remained as singletons, even when the assembly was performed at the Sulston score of 10^{-4}. Thus, these singletons were not due to the use of Sulston scores with too high a stringency. The Sulston scores used in our study are much smaller than the value (10^{-8}) used in developing BAC-based physical maps for the rice genome (Chen et al. 2002). However, higher cutoff values may be appropriate when the number of BAC clones is small and the generated contigs can be verified by other methods such as genetic maps, anchored molecular markers, and/or Southern hybridization data, as used in this report. Another reason for the large number of singletons might be the failure to isolate all the α-gliadin BAC clones. However, it is unlikely that many α-gliadin clones could have been missed in our screening—causing gaps in contig assembly—since the library screening was conducted under medium stringency conditions. Therefore, the results suggest that some intergenic regions between α-gliadin genes are so large that the entire intergenic sequence is not included in the α-gliadin gene-containing BAC clones. The existence of large BAC contigs and singletons indicates that the α-gliadin genes might not be evenly distributed in the *Gli-2* loci. Instead, there are regions containing clusters of α-gliadin genes and regions having one or no α-gliadin gene. The organization of the α-gliadin genes differs from that of ω-secalin genes in the rye *Sec-1* locus, in which each gene is contained in a 9.2-kb repeat unit and 15 repeat units are arranged in head-to-tail fashion in a 145-kb region (Clarke et al. 1996). Such organization also appears to be quite different from that of the 22-kD α-zein gene family in the maize genome, in which 22 of the entire 23 22-kD α-zein genes are arranged in tandem repeat within a 168-kb segment (Song et al. 2001), suggesting that maize and wheat have taken different paths in evolving their prolamin gene families as depicted in Fig. 5.

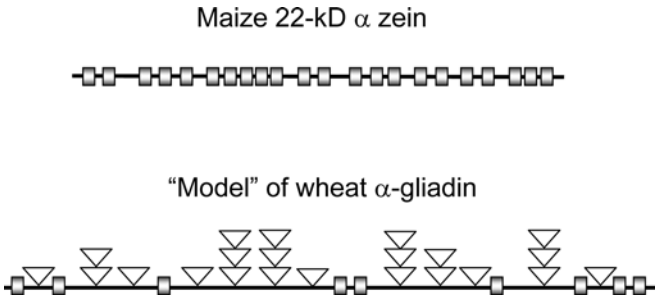

Fig. 5 A model of the structural organization of wheat α-gliadin genes compared with the maize 22-kD α-zein genes. *Boxes* represent either α-zein or α-gliadin genes. *Triangles* represent retrotransposons. The depicted structural organization of the maize α-zein genes is based on the published results (Song et al. 2001). The 22-maize α-zein genes are arranged in tandem repeat within a 168-kb segment. This region is mainly free of retrotransposable elements, resulting in the compactness of the α-zein genes. In contrast, the wheat α-gliadin genes are not evenly distributed in the *Gli-2* region; some α-gliadin genes are adjacent to one another, while others are distantly separated. Blocks of nested retrotransposons primarily contribute to the large intergenic spacing

Retrotransposons are abundant in the *Gli-2* loci

Assuming an average of two α-gliadin genes (3 kb each in size) in the average 130-kb BAC insert, over 90% of the BAC sequences would be nongenic or unrelated to α-gliadin sequences. Examination of these nongliadin sequences using BAC hybridization with retroelement probes and sequencing the ends of individual clones have indicated that repetitive DNA is likely to be the major components of the *Gli*-2 regions. A total of 280 BAC end sequences (70%) were found to be homologous to repetitive DNA elements. Among them, 260 were related to known retrotransposable elements. The complete sequences of subgenomic regions (e.g., some BAC-sized contigs) have probably identified the most common genome-specific retroelements in wheat. Of these, we found that *Sabrina*, *Fatima*, and *Wis*-related retrotransposons are the most abundant elements in the *Gli-2* regions, accounting for about 43% of the BAC end sequences. Other retrotransposons, such as *Wham* and *Barbara*, are not as abundant (Table 2). Significant variations in the amount of specific retroelement families have been reported in the maize genome (Meyers et al. 2001). Surprisingly, it was found that abundant retroelements are rarely expressed, whereas those with a low copy number are often transcribed based on the analysis of the maize EST database (Meyers et al. 2001). Although a detailed examination of the transcriptional activities of wheat retrotransposons has not yet been reported, analysis of the Triticeae EST database revealed that sequences homologous to both *Ty3-copia* and *Ty1-copia* retrotransposons are differentially expressed in the cDNA libraries from different tissues or under various biotic and abiotic stresses (Echenique et al. 2002).

