Copyright © 2002 by the Genetics Society of America

The Colinearity of the *Sh2/A1* Orthologous Region in Rice, Sorghum and Maize Is Interrupted and Accompanied by Genome Expansion in the Triticeae

Wanlong Li and Bikram S. Gill[1]

Department of Plant Pathology, Wheat Genetics Resource Center, Kansas State University, Manhattan, Kansas 66506-5502

Manuscript received September 6, 2001
Accepted for publication December 17, 2001

ABSTRACT

The *Sh2/A1* orthologous region of maize, rice, and sorghum contains five genes in the order *Sh2*, *X1*, *X2*, and two *A1* homologs in tandem duplication. The *Sh2* and *A1* homologs are separated by ∼20 kb in rice and sorghum and by ∼140 kb in maize. We analyzed the fate of the *Sh2/A1* region in large-genome species of the Triticeae (wheat, barley, and rye). In the Triticeae, synteny in the *Sh2/A1* region was interrupted by a break between the *X1* and *X2* genes. The *A1* and *X2* genes remained colinear in homeologous chromosomes as in other grasses. The *Sh2* and *X1* orthologs also remained colinear but were translocated to a nonhomeologous chromosome. Gene *X1* was duplicated on two nonhomeologous chromosomes, and surprisingly, a paralog shared homology much higher than that of the orthologous copy to the *X1* gene of other grasses. No tandem duplication of *A1* homologs was detected but duplication of *A1* on a nonhomeologous barley chromosome 6H was observed. Intergenic distances expanded greatly in wheat compared to rice. Wheat and barley diverged from each other 12 million years ago and both show similar changes in the *Sh2/A1* region, suggesting that the break in colinearity as well as *X1* duplications and genome expansion occurred in a common ancestor of the Triticeae species.

WHEAT (*Triticum aestivum* L., 2n = 6x, AABBDD; *T. turgidum* L., 2n = 4x, AABB; and *T. monococcum* L., 2n = 2x, A^mA^m), barley (*Hordeum vulgare* L. 2n = 2x, HH), rye (*Secale cereale* L. 2n = 2x, RR), rice (*Oryza sativa* L.), sorghum [*Sorghum bicolor* (L.) Moench], and maize (*Zea mays* L.) are important food crops of the grass family (Gramineae or Poaceae). Despite ∼55 million years of coevolution (Kellogg 2001) and the 40-fold variation in genome size among these taxa (Arumuganathan and Earle 1991), their gene content and gene order are conserved as demonstrated by comparative, albeit low resolution, mapping (Hulbert *et al.* 1990; Ahn and Tanksley 1993; Ahn *et al.* 1993; Moore *et al.* 1995). With advances in DNA cloning and sequencing technology, comparative genetics can be employed on a finer scale using large cloned fragments or long stretches of genomic sequences. Comparative sequence analysis of orthologous regions of rice, sorghum, and maize has provided important information on grass genome evolution, colinearity, and small rearrangements at the gene level (Chen *et al.* 1997; Messing and Llaca 1998; Tikhonov *et al.* 1999; Tarchini *et al.* 2000).

One genomic region analyzed by comparative sequence analysis in rice and sorghum is the *Sh2/A1*, a region initially investigated by maize geneticists. *Sh2* (*shrunken2*) codes for the large subunit of ADP-glucose pyrophosphorylase and *A1* (*anthocyaninless1*) encodes dihydroflavonol-4-reductase. These two genes are separated by ∼140 kb in maize (Civardi *et al.* 1994). The *Sh2* and *A1* are only ∼20 kb apart in rice and sorghum. Two putative transcription-factor genes *X1* and *X2* lie between *Sh2* and *A1* (Chen *et al.* 1997; Bennetzen and Ramakrishna 2002; GenBank accession no. AF101045). Sequence characterization detected a direct tandem duplication of *A1* in this region of rice and sorghum and several miniature inverted repeat transposable elements in the intergenic regions and introns (Chen *et al.* 1997, 1998). Although gene colinearity was maintained between maize, sorghum, and rice, intergenic regions expanded greatly in maize, which is consistent with its large genome (∼2500 Mb) compared to the smaller genomes of rice (430 Mb) and sorghum (750 Mb; Arumuganathan and Earle 1991).

The Triticeae species, including wheat and barley, have genomes much larger than those of the other grasses. For example, the barley (4873 Mb) and diploid wheat (5751 Mb) genomes are 11- to 13-fold larger than the rice genome. The amplification of retrotransposons is the major cause of genome obesity in maize (SanMiguel *et al.* 1996; Messing and Llaca 1998; SanMiguel and Bennetzen 1998; Tikhonov *et al.* 1999). A similar picture of genome organization in the Triticeae has emerged from the contiguous sequences of *mlo* (Panstruga *et al.* 1998) and *rar1* regions (Shirasu *et al.* 2000) of barley and the *Lr10* region of *T. monococcum* (Wicker

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AF434703 (A1-683), AF434704 (X1-532), AF434705 (X1-539), AF434706 (X1-554), and AF434707 (X2-611).

[1] *Corresponding author:* Department of Plant Pathology, 4024 Throckmorton Hall, Kansas State University, Manhattan, KS 66506-5502. E-mail: bsg@ksu.edu

et al. 2001). A further increase in genome size was brought about by polyploidy; bread wheat (*T. aestivum*) is hexaploid and has a genome of 15,966 Mb. Theoretically, the process of genome inflation could affect the gene order and content. Because the *Sh2/A1* interval has been used as a meter of genomic obesity in other cereals, it would be instructive to explore the fate of this region in terms of synteny and intergenic distances after genome expansion in the Triticeae. Here, we present the genome organization of the *Sh2/A1* gene region in the Triticeae.

MATERIALS AND METHODS

Clones and primers: A *Hin*dIII bacterial artificial chromosome (BAC) library of diploid wheat *T. monococcum* cv. DV92 (Lijavetzky *et al.* 2000) arrayed in high density on filters was used in this study. The BAC filters were probed with the wheat cDNA clone Aga7 (GenBank accession no. X14350) homologous to *Sh2* (provided by Dr. P. Sharp, Sydney University, Australia), the rice cDNA clone DFR (dihydroflavonol-4-reductase, GenBank accession no. AB003496) homologous to *A1* (provided by Dr. Y. Inagaki, National Institute for Basic Biology, Myodaiji, Okazaki, Japan), and the rice cDNA clone R2277 (GenBank accession no. D24626) homologous to the *X1* gene (supplied by Dr. T. Sasaki, Institute of Agricultural Resources, Japan). Primer pairs X2-1 (5′-ATTCATCAGCTTGGTGGC-3′ and 5′-GAGTTCTGAATCATCTGCCC-3′) and X2-2 (5′-ACCGTCAAGTACGCCTTCCC-3′ and 5′-CGCACACGTCGTGGCTGCAC-3′) were designed for PCR amplification of the last and the first exons of the *X2* gene from rice (*O. sativa* subsp. *japonica* cv. Nipponbare). Subclones X1-532 (550-bp insert) and X1-554 (420-bp insert) are homologous to the *X1* gene. X2-611 (300-bp fragment) is homologous to the last exon of the rice *X2* gene. The inserts were amplified by PCR from the subclones and used as probes for mapping.

