DNA Rearrangement in Orthologous *Orp* Regions of the Maize, Rice and Sorghum Genomes

Jianxin Ma*, Phillip SanMiguel‡, Jinsheng Lai**, Joachim Messing** and Jeffrey L. Bennetzen*†

* Department of Genetics, University of Georgia, Athens, GA30602, USA
† Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
‡ Genomics Core Facility, Purdue University, West Lafayette, IN 47907, USA
** Waksman Institute, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Sequence data from this article have been deposited with the GenBank Data Libraries under accession nos: AF466200, AY555142 and AY555143
Running title: Cereal Genome Structure and Evolution

Key words: Grass Genome, Microcolineary, Orthologous Loci, Transposable Elements

Corresponding Author:
Jeffrey L. Bennetzen
Department of Genetics
University of Georgia
Athens, GA 30602, USA
Tel: 706-542-3698
Fax: 706-583-0972
E-mail: maize@uga.edu
The homoeologous *Orp1* and *Orp2* regions of maize and the orthologous regions in sorghum and rice were compared by generating sequence data for over 486 kb of genomic DNA. At least 3 genic rearrangements differentiate the maize *Orp1* and *Orp2* segments, including an insertion of a single gene and two deletions that removed one gene each, while no genic rearrangements were detected in the maize *Orp2* region relative to sorghum. Extended comparison of the orthologous *Orp* regions of sorghum and *japonica* rice uncovered numerous genic rearrangements and the presence of a transposon-rich region in rice. Only 11 of 27 genes (40%) are arranged in the same order and orientation between sorghum and rice. Four of the 8 genes that are uniquely present in the sorghum region were found to have single-copy homologs in both rice and *Arabidopsis*, but none of these genes are located near each other, indicating frequent gene movement. Further comparison of the *Orp* segments from two rice subspecies, *japonica* and *indica*, revealed that the transposon-rich region is both an ancient and current hotspot for retrotransposon accumulation and genic rearrangement. We also identify unequal gene conversion as a mechanism for maize retrotransposon rearrangement.
THE grasses, including major cereals such as rice, maize, wheat, barley and sorghum, are the most agronomically and economically important plant species. Despite their fairly recent origin from a common ancestor, the grasses exhibit broad variation in genome size. On the other hand, comparative genetic mapping of rice, maize, wheat, sorghum and other grasses has revealed extensive conservation of gene content and gene order in all species investigated to date (Gale and Devos 1998), although some large chromosomal rearrangements were also observed (reviewed in Paterson et al. 2000). These studies have provided a foundation for understanding grass genome evolution, and have led to the map-based isolation of agronomically important genes (Brueggeman et al. 2002; Feuillet et al. 2003; Yan et al. 2003; Yan et al. 2004; Yahiaoui et al. 2004).

With the near completion of the rice genome sequence (Sasaki et al. 2002; Feng et al. 2002; Rice Chromosome 10 Sequencing Consortium 2003), cross-species sequence comparisons in the grasses become increasingly feasible. So far, several orthologous grass genome segments containing more than one gene have been compared at the level of DNA sequence, including the sh2/a1-homologous regions of maize, sorghum and rice (Chen et al. 1997), the adh1-homologous regions of maize, sorghum and rice (Tikhonov et al. 1999; Ilic et al. 2003), the LrK-homologous regions of barley, maize, rice and wheat (Feuillet and Keller 1999), the genomic regions near Vrn1 and its orthologues in wheat, barley, sorghum and rice (Ramakrishna et al. 2002a), the Zein gene cluster of maize and its orthologues in sorghum and rice (Song et al. 2002), the Rp1-homologous regions of maize and sorghum (Ramakrishna et al. 2002b), the Rph7-homologous regions of barley and rice (Brunner et al. 2003), and the lg2/lrs1-homoeologous regions...
of maize and its orthologue in rice (Langham et al. 2004). These studies uncovered little or no retention of sequence homology in intergenic spaces but indicate general conservation of gene content and gene order between orthologous genomic segments of grass genomes. In addition, many exceptions to genome microcollinearity such as gene deletion, insertion, duplication, inversion and translocation were observed (reviewed in Bennetzen and Ma 2003).

