Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 3431–3435, April 1997
Plant Biology

Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes

(genome microsynteny/gene duplication/repetitive DNA/intron evolution)

M. CHEN*, P. SANMIGUEL*, A. C. DE OLIVEIRA*, S.-S. WOO†, H. ZHANG†, R. A. WING†, AND J. L. BENNETZEN*‡

*Department of Biological Sciences, Purdue University, West Lafayette, IN 47907; and †Crop Biotechnology Center, Texas A&M University, College Station, TX 77843

Communicated by John D. Axtell, Purdue University, West Lafayette, IN, January 6, 1997 (received for review April 19, 1996)

ABSTRACT **Large regions of genomic colinearity have been demonstrated among grass species by recombinational mapping, but the degree of chromosomal conservation at the sub-centimorgan level has not been extensively investigated. We cloned the rice and sorghum genes homologous to the sh2 locus of maize on bacterial artificial chromosomes (BACs), and observed that a homologue of the maize *a1* gene was also present on each of these BACs. In sorghum, we found a direct duplication of *a1* homologues separated by about 10 kb. In maize, *sh2* and *a1* are approximately 140 kb apart and transcribed in the same direction, with *sh2* upstream of *a1*. In rice and sorghum, this arrangement is fully conserved. However, the *sh2* and *a1* homologues are separated by about 19 kb in both rice and sorghum. We found low-copy-number and repetitive DNAs between the *sh2* and *a1* homologues of sorghum and rice. The *sh2* and *a1* homologues cross-hybridized, but the repetitive DNA and most low-copy-number sequences between these genes did not. These results indicate that maize, sorghum, and rice have conserved gene order and composition in the *sh2-a1* region, but have acquired extensive qualitative and quantitative differences in the sequences between these genes.**

The use of common sets of DNA markers to map higher plant genomes has uncovered extensive conservation of both gene content and gene order in the grasses (1–5). Hence, despite tremendous variation in genome size (6) and chromosome number, the grasses have maintained the arrangement of much of their nuclear genomes. However, some rearrangements are observed, including inversions and translocations of most or all of individual chromosome arms (7). Particularly common, and troublesome for gene mapping programs, are segmental duplications at adjacent sites on the same chromosome arm (8). Numerous laboratories have observed that the highly repetitive sequences that make up more than 50% of many grass genomes are not highly conserved.

Two recent studies have demonstrated microcolinearity between rice chromosomal DNA on yeast artificial chromosomes (YACs) and fine scale recombinational maps in the Triticeae (9, 10). At a molecular level, though, it is not clear whether short insertions, inversions, deletions, or other rearrangements might differentiate these regions of these genomes.

In maize, the *sh2* and *a1* loci have been cloned on a single YAC, and map about 140 kb apart (11). The close physical and recombinational (0.1–0.2 cM) linkage of these two genes makes this region particularly informative for studies of intergenic chromosomal organization. We have mapped homologues of these two genes to tightly linked sites in the homeologous region of the sorghum genome (A. Melake-Berhan, and J.L.B., unpublished observation).

The construction of separate bacterial artificial chromosome (BAC) libraries containing either sorghum (12) or rice (13) chromosomal DNA suggested that the regions homologous to the *sh2-a1* domain of maize could be physically analyzed in these two grasses. The results of this analysis indicate that gene arrangement and composition are conserved in these species, but that the DNA between the genes has evolved much more rapidly.

MATERIALS AND METHODS

Materials. Libraries and host strains for BAC screening were as previously described (12, 13). Restriction enzymes for mapping studies were from New England Biolabs, Promega, and GIBCO/BRL. The [^{35}S]dATP used for DNA sequencing was obtained from Amersham. Hybridization probes for *sh2* (14) and *a1* (15) were provided by L. C. Hannah (Gainesville, FL) and H. Saedler (Cologne, Germany), respectively.

BAC Library Screening. The construction and screening of sorghum and rice BAC libraries have been described (12, 13). The hybridization probe used was the 900-bp *Eco*RI fragment of a maize *sh2* cDNA (14).