Plant materials: For mapping the genes on chromosomes, the following cytogenetic stocks of *T. aestivum* cv. Chinese Spring (CS) were used: nullitetrasomic (NT) lines, in which a missing pair of chromosomes is compensated by four doses of its homeolog (Sears 1966); ditelosomic lines, where one chromosome arm is missing (Sears and Sears 1978); deletion lines arising from single breaks and loss of distal acentric segments (Endo and Gill 1996); and wheat-alien addition lines, where an alien chromosome pair is added to wheat. CS-Imperial rye and CS-Betzes barley addition lines were obtained from Dr. T. E. Miller, John Innes Centre, United Kingdom, and Dr. A. K. M. R. Islam, University of Adelaide, Australia, respectively. For genetic mapping, a population of recombinant inbred lines (RILs) derived from a cross between the common wheat cv. Opata 85 and the synthetic hexaploid wheat W-7984 as described in Nelson *et al.* (1995) was used (provided by Dr. M. E. Sorrells, Cornell University, Ithaca, NY).

Filter hybridization: Plant DNA was isolated following the protocol described in Faris *et al.* (2000). BAC plasmids were isolated according to Sambrook *et al.* (1989). BAC and plant genomic DNA were digested with restriction endonucleases (Promega, Madision, WI; New England BioLabs, Beverly, MA), separated by agarose gel electrophoresis, and blotted onto N^+ Hybond membrane (Amersham Biosciences, Piscataway, NJ) following the manufacturer's instructions. BAC plasmids, digested with rare cutters *Bgl*I, *Bss*HII, *Not*I, *Pme*I, *Sfi*I, *Sgr*AI, and *Swa*I (New England BioLabs) were separated by pulse field gel electrophoresis using a CHEF-DRII System (Bio-Rad, Emeryville, CA) at a field strength of 6 V/cm for 16 hr at 12° with an initial pulse time of 5 sec and final pulse time of 15 sec. The size of the BAC inserts was determined using a λ ladder (New England BioLabs) as reference. Hybridization, probe labeling, and filter washing were performed as described previously (Faris *et al.* 2000).

Subcloning and sequence analysis: Based on Southern hybridization, specific bands homologous to *A1*, *X1*, and *X2* were purified from an agarose gel, ligated in pUC18, and transformed in competent cells of the *Escherichia coli* strain DH10B by electroporation. White (recombinant) colonies were inoculated into 96-well plates. Colony-blot hybridization was performed to select positive clones and grown in a Luria-Bertani broth medium containing 100 μg/ml carbenicillin. Plasmids were purified and used as templates for sequencing from both directions. Ligation, colony-blot hybridization, plasmid isolation, and purification were done using standard protocols described in Sambrook *et al.* (1989).

Deduction of open reading frames (ORFs) and amino acid sequences, prediction of protein secondary structure, and multiple sequence alignments were performed using the Baylor College of Medicine (BCM) Search Launcher (www site: http://www.dot.imgen.bcm.tmc.edu). Multiple sequence alignment results were output by using the BOXSHADE program (version 3.2) with fraction of sequences set at 0.5 (http://www.ch.embnet.org/software/BOX_form.html). Homology searches were made using the BLAST 2.0 program of the National Center of Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov).

Genetic mapping: The mapping population, used extensively by investigators of the International Triticeae Mapping Initiative (ITMI), consists of 114 RILs. Mapping data were obtained from the GrainGenes database (http://genome.cornell.edu/cgi-bin/WebAce/webace?db=graingenes). The first 60 RILs were used for genetic mapping in this study. Linkage analysis and genetic distances were estimated with MAPMAKER software (Lander *et al.* 1987). Recombination frequencies were converted into map distances using the Haldane mapping function (Haldane 1919).

RESULTS

BAC library screening: Four high-density filters containing 73,728 BAC clones (1.15 genome equivalents) from diploid wheat were screened by hybridization to candidate clones. Two BACs each for *Sh2* and *A1* and three for *X1* were isolated. The BAC insert size ranged from 45 to 155 kb (Table 1). An agarose gel electrophoresis of the *Hin*dIII-digested DNA of seven BAC clones showed an overlap between the two BACs containing *Sh2* or *A1*, but no overlap among the three containing *X1*. Southern hybridization analysis with the probes Aga7 (*Sh2*), DFR (*A1*), and R2277 (*X1*) further confirmed the above results.

Using the first and last exons of the *X2* gene as probes, Southern hybridization of *Hin*dIII- and *Not*I-digested BACs showed that the *X2* homolog was present in BACs 611L12 and 683A21, which also contained the *A1* gene homolog. BACs 611L12 and 683A21 contain identical copies of the *A1* and *X2* homologs because hybridization patterns resulting from *Hin*dIII digestion were identical. Southern analysis, using the last exon of *X2* and the 5′ portion of *A1* as probes, showed that *X2* and *A1* are

TABLE 1

Summary of library screening, subcloning, and mapping of wheat homologs of the *Sh2*, *X1*, *X2*, and *A1* genes

BACs	Homology	Insert size (kb)	Subclone	Chromosomal location of subclones
532J13	*X1*	151	X1-532	1A, 1DL, 1HL
539B21	*X1*	112	X1-539	7A, 7B, 7D, 7H, 6R
554G10	*X1*	125	X1-554	3A, 3B, 3D
611L12	*A1-X2*	155	X2-611	3A, 3B, 3D
683A21	*A1-X2*	85	A1-683	3A, 3B, 3D, 3H, 6H, 3R
655N4	*Sh2*	125		1A, 1B, 1D, 1H, 1R
692D11	*Sh2*	45		

present in a *Not*I fragment of ∼50 kb in both BACs (data not shown).

Sh2 homologs were detected in BACs 655N4 and 692D11. A 7-kb fragment and two smaller ones were seen in BAC 655N4 but only the 7-kb fragment was present in BAC 692D11. *Sh2* is a large gene with 15 exons (Chen *et al.* 1998), and it is likely that two overlapping pieces of the gene were cloned into independent BACs during library construction.

For the *X1* gene, Southern blot analysis of BACs 532J13, 539B21, and 554G10 probed with R2277 showed different-sized fragments of varying intensities (Figure 1), suggesting that *T. monococcum* carries at least three copies of *X1* with variable sequence homologies.