Traditional cytological analyses suggested that maize originated from a tetraploid (McClintock, 1930), while other genetic and molecular data also indicate that the maize genome contains many duplicated genes and duplicated segments with colinear gene arrangements (Rhoades 1951; Helentjaris et al. 1988; Ahn and Tanksley 1993; Davis et al. 1999). Some duplicated genes in maize have been isolated and sequenced (Gaut and Doebley 1997; Ilic et al. 2003; Langham et al. 2004; Swigoňová et al. 2004). By examining the patterns of sequence divergence among 14 pairs of duplicated genes in maize, Gaut and Doebley (1997) proposed that the modern maize genome originated from an ancient segmental allotetraploid event that occurred between 16.5 and 11.4 million years ago (mya), after the divergence of sorghum from one of the two maize diploid progenitor lineages that themselves diverged approximately 20 mya. However, by analyzing 11 genes from clearly orthologous segments of maize, sorghum and rice, Swigoňová and colleagues determined that the two maize progenitors and sorghum diverged contemporaneously from a common ancestor about 11.9 mya (Swigoňová et al. 2004).
In a recent manuscript, Ilic and co-authors presented a detailed genomic sequence comparison of an orthologous segment of the rice, sorghum and two maize sub-genomes (Ilic et al. 2003). This first comparative sequence analysis involving homoeologous segments of maize and corresponding colinear regions in sorghum and rice provides numerous insights into the nature and timing of local genomic rearrangements that occurred in these three important grass lineages. Ilic and coworkers identified extensive gene loss by an accumulation of small deletions in the two homoeologous segments of maize analyzed, and these two segments seem to be equally unstable compared to the orthologous regions of rice and sorghum. The progressive accumulation of small deletions, most caused by illegitimate recombination, also are responsible for rapid loss of retrotransposons, other intergenic space, and some portions of genes (e.g., introns) in the Arabidopsis, wheat, and rice genomes (Devos et al. 2002; Wicker et al. 2003; Ma et al. 2004; Ma and Bennetzen 2004).

Additional studies are needed to help identify the full spectrum of local genome rearrangement in plants and to determine their frequencies and relative contributions. Here we use comparative sequence analysis to investigate genome structure and change in orthologous Orp regions of maize, sorghum and rice, thereby uncovering rapid gene movement without gene loss, a hotspot for transposon accumulation, and a propensity for genic rearrangement within a transposon-rich region.

MATERIAL AND METHODS
**BAC selection:** An *Orp* probe was obtained from maize by PCR using primers based on the complete coding sequence of the *Orp2* gene (GenBank Acc. No. M76685). A *Hind*III BAC library with inserts from cultivar BTx623 sorghum DNA (http://www.tamu.edu/bacindex.html) and a *Mbo*I BAC library made from B73 maize DNA (Yim *et al.* 2002) were screened with the *Orp* probe as described previously (Song *et al.* 2002). The positive BAC clones detected in both libraries were fingerprinted with restriction enzyme *Hind*III and further confirmed by gel-blot hybridization analysis. Meanwhile, the identified BAC clones were digested with 8-bp specificity restriction enzymes *Asc*I, *Not*I, *Pac*I, and *Swa*I. The restriction fragments were separated by pulsed-field gel electrophoresis, transferred to nylon membranes, and hybridized with the *Orp* probe to estimate the BAC insert sizes and to construct restriction maps. This information helped determine the appropriate BACs for sequencing, and also experimentally validated the computer sequence assemblies of analyzed BACs.

The largest sorghum BAC, SB18C08, that contained an *Orp* homologue was sequenced and analyzed. The predicted genes in SB18C08 were used as probes to hybridize with the positive maize BAC clones that we previously identified. Two maize BACs, ZM573L14 and ZM573F08, sharing the greatest genic homology with each other and with sorghum BAC SB18C18 were finally chosen for sequencing.