Gel Blot Hybridization. Genomic DNA isolation, gel blot analysis, and hybridization were all as previously described (16). Gel blot hybridizations of labeled total sorghum DNA to sorghum BAC fragments and of total rice DNA to rice BAC fragments were as previously described (17, 18). Other hybridization probes used were the maize *sh2* cDNA (14) and a 4.3-kb *Hin*dIII/*Eco*RI fragment containing the maize *a1* gene (15).

Restriction Mapping and DNA Sequencing. Restriction enzymes were used according to the conditions specified by the manufacturer. Sequencing and sequence analysis were performed as described (16). Some sequencing reactions were analyzed over an ALF express automated sequencer (Pharmacia).

RESULTS

Cloning of *sh2* Homologues from Rice and Sorghum. A 900-bp *Eco*RI fragment containing the first eight exons of an *sh2* cDNA from maize (14) was used to screen BAC libraries containing rice or sorghum DNA. Three clones strongly homologous to *sh2* were isolated from 11,520 rice BACs, and all three also hybridized to an *a1* probe from maize. Two clones highly homologous to *sh2* were detected out of 13,440 screened sorghum BACs. One of the two *sh2*-homologous clones from sorghum (BAC 130K4) was found to hybridize strongly to the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/943431-5$2.00/0
PNAS is available online at **http://www.pnas.org**.

Abbreviations: BAC, bacterial artificial chromosome; YAC, yeast artificial chromosome.

‡To whom reprint requests should be addressed. e-mail: maize@bilbo.bio.purdue.edu.

maize *a1* probe, while the other clone terminated between the sites of *sh2* and *a1* homology (data not shown). Hence, we had performed a "chromosome walk" to *a1* homologues in these two species in a single cloning step.

One positive clone from rice (BAC TQ27A9G, containing a 130-kb insert) and one from sorghum (BAC 130K4, containing an 80-kb insert) were selected for further analysis. Restriction maps of the central portions of these BACs indicated that the regions of hybridizational homology between *sh2* and *a1* homologues were less than 20 kb apart in each species (Fig. 1).

Confirmation of Rice *sh2* and *a1* Homology. Several fragments of the rice BAC that hybridized to *a1* and *sh2* probes were subcloned and subjected to sequencing analysis. The sequence from the *Nsi*I end of a 2.0-kb *Hpa*I/*Nsi*I fragment, inside the region of *sh2* homology, indicated that the cloned sequences were highly homologous to maize *sh2* (75% identity over 256 bp; Fig. 2). The sequence from an internal portion of a 2.5-kb *Sac*I/*Eco*RI fragment (R10; Fig. 3) indicated excellent homology with maize *a1* (82% identity over 600 bp; Fig. 2). The rice *a1* homologue, however, is missing the second intron found in the maize and sorghum genes (Fig. 2*B*). The sequence data also unambiguously positioned the homologues on the subcloned fragments, thereby indicating that the rice *sh2* and *a1* homologues reside in the same orientation as in maize, with *sh2* upstream of *a1* (Fig. 3). Integration of sequencing data with the restriction maps indicated that the two genes are about 19 kb apart (predicted 3′ end of the *sh2* homologue to the predicted 5′ end of the *a1* homologue).

Tandem Duplication of *a1* Homologues in Sorghum. Hybridization with an *a1* probe to the sorghum BAC indicated a very large region (>12 kb) of cross-hybridization for a gene that is less than 2 kb in length in maize. Further hybridizations and subcloning indicated that the *a1* probe hybridized at sites separated by about 10 kb. Segments inside the two regions of *a1* homology, the 0.9-kb *Sna*BI fragment (S10; Fig. 3) and the 3.3-kb *Pme*I/*Bam*HI fragment (S16; Fig. 3) were sequenced, and found to have overlapping homology with each other and the *a1* gene of maize (Fig. 2). These data indicate a direct duplication of *a1* homologues at this genomic location in sorghum (Fig. 3). We do not know the nature or mechanism of this duplication, although we have observed that the sequences between the genes we call *a1*-a and *a1*-b do not hybridize elsewhere on this sorghum BAC.