Homology confirmation: Several cDNA clones coding for the large subunit of ADP-glucose pyrophosphorylase (the product of the *Sh2* gene) have been isolated from wheat (GenBank accession nos. AF026539, U61178, U61179, X14349, X14350, and Z21969) and barley (GenBank accession nos. U66876, X62242, and X67151). All of these cDNA clones, among which Aga7 (X14349) was used as a probe in this study, showed >80% identity at the nucleotide level to the protein-coding region of the *Sh2* gene of maize, *Sh2* homologs of rice and sorghum, and an identity/similarity of >70%/>85% at the amino acid sequence level.

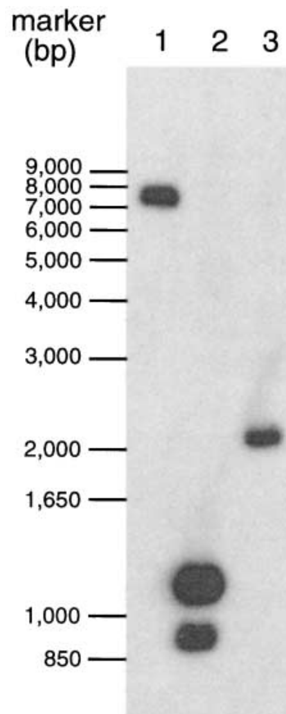

Figure 1.—Southern blot of *Hin*dIII digestion of BAC 532J13 (lane 1), BAC 539B21 (lane 2), and BAC 554G10 (lane 3) probed with rice cDNA R2277.

We subcloned and sequenced the homologs of genes *A1*, *X1*, and *X2* from the corresponding BACs of *T. monococcum*, because no sequences were available for these genes in wheat.

The wheat *A1* homolog (A1-683) was cloned from BAC 683A21 and is predicted to encode a protein of 374 amino acids. As expected, A1-683 showed the highest (∼90%) sequence identity to the barley *A1* gene homolog from the promoter through the 5′ untranslated region (UTR), the protein-coding region, and the 3′ UTR to the 3′ downstream region beyond the poly(A) signal except for introns. A1-683 showed >80% sequence identity to the *A1* homologs of sorghum, rice, and maize but only in the protein-coding regions. The inferred amino acid sequence of A1-683 was similar to those of the dihydroflavanol-4-reductases from diverse plant species. Alignment against the deduced A1 protein sequences of maize, rice, sorghum, and barley revealed an insertion of 20-amino-acid residues in A1-683 of wheat between positions 97 and 116 (data not shown).

Three *X1* homologs, X1-532, X1-539, and X1-554, were subcloned from three separate BAC clones (532J13, 539B21, and 554G10). The X1-539 subclone had the highest (75–89%) identity to all six exons of the *X1* gene of rice (AF101045) and sorghum (AF010283). A protein of 639 amino acids was deduced from the nucleotide sequence of X1-539 and showed 79 and 82% sequence similarity to the *X1* gene products of rice and sorghum, respectively. Comparison with the amino acid sequences of the X1 proteins of rice and sorghum detected an insertion of 10-amino-acid residues at positions 97–106 in the wheat X1 protein encoded by X1-539

```
wheat     1 MEYSSDDDSDISDSEIDEYGAKIQARLLSGDLKFRNGDSYSCPFCTGRKNKDYNMQSLLQ
rice      1 MSHSSDEDSEISDSEIDEYADKFYARLVAGEFKVKDGQSYSCPFCSGKKKKDENLNNLLQ
sorghum   1 --------------------------YSGGLKVKNNGENYSCLFCSSKKKNNYSKSSLVQ

wheat    61 HSSGVGAAPNRPAKDKASHRALAKHLKNGVAKPSDPQQAQQIAVEPQQPQQISVEPQPLP
rice     61 HASGVGAASNRQAKDKATHRALAKHLKNGLTKSSG-----------QQSQTAVVEPQPLP
sorghum  35 HASGVSAAPNRKAKEKAAHRALFKYLKNDLAKSSEP----------Q-PLVIPVEPQPLQ

wheat   121 NRHEKFVWPWMGVLVNVPTEWKDGRQVGESGNRLKGELSQFCPLKVIPLWNFRGHTGNAI
rice    110 NRDEKFVWPWMGVLVNVPTEWKDGRQIGRSGNHLKEQLSRFCPLKTIPLWNFRGHSGNAI
sorghum  84 NRDEKFVWPWMGILVNVPTEWKDGRQIGESGNRLKEQLSHFCPLKVIPLWTFRGHTGNAI

wheat   181 VEFAKNWNGFRNALAFEKYFEAGGCGRRDWKQNMNQGSKLCGWVARAEDYNFPGLIGDHL
rice    170 VEFGKDWHGFRNALAFEDYFGKEGYGRRDWKEKQNQGSNLFGWVARAEDHTSPGLIGDHL
sorghum 144 VEFGKDWNGFRNARTFESHFAAGGEGKKDWTGKKNQGSELYGWLARAEDYNSPGIIADYL

wheat   241 RKNADLKTIDDLENEGTRKNNKLVANLANQIEVKNKYLQELELRYNETTVSLEKMMGQRE
rice    230 RKNGDLKTINDLENEGARKTDKLVANLANQIEVKNRHLQELEVTYNERTTSLEKMMGQRE
sorghum 204 RKNGDLKSVNDLAKEGARKTDRLVANLANQIEVKNRYLQELESKYSETTASLEKMMGQRE

wheat   301 QRLQAYNEEIRKMQQLARRHSEKIIDENQNLRSELESKMSELNARSKELDDLAAKSSHDK
rice    290 QLLQKYNEEIRKMQQLAQRHSQKIIDENQKLRSELESKMSELNTRSKELDEIAAKSDYDR
sorghum 264 QLLQSYNEEISKMQQLARRHSQKVIDENQKLRSELEAKMNDLLVRSKQLDELAAKSDYDR

wheat   361 SNLEQEKQKNAIKSNHLKLATAEQQRADEEVVKLVRDQKREKVAALNKILELEQQLEAKQ
rice    350 RIEDQEKQKNAIKSSHLKLATLEQERADENVLKLVREHKREKEAAVKKILKLEQQVDAKQ
sorghum 324 RNLEQEKQKNAIKSSHLKLATLEQQKADENVLKLR-----EKHAALKKILMLEQQLDAKQ

wheat   421 TLELEIQQLKGKLEVMKHMPGHEDSVLKDKINELSEELQDKMDELDAMESLNQTLVIKES
rice    410 KLELGIQQLKGKLEVMKHMPGDEDSALKNKIDELSEELQEKMDELDAMESLNQTLVIKER
sorghum 379 KLELEIQQLKGKLKVMEHMPGDEDSASKNKINELSEALQEKIDELDGMESLNQTLVIKES

wheat   481 KSNTEMQEARKELENGLLNLPGGRAHIGIKRMGELDLKAVSNVLGQKLSKEDAEVTAAIL
rice    470 KSNTEMQDARKELENGLLDLLDGQSHIGIKRMGELDLEAFSKACRKMSSEEDAEITAAIL
sorghum 439 KSNIELQEARKELEN-----VCGQAHIGIKRMGELDLKAFSKACQKE-RTEDAEVTAAFL

wheat   541 CSKWEAEIKNPEWHPFRAVMVDGKEMERINADDAKLRELKEEHGEEIYSLVTKALREYNV
rice    530 CSKWQAEIKNPDWHPFRFVLVDGQEKE-IIEDDAKLQELKEEHGEDIYRLVRDALCEINE
sorghum 493 CSKWEAEIKNPDWHPFR----VEIIED-----DAKLRALKEEHGEEIYALVTKALLEINE

wheat   601 N--STRYPVGELWNFREERKASLKEAVQVVLRQWRANRRKR
rice    589 YNPSGRFPVGELWNFKDKRKATLKETVQFVLRQWRANRRKR
sorghum 544 YKSKGSYPVGELWNFKENRKVTLKEAVQFYWSEALKSFLDH
```