**BAC sequencing:** Shotgun libraries for BACs SB18C08, ZM573L14 and ZM573F08 were constructed as described previously (Dubcovsky *et al.* 2001; Song *et al.* 2001). Subclones were sequenced from both directions using ABI PRISM BigDye Terminator...
Chemistry (Applied BioSystems, Foster, CA) and run on an ABI3700 capillary sequencer. Base calling and quality assessment were done using PHRED (Ewing and Green 1998). Reads were assembled with PHRAP and edited with CONSED (Gordon et al. 1998). Sorghum clone SB18C08 was sequenced at ~8 fold redundancy, while maize clones ZM573L14 and ZM573F08 were each sequenced at ~12 fold redundancy. Gaps were filled by a combination of several approaches, as described earlier (Ramakrishna et al. 2002a). The final error frequency estimated by CONSED was less than one base per 10 kb. The finished assemblies of BAC sequences were found to agree completely with their restriction maps.

The *Orp*-orthologous regions in two rice subspecies, *japonica* (c.v. Nipponbare) (http://rgp.dna.affrc.go.jp/IRGSP/) and *indica* (c.v. 93-11) (Yu et al. 2002; Zhao et al. 2004), were identified by homology comparisons of the genomic sequences in GenBank deposited by May, 2004. Sequence alignments were conducted by using BLASTN (NCBI BLAST 2.0, BLAST2 (Tatusova and Madden 1999) and CROSS_MATCH (http://www.phrap.org). We considered it as an ortholog when a sequence/contig between *japonica* and *indica* has a unique match in the *japonica* genomic sequence and the assembled *indica* shotgun sequences.

**Sequence analysis and annotation:** Gene finding programs FGENESH (http://www.softberry.com/berry.phtml?topic=gfind&prg=FGENES) with the monocot training set, GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi) with maize and/or rice training sets, and GENSCAN (http://genes.mit.edu/
GENSCAN.html) with the maize training set were used to predict potential genes in rice, sorghum and maize *Orp* BAC sequences. The genes predicted by these programs and remaining regions (excluding the identified transposable elements) of these BAC sequences were investigated by BLASTX searches against the GenBank protein database (http://www.ncbi.nlm.nih.gov/BLAST/). Sequences identified as candidate genes by the gene finding programs were further investigated to determine whether they were actually genes. From our earlier rice genome annotation studies, for instance, we have found that >30% of the candidate genes identified by these programs were actually transposons or transposon fragments (Bennetzen et al., 2004). So, we used conservation in a distantly related species as an additional criterion for gene certification. Hence, candidate rice genes were used as queries in BlastX searches against the full GenBank database, but were only considered likely genes if they detected homology at an expect value of less than $e^{-05}$ in some species other than rice. The recent release of genome sequence for maize (Whitelaw *et al.* 2003) provided a particularly useful dataset for this analysis.

Shared genes were detected by orthologous sequence comparisons and multiple sequence alignments using CROSS_MATCH, BLAST2, and ClustalX (Thomposon *et al.* 1997). Genes that were not shared in the orthologous regions were further investigated by BLASTX searches against the *Arabidopsis* protein database at The *Arabidopsis* Information Resources (TAIR) (http://www.arabidopsis.org) and the rice predicted protein database at The Institute for Genomic Research (TIGR) (http://www.tigr.org/tdb/e2k1/osa1/) to determine the copy numbers and distribution of corresponding homologous genes in the *Arabidopsis* and rice genomes. The genes in
sorghum were named in numerical order by their position on the sequenced BAC, while the genes in rice and maize were numbered according to their homology to the shared genes in sorghum. Three unshared genes in rice and one unshared gene in maize were given alphabetical designations.