Confirmation of *sh2-a1* Homology and Microcolinearity in Sorghum. One end each of the 2.2-kb and 7-kb *Pac*I fragments (inside the region of *sh2* homology) were also sequenced (Fig. 1), confirming that the sorghum gene is a homologue of maize *sh2* (Fig. 2). In the regions sequenced, maize and sorghum exhibited extensive homology: 80% identity over 592 bp for *a1*-a, 81% identity over 582 bp for *a1*-b, and 92% identity over 365 bp for *sh2*. As in rice and maize, the *sh2* homologue was found to be upstream of the *a1* homologues and transcribed in the same direction (Fig. 1). The sequence data allowed precise placement of the *sh2* and *a1* homologues on the sequenced fragments, thereby indicating that both the transcriptional orientation and the *sh2-a1* separation is about 19 kb in sorghum, as it is in rice.

Sequence Conservation in Exons and Introns. As with most interspecies gene sequence comparisons, the exons were found to be more highly conserved than were the introns, but this difference was much more notable in the introns of the *a1* homologues (Fig. 2*B*) than in the *sh2* homologues (Fig. 2*A*). The ratios of exon to intron sequence identity to the maize *sh2* gene were 93% to 92% and 87% to 65% for sorghum and rice, respectively. The ratios of exon to intron sequence identity to the maize *a1* gene were 95% to 60% (sorghum *a1*-a), 96% to 59% (sorghum *a1*-b), and 85% to 50% (rice *a1*).

Comparison of Clone Structures to Chromosomal Organization. Very little information currently exists regarding the stability of higher plant DNA in BAC vectors. The high levels of interspersed repetitive DNA in plants, including both direct and inverted tandem repeats, suggest that many clones may be