Figure 2.—Alignment of the amino acid sequences of proteins encoded by the *X1* genes of rice and sorghum and wheat *X1* homolog X1-539.

(Figure 2). In addition, the deduced protein product of X1-539 showed high similarity to the *X1* gene product of rice and sorghum in secondary structure and is predicted to possess coiled-coil domains (Chen and Bennetzen 1996).

X1-532 and X1-554 showed sequence similarity only to the first exon, not to the other five exons of the *X1* gene of rice and sorghum. At the nucleotide sequence level, these two clones showed sequence identity of >71% (126 and 143 bp of X1-554 and 316 bp of X1-554) to the rice *X1* gene. BLASTX (translation alignment) detected sequence similarity of >55% in clone X1-532 spanning 200-amino-acid residues (∼600 bp in nucleotide sequence) and in clone X1-554 spanning 130-amino-acid residues (∼400 bp) to the *X1* gene of rice and sorghum.

Using the last exon of the rice *X2* gene as a probe, a subclone (X2-611) was isolated from BAC 611L12 containing the *X2* homolog. Sequence analysis showed 85% identity to the last exon of the predicted *X2* gene of rice and sorghum.

Using the coding sequences of the wheat homologs as queries, a BLAST search was performed against a wheat expressed sequence tag (EST) database at The Institute of Genome Research (TIGR) Gene Indices (http://www.tigr.org/tdb/tgi.shtml). Two *A1* homologs were found in the wheat cDNA library made from spikes at 5–15 days after pollination (DAP). Many *X1* homologs were found in cDNA libraries made from wheat tissues including root, leaf, seedling, spikelet, preanthesis spike, 5–15 DAP spike, and endosperm, indicating that *A1* and *X1* (X1-539) are actively transcribed in wheat. No EST match was found for the *X2* gene.

Chromosomal localization: To investigate the colinearity between genomes of the Triticeae species and those of rice, sorghum, and maize in the *Sh2/A1* region, the wheat clones were mapped using the CS NT, ditelosomic, deletion lines, wheat-alien addition lines, and the ITMI mapping population.

On the basis of restriction fragment length polymorphism (RFLP) analyses of NT lines and CS-alien addition lines, Aga7 was located on wheat chromosomes 1A, 1B, and 1D (Figure 3); rye chromosome 1R (Figure 4); and the long arm of barley chromosome 1H (Figure 4). Aga7 detected a polymorphic fragment between the parents of the ITMI mapping population, Opata 85 and

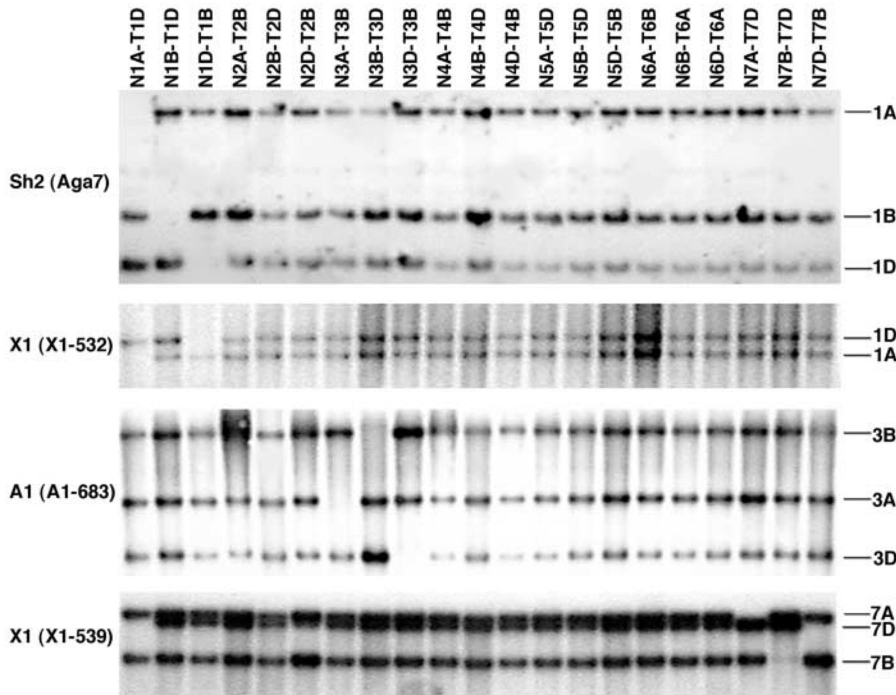

Figure 3.—Chromosome localization of the wheat homologs of genes *A1*, *X1*, and *Sh2* by Southern blot hybridization using NT lines. The 21 NT lines are indicated on the top of the lanes of Southern blots. The genes (clones) are listed to the left and the chromosome locations of the bands are indicated on the right.

Synthetic wheat W-7984, and was genetically mapped to the distal region on the long arm of chromosome 1D (Figure 5).

The wheat *A1* homolog (A1-683) was mapped to chromosomes 3A, 3B, 3D (Figure 3), and 3R (Figure 4). Ditelosomic and deletion line analysis indicated that the *A1* gene homologs are located in the proximal region of long arms of the group 3 chromosomes, between breakpoints FL0.26 and FL0.42 (FL, fraction length of distance from centromere). Two copies of the *A1* homologs exist in the barley genome on chromosomes 3H and 6H (Figure 4). No polymorphism for A1-683 was detected between Opata 85 and W-7984 even though 15 restriction enzymes were used.