Transposable elements (transposons and retrotransposons) were identified using a combination of structural analysis of repetitive DNA and homology-based searches against GenBank nucleotide and protein databases and the TIGR cereal repeat database (http://www.tigr.org/tdb/rice/blastsearch.shtml). The programs Repeat and Gap from the Wisconsin Package Version 10.1, Genetic Computer Group (GCG), were used to identify LTR-retrotransposons as described earlier (Devos et al. 2002; Ma et al. 2004). Newly identified retrotransposons were named according to the retrotransposon nomenclature previously described by SanMiguel et al. (2002). The approximate dates of LTR-retrotransposon insertion and gene duplication in rice were estimated in a manner similar to SanMiguel et al. (1998) and Ramakrishna et al. (2002a), respectively. For dating LTR-retrotransposon insertion times, the molecular clock was set at an average substitution rate of $1.3 \times 10^{-8}$ mutation per site per year that we estimated for intergenic regions in rice (Ma and Bennetzen, 2004).

RESULTS

Isolation of Orp segments of sorghum, maize and rice: The maize genes Orp1 and Orp2, encoding the β subunit of tryptophan synthase, have been cloned and mapped to the short arms of chromosomes 4 and 10, respectively (Wright et al. 1992). From earlier
comparative maps of the cereals (e.g., Gale and Devos, 1998), it appears that these two chromosome arms are homoeologues. That is, they are orthologous regions descended from two different diploid ancestors of the tetraploid progenitor of maize (Swigoňová et al. 2004). Two contiguous series (contigs) of maize BAC clones that hybridized to the Orp probe were generated by fingerprinting and restriction map analysis. Only one contig of BACs that contain an Orp gene was detected in sorghum. Because the maize genome is primarily composed of large blocks of LTR-retrotransposons often organized in a nested insertion pattern (SanMiguel et al. 1996; Fu et al. 2002; Song et al. 2002; Song and Messing 2002; 2003), it was difficult to predict which BACs of maize and sorghum would provide the best alignment of colinear genes. Therefore, we first sequenced an ~160 kb sorghum BAC, SB18C08, the largest among the overlapping BACs containing the sorghum Orp gene. After analyzing the gene content of BAC SB18C08, probes from 10 additional genes physically linked to the sorghum Orp gene were obtained by PCR and hybridized with the previously identified positive BACs of maize. Two maize BACs, ZM573F08 and ZM573L14, sharing the most genes with the orthologous region of sorghum, were then chosen and completely sequenced.

BLASTN searches against the non-redundant database at GenBank using the predicted genes in sorghum as queries were conducted to identify potential homologous segments of rice. A contig of 5 overlapping finished BAC sequences (GenBank Accession Nos. AP003896, AP005620, AP005618, AP005250, AP004591) from japonica cultivar Nipponbare were found to contain most of the genes homologous to the genes predicted
on sorghum clone SB18C08, defining this contig as an orthologous Orp region in rice. 
Therefore, a 313 kb contiguous Orp region in rice was selected for further analysis.

**Sequence organization of the Orp regions of sorghum, maize and rice:** The complete sequence of the Orp segment in sorghum clone SB18C08 is 159669 bp (GenBank Accession No. AF466200). With our criteria for gene identification (see MATERIALS AND METHODS), we identified 22 sorghum genes on this BAC (Table 1). The average gene density is one gene per 7.3 kb, similar to that previously observed in the sorghum sh2/a1 region (Chen et al. 1997), the sorghum adh region (Tikhonov et al. 1999), and the region near the Vrn1-orthologue of sorghum (Ramakrishna et al. 2002a), but higher than the density of one gene per 10.8 kb in the 215-kb region comprising the kafirin gene (Song et al. 2002). No intact transposable elements were annotated in the sorghum region, but 2 non-LTR retrotransposon fragments (-f) and one DNA transposon fragment (TNP2-f) were detected by homology-based searches (Figure 1).