Analysis of 73,000 BAC end sequences from the rice genome revealed that only 4.5% of the sequences showed homology to the known retrotransposons (Mao et al. 2000). This reflects the existence of a small number of retroelements in the compact rice genome. Retrotransposons are primarily responsible for the genome expansion of many cereal crops such as maize, barley, and wheat (SanMiguel et al. 1996, 2002; Anderson et al. 2003; Gu et al. 2003; Wicker et al. 2003a; Yan et al. 2003). Our results indicate that in the α-gliadin regions—as in other wheat genomic regions—retrotransposons comprise the majority of the sequences. The identification of large numbers of retroelements in the *Gli-2* regions implies that they have actively participated in shaping the *Gli-2* regions during the evolution of the wheat genome. This is in contrast with the rye *Sec-1* locus region of 145 kb (Clarke et al. 1996; Clarke and Appel 1999) and the maize 22-kD α-zein region of 168 kb, which generally lacks repetitive DNA

(Fig. 5; Song et al. 2001). The compactness of the 22-kD α-zein gene cluster seems to be unique when compared to other multigene families in plants. The major disease-resistance complex in lettuce is estimated to contain 24 copies, but it is spread over 3.5 Mbp (Meyers et al. 1998). It has been proposed that the 22-kD α-zein gene cluster represents a stable region (Song et al. 2001). Alternatively, this gene cluster might have evolved more recently as compared to other gene cluster complexes such as the α-gliadin loci. On the basis of our contig assembly result, it is still difficult to predict the size of the *Gli-2* loci. However, given the large number of repetitive DNA, it is likely that retrotransposable elements have greatly expanded this region.

Dynamic evolution of the α-gliadin gene family

Unlike other wheat prolamins, α-gliadins are present only in wheat and closely related species of *Triticum* and *Aegilops*, not in rye or barley (Shewry and Tatham 1990). Therefore, α-gliadin genes are likely to have evolved after the divergence of wheat and rye, but before the separation of the various wheat genomes, an event estimated to have occurred within the last 4–5 million years (Huang et al. 2002). Given the structural relatedness to other gliadin storage proteins (e.g. γ- and ω-gliadin, Shewry and Tatham 1990), the α-gliadin genes in chromosome 6 probably originated from the gliadin genes in chromosome 1 through a duplication and/or translocation event. Since then, further evolutionary events have changed the *Gli-2* loci into complex genomic regions. Several mechanisms underlying genome rearrangement may act on the *Gli-2* regions. Apparently, one such mechanism is the frequent duplication of α-gliadin genes, leading to the large number of the α-gliadin genes in the genomes. However, depending on the genotype, the estimated number of α-gliadin genes ranges from 25 to 150 copies (Anderson et al. 1997). Such variation could have been caused by differential amplification of the prolamin genes in the wheat genomes. For example, analysis of orthologous regions of different wheat genomes revealed that some LMW glutenin genes in the Am genome have undergone more rounds of duplication than the colinear copy in the tetraploid A genome (Wicker et al. 2003a). The large variation in copy number suggests the *Gli-2* regions are genetically unstable. Such instability is likely to have resulted from clusters of homologous genes that frequently undergo imperfect pairing and recombination, or "unequal crossover," resulting in the deletion of gene family members. In the maize *Rp1-D* region, a set of 27 *Rp1-D* mutants displayed at least nine different deletions of *Rp1-D* gene family members that are consistent with unequal crossover events (Collins et al. 1999).

In addition to the frequent gene amplification/deletion events, retrotransposable elements contributed significantly to the dynamic changes in the *Gli-2* regions. First, as discussed above, these regions appear to have been greatly expanded by the rapid amplification of retroelements. Second, retrotransposons preferably insert themselves into older elements, resulting in large blocks of nested retroelement structures up to several hundred kb in size. Such an insertion pattern can cause localized expansion, leading to the uneven distribution of the α-gliadin genes in the *Gli-2* regions. Third, in addition to the rapid amplification, the retroelements also may have undergone illegitimate recombination, which has been proposed to have a role in counterbalancing the genome obesity caused by retrotransposon amplification. The higher-than-expected ratio of LTR:integrase suggests a high frequency of genome rearrangements caused by such recombination. Some of these events could be associated with changes in the α-gliadin copy number if genes were located within the recombination region (Devos et al. 2002). The rapid amplification and frequent illegitimate recombination events associated with retrotransposons are primarily responsible for the extensive sequence divergence present in closely related tetraploid A and diploid A^m wheat genomes (Wicker et al. 2003a). Because of the resulting sequence divergence, it is possible to separate orthologous regions of the wheat genomes by BAC contig assembly as demonstrated here and in other studies (Cenci et al. 2003; X. Kong et al., unpublished data).

In addition to the dynamic evolution of the structural organization in the *Gli-2* loci, the coding sequences of the α-gliadin genes have undergone considerable changes during wheat evolution, particularly in the two polyglutamine domains, which are the characteristic features of the α-gliadin storage proteins and account for most of the variation in protein size (Anderson et al. 1997). It was reported that C→T substitution is the most common single-base change and has contributed to the generation of stop codons and consequently, to the observation that approximately 50% of the α-gliadin genes are pseudogenes (Anderson et al. 1997). The recent advances in the wheat EST sequencing projects will provide a useful resource for studying the transcriptional activities of the members of the α-gliadin family. Together with the further characterization of these α-gliadin contigs by sequencing overlapping BAC clones, we will have a better understanding of the distribution, regulation, and evolution of the α-gliadin gene family in wheat.