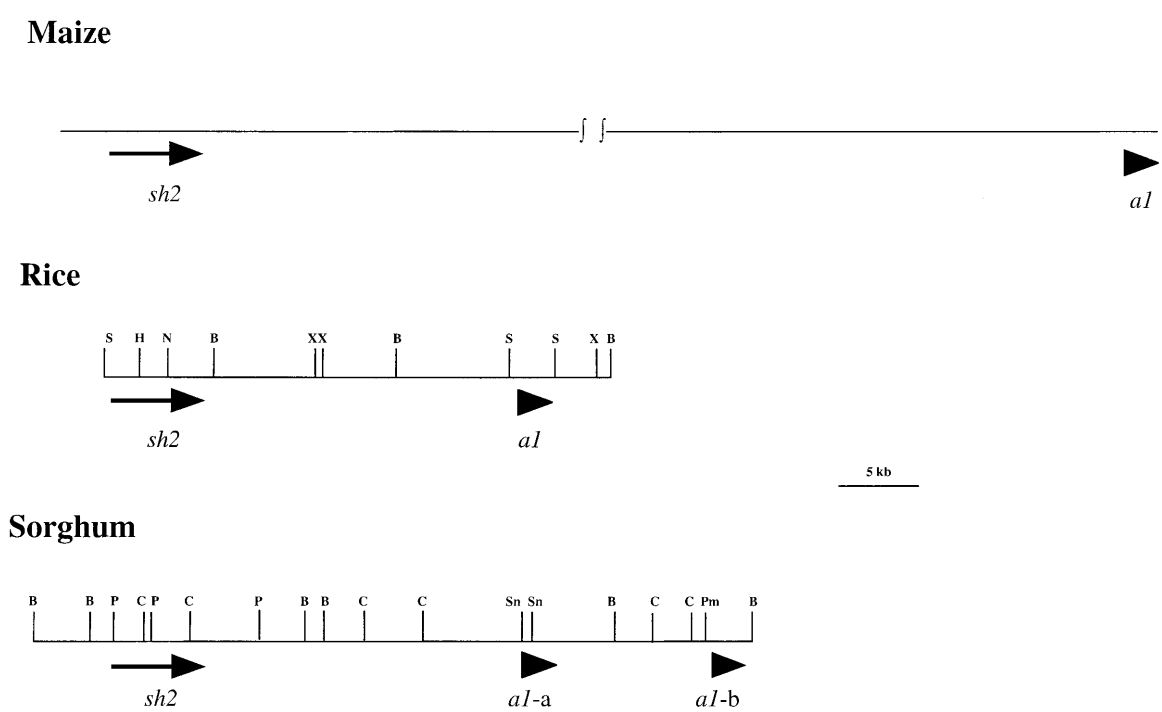

FIG. 1. Restriction maps of regions containing *sh2* and *a1* homologues from maize, sorghum, and rice. Arrows indicate predicted transcripts and transcriptional orientations. Transcript sizes in sorghum and rice are approximations based on the analogous maize transcripts. The sorghum and rice maps are drawn to scale, while the gap in the maize map indicates about 100 kb of additional DNA not shown. Within the sorghum or rice maps, all sites for the mapped enzymes are indicated, except *Hpa*I and *Nsi*I sites in rice. The mapped segments in rice and sorghum are from central portions (32 kb and 45 kb, respectively) of the much larger (130 kb and 80 kb, respectively) BAC inserts. B, *Bam*HI; C, *Cla*I; H, *Hpa*I; N, *Nsi*I; P, *Pac*I; Pm, *Pme*I; S, *Sac*I; Sn, *Sna*BI; X, *Xho*I.

FIG. 2. Sequence confirmations and comparisons for the rice and sorghum homologues of the maize *sh2* and *a1* loci. Boxes indicate exons, relative to the maize genes, and vertical lines connect identical nucleotides. (*A*) *sh2* comparisons, showing overlap between maize and sorghum (nt 1–394) and between maize and rice (nt 245–500). The different regions of overlap with the maize sequences are due to the different fragments sequenced in each species. (*B*) *a1* comparisons, showing overlap between maize, sorghum, and rice (nt 1–600).