T. monococcum has three copies of *X1* homologs, X1-532, X1-539, and X1-554. Both 3′ and 5′ regions of X1-539 were located on chromosomes 7A, 7B, 7D (Figure 3), 7H (Figure 4), and 6R (Figure 4). A specific fragment

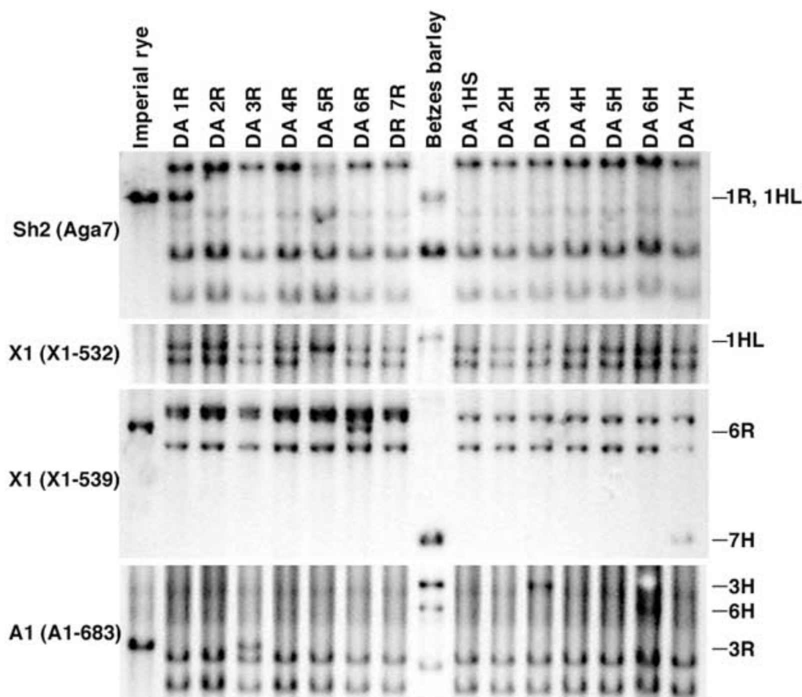

Figure 4.—Chromosome localization of wheat homologs of the *A1*, *X1*, and *Sh2* genes by Southern blot hybridization using CS-alien addition lines. The 14 CS-alien disomic addition (DA) lines and their alien parents are indicated on the top, genes (clones) are listed to the left, and the alien chromosomes on which the genes are located are indicated to the right. The remaining bands are from CS wheat.

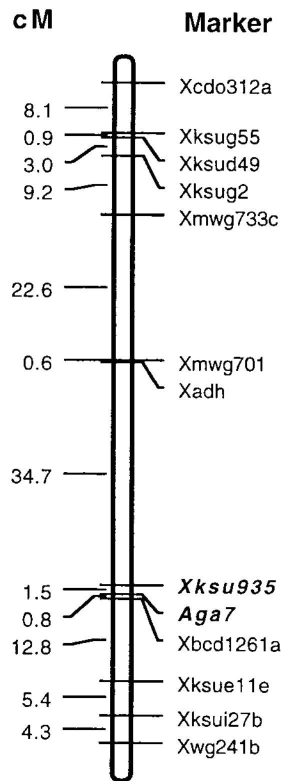

Figure 5.—Linkage map of wheat chromosome arm 1DL. The telemere is toward the bottom. Centimorgan (cM) distances are indicated to the left of the chromosome and marker loci to the right. The positions of clones mapped in this experiment are indicated in boldface type. The symbol for the marker locus detected by clones X1-532 is Xksu935.

missing in N7D-T7B also was missing in N1A-T1D (Figure 3); in the latter the distal region of chromosome 7D was deleted during its development (Devos *et al.* 1993). Similarly, X1-539 is present on 6R and not on 7R (Figure 4), because the distal region of the long arm of the original 7R chromosome was translocated to the long arm of 6R (Devos *et al.* 1993). Therefore, X1-539 can be localized to the distal region of the long arms of group 7 chromosomes of the Triticeae. The 7A band was missing from all of the CS-Betzes addition lines (Figures 3 and 4), suggesting that a deletion proximal to the *X1* locus occurred on chromosome 7A of the CS used before or during development of the addition lines. X1-539 was monomorphic between Opata 85 and W-7984.

X1-532 was located on wheat chromosomes 1A and 1D (Figure 3) and on the long arm of barley chromosome 1H (Figure 4). No homolog was detected in the rye genome. A polymorphic band was mapped to chromosome 1D, 1.5 cM proximal to the *Sh2* homolog Aga7 (Figure 5).

X1-554 was mapped to chromosomes 3A, 3B, and 3D of wheat. Deletion line analysis localized the 3A fragment to the proximal region of the short arm (data not shown).

As expected, a wheat fragment (300 bp) homologous to the last exon of the *X2* gene was localized to wheat group 3 chromosomes (data not shown).

DISCUSSION

Intergenic expansion: Initial analysis of the genomic sequences of the orthologous *Sh2/A1* region of rice and sorghum identified three genes, *Sh2*, *X*, and *A1*, which span ~30 kb. A direct tandem duplication of *A1* was found in sorghum and rice (Chen *et al.* 1997, 1998). Subsequent annotation concluded that the original *X* "gene" consists of two separate genes, *i.e.*, *X1* and *X2* (GenBank accession no. AF101045; Bennetzen and Ramakrishna 2002). We sequenced the rice cDNA clone R2277; it showed 100% identity to rice *X1* and 0% identity to *X2*. On the basis of current knowledge, five genes (*Sh2*, *X1*, and *X2*) and two tandem *A1* homologs exist in this region of the rice and sorghum genomes. The same scenario also has been revealed in maize by sequencing of the orthologous region (see Bennetzen and Ramakrishna 2002).

In the *T. monococcum* BAC library, we identified two BACs containing *Sh2* homologs, two containing *A1* homologs, and three containing *X1* homologs. Homologs of *A1* and *X2* exist in the same BACs separated by ~50 kb. They are separated by ~11.9 kb in rice and by 7.4 kb in sorghum. These data predict an expansion of approximately fourfold in the *A1/X2* interval in wheat compared to rice.