The complete sequence of the Orp1 segment in maize clone ZM573F08 is 181627 bp (GenBank Accession No. AY555142). It contains 4 identified genes (Table 1). The average gene density is one gene per 45.4 kb. LTR-retrotransposons in this region are more abundant than in the average sequenced regions of maize (SanMiguel et al. 1996; Fu et al. 2002; Song et al. 2002; Ramakrishna et al. 2002b; Ilic et al. 2003). A total of 14 LTR-retrotransposons, 1 solo-LTR, and 3 retrotransposon fragments were identified (Table 1), constituting ~138 kb of DNA (~76% of the region). The majority of the retrotransposons in this region are organized in typical nested fashion (SanMiguel et al. 2003).
The 4 predicted genes are separated into 2 gene pairs by the largest (~53 kb) retrotransposon block.

The complete sequence of the Orp2 segment in maize clone ZM573L14 is 144792 bp (GenBank Accession No. AY555143) and contains 4 apparent genes (Table1). The average gene density is one gene per 36 kb. Three of these 4 genes are clustered together and separated from the other gene by a cluster of intact retrotransposons, retrotransposon fragments and a newly identified CACTA-like transposon, named fanal-1, which inserted into retrotransposon mil-1. The transposable elements on this BAC include 6 retrotransposons, 5 retrotransposon fragments, 1 DNA transposon (fanal-1), and 2 DNA transposon fragments (TNP-f and Tam3-f), together accounting for ~ 50% of this region.

The 313 kb rice genomic sequence contains 19 identified genes, including a triplication of one locus (genes 12-1, 12-2 and 12-3). The average gene density is one gene per 16.5 kb, much lower than estimated for the whole rice genome (one gene per 7-9 kb; Song et al. 2002; Sasaki et al. 2002; Feng et al. 2002; Goff et al. 2002; Yu et al. 2002; Rice Chromosome 10 Sequencing Consortium 2003). This low gene density is mainly due to the presence of a large cluster of repetitive DNA that harbors only 4 predicted genes (Figure 1). This repetitive domain is predominantly composed of LTR-retrotransposons, including 9 intact elements, 5 solo-LTRs and 5 truncated fragments, three of which (ifisi, ovikoh and pawepe) are discovered and named in this study. These elements constitute about 105 kb of DNA, accounting for ~55% of the retrotransposon-rich area or 33% of
the whole region investigated. We also identified four DNA transposons and/or fragments in the rice region, constituting 22 kb of DNA.

**Sequence comparison of colinear Orp regions of sorghum, maize and rice:** Three genes, 3, 8 and 9, are shared among rice, sorghum and maize Orp1 regions, distributed across 25 kb in rice, 53 kb in sorghum and 68 kb in maize. The maize Orp2 region also shares three predicted genes, 2, 3 and 5, with rice and sorghum. These genes are distributed across 17 kb in rice, 35 kb in sorghum and 18 kb in maize, respectively. In addition to genes 2, 3 and 5, one more gene (gene 7) is shared between sorghum and the maize Orp2 region. Several genes are missing from one or more of the four otherwise colinear segments (Figure 1). This dramatic variation of gene organization and intergenic distance is due to both the variable amount of intergenic repetitive DNAs and local genic rearrangements, such as deletions or insertions of genes. The maize Orp1 and Orp2 regions share only genes 3, the Orp loci themselves. Hence, as observed at adh1 and lg2/lrs1 loci (Illic et al. 2003; Langham et al. 2004), most of the duplicated genes present from the tetraploidization of the maize ancestor about 11.9 mya (Swigoňová et al. 2004) have been reduced by deletion to a near-diploid state. The colinearity of the Orp genes and other shared genes between maize and sorghum and between maize and rice (Figure 1) does indicate that the maize Orp1 and Orp2 regions are homoeologous segments derived from the two diploid progenitors of maize.

Among the orthologous regions (from gene 2 to gene 9) shared by sorghum, rice, and maize, several gene rearrangements can be attributed to specific lineages because we can
compare four chromosomal segments. (1) Genes 5 adjacent to Orp1 was deleted in maize. (2) Genes 4 and 6 were inserted into the sorghum region after the divergence of sorghum and maize ancestors. (3) Gene d was acquired by the maize Orp1 region after the divergence of sorghum and maize ancestors (Figure 1). In addition, 3' to Orp1 in maize, genes 1 and 2 were found to be deleted by analyzing the next maize BAC that is downstream of Orp1 and contains the Fie1 locus (Lai et al. 2004).