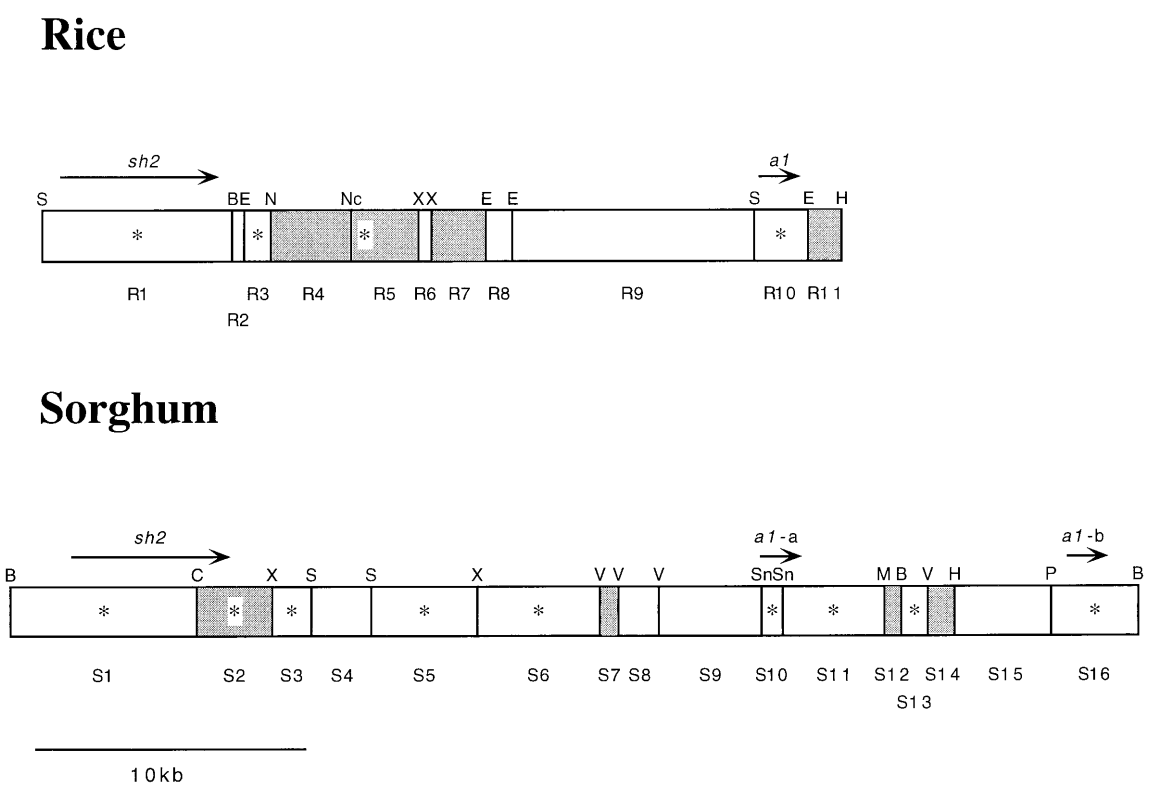

FIG. 3. Organization of repetitive and conserved sequences in the *sh2-a1* regions of sorghum and rice. The arrows above the bars indicate predicted transcripts, their orientations, and their approximate sizes. The fill indicates copy number, as determined by gel blot hybridization to total genomic DNA. The asterisks within the bars indicate fragments that cross-hybridized between these regions of the sorghum and rice BACs. The mapped segments in rice and sorghum are from central portions (30 kb and 45 kb, respectively) of the much larger (130 kb and 80 kb, respectively) BAC inserts. These maps have some different sites shown than does the map in Fig. 1. B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; M, *Mlu*I; N, *Nhe*I; Nc, *Nco*I; P, *Pme*I; S, *Sac*I; Sn, *Sna*BI; V, *Eco*RV; X, *Xho*I.

unstable. Hence, the structure of any BAC insert should be compared with the structure of the same DNA in the genome.

Gel blot hybridization experiments were performed to compare rice genomic DNA from cultivar Teqing with the rice BAC and sorghum genomic DNA from line BTX623 with the sorghum BAC, using both *sh2* and *a1* probes. Digestions with all restriction enzymes employed (*Bam*HI, *Eco*RI, *Eco*RV, *Hin*dIII, *Sac*I, and *Xho*I for rice; *Bam*HI, *Kpn*I, *Xho*I, and *Xba*I for sorghum) yielded the same size fragments, in every case, for both the BAC inserts and the genomic DNA (data not shown). Hence, these clones appear to be unrearranged relative to their genomic sources.

Repetitive DNAs in the Microsyntenous *sh2-a1* Regions of Rice and Sorghum. Gel blot hybridization of cloned sorghum BAC fragments to labeled total sorghum DNA identified four fragments (S2, S7, S12, and S14) with detected repetitive DNAs (Fig. 3; data not shown). The analogous experiment with the rice BAC, involving hybridization with total labeled rice DNA, identified four fragments (R4, R5, R7, R11) that carried repetitive DNAs (Fig. 3; data not shown). Several of these fragments were used as radioactive probes in gel blot hybridization to total DNA from the homologous genome and yielded results indicating that they were highly or middle repetitive (data not shown). One fragment that did not hybridize detectably to total rice DNA (fragment R6, Fig. 3) was found to hybridize as a single copy band when used as a probe in DNA gel blot hybridization to rice genomic DNA. Sorghum fragments S4 and S9 (Fig. 3) also yielded single-copy-number results when used as probes in gel blot hybridization to total sorghum DNA (data not shown). In this analysis, some single or low-copy-number sequences are expected to be present on fragments that also contain highly repetitive DNAs, and this is a likely explanation of the homology observed between fragments R1 of rice and S2 of sorghum, for instance.

Conserved Sequences in the *sh2-a1* Regions of Rice and Sorghum. Four subcloned fragments (*Sac*I/*Xho*I, *Xho*I/*Xho*I, *Xho*I/*Bam*HI, and *Bam*HI/*Bam*HI) covering the *sh2-a1* region of rice shown in Fig. 1 were individually labeled, pooled, and hybridized to various restriction-digested subclones of the entire *sh2-a1* homologous region of the sorghum BAC shown in Fig. 1. The reciprocal experiment was performed using pooled sorghum fragments covering the 45 kb shown in Fig. 1 as hybridization probes to subclones of the 32 kb of rice DNA shown in Fig. 1. Four rice fragments were found to cross-hybridize: fragment R1, containing the *sh2* homologue; fragment R3; a 1.0-kb portion at the left end of fragment R5; and fragment R10, carrying the *a1* homologue (Fig. 3; data not shown). Nine sorghum fragments were found to cross-hybridize with the rice BAC insert. These included five fragments that contained either *sh2* or *a1* homologues (S1, S2, S10, S11, and S16) and four others (S3, S5, a small segment within S6, and S13) (Fig. 3; data not shown).