No overlap was found among BACs containing *Sh2* and *X1* in wheat, suggesting that their physical distance may be >115 kb, the average insert size of the BAC library. Tight genetic linkage of 1.5 cM, however, was observed between X1-532 and *Sh2* (Aga7) in the distal region of wheat chromosome arm 1DL. Triticeae species have large genomes and low recombination frequency, overall ~4.4 Mb cM^{-1}. In the gene-rich regions, recombination can be very high, ranging from 20 to 270 kb cM^{-1} in 1DS of *Aegilops tauschii* (Spielmeyer *et al.* 2000), 50 kb cM^{-1} in $1A^{m}S$ (Wicker *et al.* 2001), and 260 kb cM^{-1} in $5A^{m}S$ of *T. monococcum* (Tranquilli *et al.* 1999) and in 5AL of *T. aestivum* (J. D. Faris and B. S. Gill, unpublished data). On the basis of the above data, *Sh2* and *X1* orthologs may be separated by a physical distance between 115 and 390 kb, the latter estimate derived by multiplying the estimated 260 kb cM^{-1} value by 1.5 cM genetic distance. However, the physical distance between *Sh2* and *X1* may be even greater because in a DNA fiber fluorescent *in situ* hybridization (FISH) experiment, the *Sh2* and *X1* orthologs did not hybridize to the same DNA fibers of *Ae. tauschii* (P. Zhang, W. Li, B. Friebe and B. S. Gill, unpublished data). This technique can usually measure distances as far as 660 kb (Fransz *et al.* 1996). *Sh2* and *X1* are separated by 2 kb

in rice and 8.2 kb in sorghum. Therefore, the estimated expansion in the *Sh2/X1* interval of wheat is 195-fold of that of rice.

Colinearity interruption: The *Sh2* and *A1* orthologs map to chromosome 1 of rice (A. Reddy and J. L. Bennetzen, personal communication) and chromosome 3 of maize (Davis *et al.* 1999). These chromosomes are homeologous with the group 3 chromosomes of the Triticeae (Ahn *et al.* 1993; Moore *et al.* 1995). Several markers (bcd134, cdo455, and cdo118) flanking *Sh2/A1* on chromosome arm 3L of maize (Davis *et al.* 1999) also were mapped to homeologous chromosomes: chromosomes 1 of rice (Ahn *et al.* 1993) and 3L of wheat (Anderson *et al.* 1992; GrainGenes database). The markers bcd134 and cdo118 were mapped genetically to the proximal region of chromosome arm 3L of wheat (Nelson *et al.* 1995). On the basis of these data, *Sh2*, *X1*, *X2*, and *A1* should be located on group 3 chromosomes of the Triticeae.

As expected, wheat homologs of *A1* (A1-638) and *X2* (X2-611) mapped to the proximal region of the 3L arm of the Triticeae. Each detected a single copy in the A, B, and D genomes of wheat. Therefore, the *A1* and *X2* genes constitute an orthologous set and have maintained a syntenic position on homeologous chromosomes in wheat, maize, sorghum, and rice even after 55 million years of coevolution.

However, contrary to the expected synteny, a wheat homolog of *Sh2* (Aga7) has been mapped to the distal regions of the long arms of group 1 chromosomes: $1A^{m}L$ of *T. monococcum* (Dubcovsky *et al.* 1996); 1DL of *Ae. tauschii* (Lagudah *et al.* 1991); 1HL of barley (Kleinhofs *et al.* 1993); and 1AL, 1BL, and 1DL of *T. aestivum* (Ainsworth *et al.* 1995). Our results confirmed these locations of *Sh2* in the wheat genome. Because Aga7 detected a single copy in the A, B, and D genomes of bread wheat, it should be orthologous to the gene *Sh2* of maize. We show that Aga7 is closely linked to X1-532, a presumed ortholog of the *X1* gene of rice (see later section). Thus, it appears that although these two genes are syntenic, they are located on nonhomeologous chromosomes in the Triticeae compared to those in maize, sorghum, and rice.

As discussed earlier, *Sh2*, *X1*, *X2*, and *A1* genes are syntenic in maize, sorghum, and rice. However, *Sh2* and *X1* were mapped on group 1, and *X2* and *A1* mapped on group 3 chromosomes in the Triticeae. Therefore, colinearity in the *Sh2/A1* region was interrupted by a break between the *X1* and *X2* genes and another break between *Sh2* and bcd134 in the Triticeae. Next, the *Sh2-X1* segment was translocated or transposed at an interstitial position in group 1 chromosomes in the Triticeae. This scheme is consistent with that from wheat-rice comparative mapping, where most markers flanking but excluding Aga7 on the consensus map of chromosome arm 1L in the Triticeae align to their counterparts on chromosome 5 of rice (Van Deynze *et al.* 1995).

Wheat and barley diverged from the same ancestor ~12 million years ago (Huang *et al.* 2002). Both wheat and barley share a break in colinearity in the *Sh2/A1* region and associated microrearrangements in relation to rice, sorghum, and maize. Wheat, barley, and all other species of the Triticeae also have large genomes organized into a basic set of seven chromosomes irrespective of the ploidy, which ranges from 2x to 10x. Therefore, we postulate that a break in colinearity accompanied by genome expansion occurred 12 million years ago in an ancestral species of the Triticeae.

Tandem duplication of *A1*: Nearly 50 years ago, Laughnan (1952) demonstrated the tandem duplication of functional *A1* genes in maize. Genomic sequencing of the *Sh2/A1* homologous region revealed a tandem duplication of *A1* homologs ~10 kb apart in sorghum and ~5 kb apart in rice (Chen *et al.* 1997, 1998). In the Triticeae, however, there is no evidence for tandem duplication of *A1* homologs. *A1* homologs are present as a single copy in the proximal region of the long arms of chromosomes 3A, 3B, and 3D in common wheat and in the diploid species *T. monococcum*, *T. urartu*, *Ae. tauschii*, and rye. *Ae. speltoides* showed two hybridizing fragments that were caused by heterozygosity rather than duplication of the *A1* locus because it is an outcrossing species. A nontandem duplication of the *A1* homolog was found in barley. The *A1* paralogue was located on chromosome 6H (Figure 4). Unequal recombination was probably responsible for the tandem duplication of the *A1* gene (Chen *et al.* 1998). The *A1* tandem duplication might have occurred in an ancestor of maize, sorghum, and rice (Chen *et al.* 1998), but not in the Triticeae ancestor. Alternatively, an *A1* tandem duplication occurred in the common ancestor of cereals and one copy was lost in the Triticeae ancestor.

Orthology *vs.* homology: Based on its chromosome location, X1-532 is syntenic with *Sh2* in wheat, maize, sorghum, and rice. On the basis of synteny, we conclude that X1-532 constitutes part of an orthologous set of genes in these grasses. We observed additional copies of *X1*, *i.e.*, X1-539 on 7L and X1-554 on 3S of the Triticeae. On the basis of these data, X1-539 and X1-554 should be considered paralogous to the *X1* gene of rice, sorghum, and X1-532 of wheat. However, sequencing showed that it is the paralog (X1-539) rather than the orthologous copy (X1-532) that has maintained the highest homology to the *X1* gene of rice and sorghum. X1-532 and X1-554 underwent extensive degeneration, showed only limited homology in the first exon, and have lost the other five exons of the *X1* gene. Our results indicate that an ortholog based on map position is not always the functional or the most homologous copy in a genome. The discrepancy between orthology and homology may cause misleading results in comparative mapping involving distantly related genomes.