**Extended comparison of colinear Orp regions of rice and sorghum:** The sorghum Orp segment was compared with the continuous 313 kb orthologous region of rice. We found numerous alterations in gene content, order and orientation. A total of 14 predicted genes were found to be shared, distributed across 313 kb in rice and 159 kb in sorghum, whereas 13 additional genes are not in orthologous locations. This includes 8 genes (1, 4, 6, 7, 13, 15, 16 and 22) present in this region of sorghum but absent in the orthologous region of rice, and 5 genes (a, b, c, 12-2 and 12-3) present only in the rice region. Inspection of the adjacent BACs to the rice contig that we analyzed in this study indicated no copies homologous to genes 1 and 22. For most of the non-orthologous genes, we do not know whether they were gained or deleted in sorghum or in rice based on the comparative analysis of two species. However, for genes 4, 6 and 7 (present in sorghum or maize but not in rice), the simplest explanation suggests that they inserted in these locations in the lineage that gave rise to sorghum and/or maize.

An inversion of a cluster of 4 predicted genes (gene 17, 18, 19 and 20) was detected between rice and sorghum. These 4 genes are arranged in the indica genome in the same
order as present in *japonica* (Figure 2), but it is not clear whether the inversion occurred in an ancestor of rice or sorghum.

In rice, we discovered three copies of gene 12 (12-1, 12-2 and 12-3) that were not tandemly arrayed. Gene 12-1 remains intact, while genes 12-2 and 12-3 are truncated at their N-termini when compared with rice gene 12-1 and sorghum gene 12. If the divergence time for rice and sorghum ancestors is 60 mya (Wolfe et al. 1989; Kellogg et al. 2000), we roughly estimate that the first duplication of rice gene 12 homologues occurred approximately 25 mya. However, because only gene 12-1 appears to be intact, the truncated genes 12-2 and 12-3 may be evolving more rapidly than functional loci that usually follow standard molecular clocks. Hence, this first duplication may have occurred much less than 25 mya, and, similarly, the second duplication (yielding genes 12-2 and 12-3) may have taken place more recently than the 8 mya that we calculated. Gene 12-2 is arranged in inverted orientation relative to 12-1 and 12-3 in rice and gene 12 in sorghum, an event that probably occurred after the second duplication. In addition, putative genes a and b were found between genes 12-1 and 12-2, and between genes 12-2 and 12-3, respectively. The extra three genes (a, b and c) in rice are also truncated. Altogether, these data indicate a high frequency of several different types of genic rearrangement in this specific region of rice.

**Chromosomal locations of homologs in rice and Arabidopsis:** Nearly complete genomic sequence and comprehensive sequence annotation of the *Arabidopsis* and rice genomes allowed us to investigate the nature of some local gene rearrangements at the
whole genome level. All of the genes predicted in the Orp regions of sorghum and/or maize but not shared with the orthologous region of rice were used as queries to search against nucleotide databases and protein databases of the rice genome at TIGR (http://www.tigr.org/tdb/e2k1/osa1/) and Arabidopsis genome at TAIR (http://www.tigr.org/servlets/sv). The rice and Arabidopsis homologs closest to the corresponding sorghum genes and their chromosomal locations in individual genomes are summarized in Table 2.

All of the identified genes (1, 4, 6, 7, 13, 15, 16, 22 and d) absent in the rice Orp region were found to have homologs in both rice and Arabidopsis genome protein databases (Table 2). These homologs (the best matches) are distributed along several different chromosomes, including chromosomes 2, 4, 5, 6, 7, 9 and 12 in rice, and chromosomes 1, 2, 3, 4 and 5 in Arabidopsis (Table 2). None of these genes were closely linked to each other on any rice or Arabidopsis chromosome (data not shown). Genes 1, 7, 13 and 22 only have single copies in both rice and Arabidopsis genomes, suggesting but not proving that these loci are orthologous and further suggesting that numerous independent rearrangements involving these genes must have occurred after the divergence of sorghum and rice lineages. Multiple copies were observed for genes 4, 15, 16, and d in both rice and Arabidopsis.