DISCUSSION

This paper reports direct evidence that the colinearity of genomes detected at the levels of comparative recombinational mapping (1–5), or comparison of probes from single YAC clones to recombinational maps (9, 10), is also observed when one compares the physical structure of the evolutionarily conserved components of large contiguous segments. These results also indicate the differing characteristics of the conserved and variable sequences in a syntenous region of three grass genomes.

Microcolinearity in the *sh2-a1* Region. Our results demonstrate that both *sh2/a1* gene composition and gene arrangement have been maintained over the 50 million years of independent evolutionary descent (19) that separate maize, sorghum, and rice. Although we have no evidence regarding expression of the rice or sorghum homologues, their high degree of conservation (particularly in exons) suggests that they are functioning loci.

The local tandem duplication of an *a1* homologue is one type of rearrangement that we detected, and could cause some confusion in assignment of true orthology. Such a duplication is unlikely to interfere with, or even be detected by, recombinational mapping programs. However, "distantly tandem" duplications (8) that are separated by many centimorgans can cause severe problems in map comparison, especially when only a subset of DNA marker bands are mapped in a given population (20). The mechanisms of local and distant tandem duplications may or may not be the same, but we should be able to acquire information regarding the nature of the *a1*-a/*a1*-b duplication in future sequence analyses.

The most striking exception to the *sh2-a1* synteny between maize, sorghum, and rice is the more than seven-fold greater distance between these genes in maize. The nuclear genome of maize is about 3.5-fold larger than that of sorghum and about 6-fold larger than that of rice (6). However, maize and sorghum are both replete with segmental duplications, indicating that they might be best viewed as "degenerate" tetraploids (1, 21). Hence, on average, if these species all have similar gene content (22), one expects maize genes to be about 3–4 times as far apart as those same genes would be in rice and sorghum.

In maize, we have found that most of the sequences between genes are repetitive transposable elements (23). If the repetitive DNAs that differentiate these species are mostly mobile, then it is likely that the repetitive DNA clusters (17) that separate a given pair of genes will be highly variable in both content and size between even closely related species or individuals of the same species.

Sequence Conservation and Genetic Function. Our studies have found that most repetitive DNAs in maize and sorghum do not cross-hybridize (1, 24). However, in both species the types of repeats (mobile elements) and their clustering into large 2- to 200-kb blocks appear to be quite similar (17, 22, 24). In the microsyntenous *sh2-a1* regions of sorghum and rice, none of the identified repetitive DNAs cross-hybridize, but both of the known genes do. Hence, we expect that this cross-hybridization process would be an efficient technique for the identification of conserved genetic functions (i.e., genes) in microsyntenous species. Our data would suggest, for instance, that two to three additional genes reside between *sh2* and *a1* homologues in sorghum and rice.

In cross-species comparisons of gene sequences, coding exons are found to be more highly conserved than are noncoding introns. The degree and rate to which introns diverge is very different between genes, however, for unknown reasons. In the *sh2-a1* region, we have observed that the *a1* introns have diverged to a much greater degree than have the *sh2* introns. This result for two tightly linked loci suggests that possible differences in regional mutation rate do not account for different degrees of intron divergence.

Microsynteny and Gene Isolation. The tighter physical linkage of genes in rice and sorghum, compared with that of larger genome grasses such as maize, barley, and wheat, should be a general phenomenon. This should allow a more rapid chromosome walk to genes in a surrogate small genome species (like rice) than in its big genome relatives (22). Our cloning of both *sh2* and *a1* homologues on relatively small fragments in rice and sorghum provides an example of this process. However, all regions will probably not be microsyntenous or completely unrearranged, and investigators should probably test the nature of microsynteny in the targeted region before launching a major gene isolation campaign.

The apparent high mobility and rapid divergence of the repetitive DNAs between grass genes will also assist in the gene isolation process. Putative genes might be identified by a simple process of cross-hybridization of homologous regions cloned from two microsyntenous genomes.