We propose the following hypothesis to explain these results (see also Figure 6). First, an *X1* ortholog on 3L

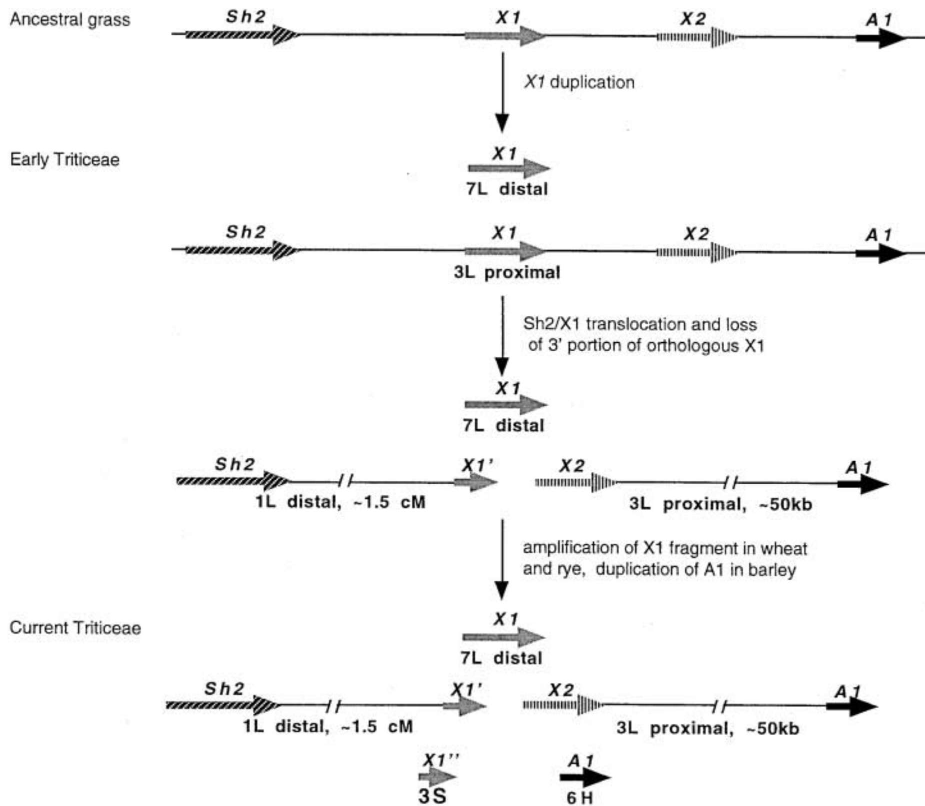

FIGURE 6.—A tentative scheme of the evolution of *Sh2/A1* counterparts in the Triticeae. *X1'* and *X1"* represent the deleted *X1* fragments, detected by X1-532 and X1-554, respectively.

was duplicated on 7L (X1-539) early during the evolution of the Triticeae. Next *Sh2/X1* was translocated or transposed to 1L followed by another round of *X1* duplication and homology degradation in the current Triticeae. The loss of the 3′ portion of X1-532 in the current Triticeae might be associated with the *Sh2/X1* translocation/transposition and low selection pressure because an intact paralog X1-539 existed somewhere else in the genome (on 7L). Both X1-532 (on 1L) and X1-554 (on 3S) lack the 3′ portion of the *X1* gene compared with X1-539 and X1-554 is more divergent than X1-532 in relation to the *X1* gene of rice and sorghum. This suggests that X1-554 was derived from X1-532 and evolved independently after the *Sh2/X1* translocation/transposition event.

The unusual evolutionary pattern of *X1* homologs implies a mechanism of colinearity breakage by duplication-deletion events. As discussed above, the ortholog X1-532 underwent extensive homology degradation and lost the five exons in the 3′ region, but the paralog X1-539 maintains a high degree of homology to the *X1* gene. An extreme situation was observed in rye where no homology was detected to X1-532, whereas a single-copy homolog was detected by X1-539 in the distal region of 7RL, which was translocated to 6RL (Figure 4). If only rye is compared with rice and sorghum, one would conclude that colinearity of the *Sh2/A1* homologous interval was interrupted by two single-gene translocations. The *X1* ortholog (X1-532) also was lost in the B genome of *T. aestivum* (Figure 3), but is present in *Ae. speltoides*, the putative B-genome donor species (data not shown), indicating that the deletion event occurred following polyploidization.

Use of a model genome: The Triticeae species have large genomes, ~80% of which are composed of repeated DNA sequences. The use of a small genome as a reference is a natural choice for positional cloning of agriculturally important genes from these species. On the basis of results of comparative mapping, rice has been proposed as a model for grass biology because it has the smallest genome among the grasses, conserved gene content, and gene colinearity with other cereal crops (Havukkala 1996). The entire rice genome is being sequenced. However, microrearrangements (small translocations, deletions, and duplications) pose a major difficulty for the application of rice as a surrogate for large cereal genomes. The situation might be more severe in polyploid species, where rapid genome restructuring can occur during speciation (see review by Paterson *et al.* 2000), and the resulting structural variation is buffered by the duplicated genomes and fixed during subsequent evolution. Our results support other recent reports documenting frequent microrearrangements between Triticeae and rice (Foote *et al.* 1997; Feuillet and Keller 1999). We further demonstrate that colinearity may break, even in orthologous regions as small as 7.2 kb in rice that are perfectly colinear in other grass species. Furthermore, a gene cloned by map position in fact may not even be a functional copy, and a gene cloned on the basis of sequence homology may not be

an ortholog, an outcome of gene amplification and gene homology degradation events, as demonstrated for the *X1* gene of the Triticeae.

We thank Dr. J. L. Bennetzen for constructive suggestions; Drs. P. Sharp, Y. Inagaki, and T. Sasaki for providing cDNA clones; and Drs. M. E. Sorrells, T. Miller, and A. K. M. R. Islam for supplying plant materials. This study is contribution No. 02-87J from the Kansas Agricultural Experiment Station, Kansas State University, Manhattan, Kansas.

LITERATURE CITED

Ahn, S., and S. D. Tanksley, 1993 Comparative linkage maps of rice and maize genomes. Proc. Natl. Acad. Sci. USA **90:** 7980–7984.

Ahn, S., J. A. Anderson, M. E. Sorrells and S. D. Tanksley, 1993 Homoeologous relationships of rice, wheat, and maize chromosomes. Mol. Gen. Genet. **241:** 483–490.

Ainsworth, C., F. Hosein, M. Tarvis, F. Weir, M. Burrell *et al.*, 1995 Adenosine diphosphate glucose pyrophosphorylase genes in wheat: differential expression and gene mapping. Planta **197:** 1–10.