A rapidly evolving retrotransposon block in rice: The rice interval contains a transposable element-rich region, mainly composed of LTR-retrotransposons (~105 kb of DNA). This regions occupies ~190 kb of DNA, but only contains 5 genes, including 2
that are duplicated (Figure 1). This segment contains a high percentage (~55%) of LTR-retrotransposons, similar to that recently observed in the centromeric region of rice chromosome 8 (Wu et al. 2004).

The assembled whole genome shotgun sequence generated from indica cultivar 93-11 (Yu et al. 2002; Zhao et al. 2004) was used in this study to investigate the timing and lineage specificities of the dramatic accumulation of retrotransposons and genic rearrangements identified in the Orp region of japonica rice. By sequence homology searches and sequence alignments, we identified 9 assembled contiguous segments (Acc. Nos.: AAAA01000069, AAAA01004112, AAAA01006364, AAAA01008548, AAAA01009470, AAAA01009525, AAAA009834, AAAA01013675, and AAAA01023118) from indica that have unique matches in both japonica genomic sequence and indica whole genome shotgun sequences, suggesting that these segments are orthologous (Figure 2).

We found 8 LTR-retrotransposons or fragments uniquely present in the Orp region of japonica, although 7 LTR-retrotransposons or fragments were shared by indica and japonica in the comparable regions (Figure 2). For all LTR-retrotransposons that are relatively intact, we employed LTR divergence as a tool to date approximate times of insertion (SanMiguel et al. 1998). We found that all intact LTR-retrotransposons uniquely present in japonica were younger than 0.44 mya (the estimated divergence time of indica and japonica, Ma and Bennetzen, 2004), and that all shared intact elements had inserted more than 0.44 mya (Figure 2). Hence, it appears that this retrotransposon block
has been continuously and independently expanding in both *indica* and *japonica* lineages by insertion of LTR-retrotransposons.

The relatively intact LTR-retrotransposons found in the maize *Orp1* and *Orp2* regions are all recent insertions. The majority of intact elements inserted less than 2 mya (Figure 3). Our estimate is consistent with the previous dating of LTR-retrotransposons in maize (SanMiguel *et al.* 1998; Swigoňová *et al.* 2004).

**The structure of a rearranged retrotransposon:** We identified a rearranged LTR-retrotransposon, *grande_573F08-1*, in the *Orp1* region of maize. Its unusual property is that a region of 2145 bp directly upstream of the 5’ LTR is very similar (>97% identical) to the sequences upstream of the 3’ LTR. The likely origin of this element structure by unequal conversion is presented in Figure 4. The bottom diagram (Figure 4C) depicts *grande_573F08-1* and sequence flanking it from the region downstream of *Orp1*. An *opie* element, likely an insertion subsequent to the events described in Figure 4, is not included in the diagram. The 13.9 kb *grande* element is 5’-flanked by 2145 bp sharing 2093 identical bp with the 3’ portion of *grande_573F08-1* immediately upstream of its 3’ LTR (but differing by four indels of 1, 9, 15 and 20 bp). A total of 44 of the mismatches are transitions and 8 are transversions. Both LTRs are 627 bp, of which 615 bp are identical (10 transitions, 2 transversions).

**DISCUSSION**
The comparative genomics approach for gene identification: In this study, 5 genes were identified by comparison of colinear regions containing the maize Orp genes and their orthologs in rice and sorghum (Figure 1; Table 3). All of these genes are also present in the Arabidopsis genome (http://www.arabidopsis.org/servlets/sv) (Table 1). Except the conserved genes, no other long sequences were shared among these regions, as has been observed for all the orthologous or colinear segments compared among maize, sorghum and rice (Bennetzen et al. 2005). However, numerous small conserved noncoding sequences (CNSs) have been identified between orthologous genes in multiple plant species, most of which are harbored in intron or promoter domains of genes (Kaplinsky et al. 2002; Inada et al. 2003; Guo and Moose 2003).