Acknowledgements The authors thank Dr. Frances M. Dupont for reading the manuscript and making suggestions and corrections. All experiments comply with US laws.

References

Altschul SF, Maden TL, Schaffer AA, Zhang JH, Zhang Z, Miller M, Lipman DJ (1997) Gapped BLAST and PSI-PLAST: a new generation of protein database search program. Nucleic Acids Res 25:3389–3402

Anderson OD, Greene FC (1997) The α-gliadin gene family. II DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet 95:59–65

Anderson OD, Litts JC, Greene FC (1997) The α-gliadin gene family. I characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and Southern analysis of gene family. Theor Appl Genet 95:50–58

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

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

Chen M, Presting G, Barbarzuk, WB, Goicoechea JL, et al (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Clarke BC, Appel R (1999) Sequence variation at the *Sec-1* locus of rye, *Secale cereale* (Poaceae). Pl Syst Evol 214:1–14

Clarke BC, Mukai Y, Appel R (1996) The *Sec-1* locus on the short arm of chromosome 1R of rye (*Secale cereale*). Chromosoma 105:269–275

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079

D'Ovidio R, Lafiandra D, Tanzarella OA, Anderson OA, Greene FC (1991) Molecular characterization of bread wheat mutants lacking the entire cluster of chromosome 6A-controlled gliadin components. J Cereal Sci 14:125–129

D'Ovidio R, Tanzarella O, Masci S, Lafiandra D, Porceddu E (1992) RFLP and PCR analyses at *Gli-1*, *Gli-2*, *Glu-1* and *Glu-3* loci in cultivated and wild wheats. Hereditas 116:79–85

Echenique V, Stamova B, Wolters P, Lazo G, Carollo L, and Dubcovsky J (2002) Frequencies of *Ty1-copia* and *Ty3-gypsy* retroelements within the Triticeae EST databases. Theor Appl Genet 104:840–844

Gu Y-Q, Anderson OD, Londeore C, Kong X, Chibbar RN, Lazo GR (2003) Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. Genome 46:1084–1097

Herberd NP, Bartels D, Thompson RD (1985) Analysis of the gliadin multigene loci in bread wheat using nullisomic–tetrasomic lines. Mol Gen Genet 198:234–242

Herberd NP, Flavell RB, Thompson RD (1987) Identification of a transposon-like insertion in a *Glu-1* allele of wheat. Mol Genet Genomics 209:326–332

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA caroxylase and 3-phosphoglycerate kinae of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99:8133–8138

Joppa LR, Du C, Hart GE, Hareland GA (1997) Mapping a QTL for grain protein in tetraploid wheat (*Triticum turgidum*) using a population of recombinant inbred chromosome lines. Crop Sci 37:1586–1589

Khan IA, Procunier JD, Humphreys DG, Tranquilli G, Schlatter AR, Marcucci-Poltri S, Fronhberg R, Dubcovsky J (2000) Development of PCR-based markers for a high grain protein content gene from *Triticum turgidum* ssp. *dicoccoides* transferred to bread wheat. Crop Sci 40:518–524

Lafiandra D, Kasards DD, Morris R (1984) Chromosomal assignment of genes coding for the wheat gliadin protein components of the cultivars Cheyenne and Chinese Spring by two-dimensional (two-pH) electrophoresis. Theor Appl Genet 68:531–539

Lijavetzky D, Muzzi G, Wicker T, Keller B, Wing, R, Dubcovsky J (1999) Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 42:1176–1182

Ma Z, Weining S, Sharp PJ, Liu C (2000) Non-gridded library: a new approach for BAC (bacterial artificial chromosome) exploitation in hexaploid wheat (*Triticum aestivum*). Nucleic Acids Res 28:E106

Mao L, Wood TC, Yu Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frisch D, Goff S, Dean RA, Wing RA (2000) Rice transposable elements: a survey of 73,000 sequence-tagged connectors. Genome Res 10:982–990

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997) High-throughput fingerprint analysis of large-insert clones. Genome Res 7:1072–1084

Meyers BC, Chin DB, Shen KA, Sivaramakishnan S, Lavelle, DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 11:1660–1676

Moullet O, Zhang H-B, Lagudah ES (1999) Construction and characterization of a large DNA insert library from the D genome of wheat. Theor Appl Genet 99:305–313

Muniz LM, Cuadrado A, Jouve N, González JM (2001) The detection, cloning, and characterization of WIS 2–1A retrotransposon-like sequences in *Triticum aestivum* L. and × *Triticosecale* Wittmack and an examination of their evolution in related Triticeae. Genome 44:979–989

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

Reeves CD, Okita TW (1987) Analysis of α/β gliadin genes from diploid and hexaploid wheats. Gene 52:257–266

San Miguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

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

Shewry PR, Tatham AS (1990) The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J 267:1–12

Shewry PR, Tatham AS, Kasarda DD (1992) Cereal proteins and coeliac disease. In: Marsh MN (ed) Coeliac disease. Blackwell, London, pp 305–348

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provide evidence for reversible genome expansion. Genome Res 10:908–915

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

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

Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. Plant Cell 11:1769–1784

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

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, 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, Keller B (2003b) CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol 132:52–63

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