Anderson, J. A., Y. Ogihara, M. E. Sorrells and S. D. Tanksley, 1992 Development of A chromosomal arm map for wheat based on RFLP markers. Theor. Appl. Genet. **83:** 1035–1043.

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

Bennetzen, J. L., and W. Ramakrishna, 2002 Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol. Biol. (in press).

Chen, M., and J. L. Bennetzen, 1996 Sequence composition and organization in the *Sh2/A1*-homologous region of rice. Plant Mol. Biol. **32:** 999–1001.

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

Chen, M., P. SanMiguel and J. L. Bennetzen, 1998 Sequence organization and conservation in sh2/a1-homologous regions of sorghum and rice. Genetics **148:** 435–443.

Civardi, L., Y. Xia, K. J. Edwards, P. S. Schnable and B. J. Nikolau, 1994 The relationship between genetic and physical distances in the cloned a1-sh2 interval of the *Zea mays* L. genome. Proc. Natl. Acad. Sci. USA **91:** 8268–8272.

Davis, G. L., M. D. McMullen, C. Baysdorfer, T. Musket, D. Grant *et al.*, 1999 A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics **152:** 1137–1172.

Devos, K. M., M. D. Atkinson, C. N. Chinoy, H. A. Francis, R. L. Harcourt *et al.*, 1993 Chromosomal rearrangements in rye genome relative to that of wheat. Theor. Appl. Genet. **85:** 673–680.

Dubcovsky, J., M.-C. Luo, G.-Y. Zhong, R. Braansteinter, A. Desai *et al.*, 1996 Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics **143:** 983–999.

Endo, T. R., and B. S. Gill, 1996 The deletion stocks of common wheat. J. Hered. **87:** 295–307.

Faris, J. D., K. M. Haen and B. S. Gill, 2000 Saturation mapping of a gene-rich recombinant hot spot region in wheat. Genetics **154:** 823–835.

Feuillet, C., and B. Keller, 1999 High gene density is conserved at syntenic loci of small and large grass genomes. Proc. Natl. Acad. Sci. USA **96:** 8265–8270.

Foote, T., M. Roberts, N. Kurata, T. Sasaki and G. Moore, 1997 Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics **147:** 801–807.

Fransz, P. F., C. Alonso-Blanco, T. B. Liharska, A. J. M. Peeters, P. Zabel *et al.*, 1996 High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. Plant J. **9:** 421–430.

Haldane, J. B. S., 1919 The combination of linkage values and the calculation of distance between the loci of linked factors. J. Genet. **8:** 299–309.

Havukkala, I. J., 1996 Cereal genome analysis using rice as a model. Curr. Opin. Genet. Dev. **6:** 711–714.

Huang, S., A. Sirikhachornkit, J. D. Faris, B. S. Gill, R. Haselkorn *et al.*, 2002 Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. Plant Mol. Biol. (in press).

Hulbert, S. H., T. E. Richter, J. D. Axtell and J. L. Bennetzen, 1990 Genetic mapping and characterization of sorghum and related crops by means of maize probes. Proc. Natl. Acad. Sci. USA **87:** 4251–4255.

Kellogg, E. A., 2001 Evolutionary history of the grasses. Plant Physiol. **125:** 1198–1205.

Kleinhofs, A., A. Kilian, M. A. Saghai-Maroof, R. M. Biyashev, P. Hayes *et al.*, 1993 A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor. Appl. Genet. **86:** 705–712.

Lagudah, E. S., R. Appels, A. H. D. Brown and D. McNeil, 1991 The molecular genetic analysis of *Triticum tauschii*—the D genome donor of hexaploid wheat. Genome **34:** 375–386.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

Laughnan, J. R., 1952 The action of allelic forms of the gene A in maize: IV. On the compound nature of A^b and the occurrence and action of its A^d derivatives. Genetics **37:** 375–395.

Lijavetzky, D., G. Muzzi, T. Wicker, B. Keller, R. Wing *et al.*, 2000 Construction and characterization of a bacterial artificial chromosome (BAC) library for A genome of wheat. Genome **42:** 1176–1182.

Messing, J., and V. Llaca, 1998 Importance of anchor genomes for any plant genome project. Proc. Natl. Acad. Sci. USA **95:** 2017–2020.

Moore, G., K. M. Devos, Z. Wang and M. D. Gale, 1995 Grasses, line up and form a circle. Curr. Biol. **5:** 737–739.

Nelson, J. C., A. E. Van Deynze, E. Autrique, M. E. Sorrells, Y. H. Lu *et al.*, 1995 Molecular mapping of wheat. Homoeologous group 3. Genome **38:** 525–533.

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

Paterson, A. H., J. E. Bowers, M. D. Burow, X. Draye, C. G. Elsik *et al.*, 2000 Comparative genomics of plant chromosomes. Plant Cell **12:** 1523–1539.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SanMiguel, P., and J. L. Bennetzen, 1998 The evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. **82:** 37–44.

SanMiguel, P., A. Tikhonov, Y.-K. Jin, N. Motchoulskaia, D. Zakharova *et al.*, 1996 Nested retrotransposon in the intergenic regions of the maize genome. Science **274:** 765–768.

Sears, E. R., 1966 Nullisomic-tetrasomic combinations in hexaploid wheat, pp. 29–45 in *Chromosome Manipulation and Plant Genetics*, edited by R. Riley and K. R. Lewis. Oliver & Boyd, Edinburgh.

Sears, E. R., and L. M. S. Sears, 1978 The telocentric chromosomes of common wheat, pp. 389–407 in *Proceedings of the 5th International Wheat Genetics Symposium*, edited by S. Ramanujam. Indian Society of Genetics and Plant Breeding, New Delhi.

Shirasu, K., A. H. Schulman, T. Lahaye and P. Schulze-Lefert, 2000 A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res. **10:** 908–915.

Spielmeyer, W., O. Moullet, A. Laroche and E. S. Lagudah, 2000 Highly recombinogenic regions at seed storage protein loci on chromosome 1DS of *Aegilops tauschii*, the D-genome donor of wheat. Genetics **155:** 361–367.

Tarchini, R., P. Biddle, R. Wineland, S. Tingey and A. Rafalski, 2000 The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell **12:** 381–391.

Tikhonov, A. P., P. J. SanMiguel, Y. Nakajima, N. M. Gorenstein,

J. L. Bennetzen *et al.*, 1999 Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. USA **96:** 7409–7414.

Tranquilli, G., D. Lijavetzky, G. Muzzi and J. Dubcovsky, 1999 Genetic and physical characterization of grain texture-related loci in diploid wheat. Mol. Gen. Genet. **262:** 846–850.

Van Deynze, A. E., J. Dubcovsky, K. S. Gill, J. C. Nelson, M. E. Sorrells *et al.*, 1995 Molecular-genetic maps of group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome **38:** 45–59.

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

Communicating editor: J. A. Birchler