Genome Structure and Function. Future analyses of cloned segments from parallel grass genomes will be targeted at understanding the precise nature of the changes that have differentiated these nuclear genomes. The *sh2-a1* region will be exceptionally suitable for studies of the degree of conservation of sites that determine chromosome folding (23), recombinational preferences (11), and regulated gene expression. This novel approach should allow identification of important (i.e., conserved) components of genome organization, at a level heretofore uninvestigated.

This work was supported by grants from the McKnight Foundation and the United States Department of Agriculture/National Research Initiative Competitive Grants Program (94-37300-0299) to J.L.B., the Texas Agricultural Experiment to R.A.W., and the Rockefeller Foundation to H.B.Z. and R.A.W.

1. Hulbert, S. H., Richter, T. E., Axtell, J. D. & Bennetzen, J. L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4251–4255.
2. Ahn, S., Anderson, J. A., Sorrells, M. E. & Tanksley, S. D. (1993) *Mol. Gen. Genet.* **241,** 483–490.
3. Devos, K. M., Millan, T. & Gale, M. D. (1993) *Theor. Appl. Genet.* **85,** 784–792.
4. Jena, K. K., Khush, G. S. & Kochert, G. (1994) *Genome* **37,** 382–389.
5. Kurata, N., Moore, G., Nagamura, Y., Foote, T., Yano, M., Minobe, Y. & Gale, M. D. (1994) *Bio/Technology* **12,** 276–278.
6. Arumuganathan, E. & Earle, E. D. (1991) *Plant Mol. Biol. Rep.* **9,** 208–218.
7. Moore, G., Devos, K. M., Wang, Z. & Gale, M. D. (1995) *Curr. Biol.* **5,** 737–739.
8. Sanz-Alferez, S., Richter, T. E., Hulbert, S. H. & Bennetzen, J. L. (1995) *Theor. Appl. Genet.* **91,** 25–32.
9. Dunford, R. P., Kurata, N., Laurie, D. A., Money, T. A., Minobe, Y. & Moore, G. (1995) *Nucleic Acids Res.* **23,** 2724–2728.
10. Kilian, A., Kudrna, D. A., Kleinhofs, A., Yano, M., Kurata, N., Steffenson, B. & Sasaki, T. (1995) *Nucleic Acids Res.* **23,** 2729–2733.
11. Civardi, L., Xia, Y., Edwards, K. J., Schnable, P. S. & Nikolau, B. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8268–8272.
12. Woo, S.-S., Jiang, J., Gill, B. S., Paterson, A. H. & Wing, R. A. (1994) *Nucleic Acids Res.* **22,** 4922–4931.
13. Zhang, H.-B., Choi, S.-D., Woo, S.-S., Li, Z.-K. & Wing, R. A. (1996) *Mol. Breeding* **2,** 11–24.
14. Bhave, M. R., Lawrence, S., Barton, C. & Hannah, L. C. (1990) *Plant Cell* **2,** 581–588.
15. O'Reilly, C., Shepherd, N. S., Pereira, A., Schwarz-Sommer, Z., Bertram, I., Robertson, D. S., Peterson, P. A. & Saedler, H. (1985) *EMBO J.* **4,** 877–882.
16. Jin, Y.-K. & Bennetzen, J. L. (1994) *Plant Cell* **6,** 1177–1186.
17. Springer, P. S., Edwards, K. J. & Bennetzen, J. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 863–867.
18. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. & Bennetzen, J. L. (1995) *Genetics* **140,** 315–324.
19. Doebley, J., Durbin, M., Golenberg, E. M., Clegg, M. T. & Ma, D. P. (1990) *Evolution* **44,** 1097–1108.
20. Bennetzen, J. L. (1996) in *The Impact of Plant Molecular Genetics*, ed. Sobral, B. W. S. (Birkhauser, Boston, MA), pp. 71–85.
21. Helentjaris, T., Weber, D. L. & Wright, S. (1988) *Genetics* **118,** 353–363.
22. Bennetzen, J. L. & Freeling, M. (1993) *Trends Genet.* **9,** 259–261.
23. Avramova, Z., SanMiguel, P., Georgieva, E. & Bennetzen, J. L. (1995) *Plant Cell* **7,** 1667–1680.
24. Bennetzen, J. L., Schrick, K., Springer, P. S., Brown, W. E. & SanMiguel, P. (1994) *Genome* **37,** 565–576.