In addition to the conserved orthologous genes, we identified 10 more genes in maize, rice or sorghum that exhibited significant similarity to one or more annotated Arabidopsis genes based on BLASTX searches (Table 1). Because the lineage that gave rise to Arabidopsis has evolved independently from the grass lineage for >150 million years (Wolfe et al. 1989), it is likely that the conserved sequences between Arabidopsis and grasses are genes.

Gene finding programs such as FGENESH, GENSCAN and/or Genemark.hmm are useful but imperfect tools for gene identification. These programs predicted 40, 4, 16 and 27 additional genes in the Orp regions of rice and sorghum and the Orp2 and Orp1 regions of maize, respectively, beyond those we consider valid gene candidates (Table 3). Of these predicted genes, 28 (70%), 2 (50%), 15 (94%) and 27 (100%), respectively,
were found to have the structure and/or highest sequence similarity to transposable elements (Table 3). However, 12 predicted genes in rice (11 scattered in the transposon element-rich area of the Orp segment), 2 predicted genes in sorghum and 1 predicted gene in maize are unclear in origin, so we did not annotate them as genes. Because these predicted genes have no homologs in Arabidopsis or any other genome, we think they are rapidly evolving transposable elements or some other non-genic DNA.

**Gene content instability in the two maize subgenomes:** Our results are consistent with the hypothesis of a recent tetraploid origin for maize (Swigoňová *et al.* 2004). Although the two maize segments analyzed in this study only share Orp1 and Orp2 homoeologous genes, their comparisons to the orthologous regions of sorghum and rice indicate that they are two homoeologous segments. There have been at least two gene deletions near the Orp1 locus. Also, independent insertion of large blocks of retrotransposons in both Orp1 and Orp2 segments have occurred in the last few million years. The maize Orp2 segment remains relatively “intact”, with no gene deletion detected in this region. This result parallels the recent finding by Langham *et al.* (2004). By comparing the maize lg2 region and its homoeologous lrs1 region, Langham and coworkers found that a cluster of 4 predicted genes 3’ to the lrs1 locus have been deleted, leading to “zero retention” of duplicated factors excluding the lg2/lrs1 gene pair (Langham *et al.* 2004). In contrast, over 40% of the total genes from each homoeologous region were found to have been deleted by several separate deletion events in the maize adh1 region and its homoeologue, indicating that both regions have been equally unstable compared to their orthologs in sorghum and rice (Ilic *et al.* 2003). However, at least one
copy of all orthologous genes appears to be conserved between the two homoeologous regions in all cases investigated, suggesting that natural selection has acted against loss of all copies of any of these genes.

Timing of gene loss in the Orp region of maize: We cannot precisely determine the times of gene deletion or insertion events in the Orp1 region of maize, although our comparative data indicate that they took place after the divergence of maize and sorghum. The extensive deletion of genes and low-copy-number sequences appear to be a common feature of genomes with polyploid origins, such as Arabidopsis (The Arabidopsis Genome Initiative 2000) and maize (Ahn et al. 1993; Song et al. 2002; Ilic et al. 2003). The elimination of low-copy-number sequences has also been detected in newly formed polyploids (Song et al. 1995; Feldman et al. 1997; Ozkan et al. 2001), indicating that genome changes often happen in the first few generation in response to the formation of a polyploid. Gene deletion and transposon accumulation have also been seen to differentiate haplotypes in the allelic regions of different maize inbreds (Fu and Dooner 2002; Song and Messing 2003).

Genic rearrangements: deletion, insertion and/or translocation? Comparison of orthologous regions of the rice and sorghum genome reveals numerous small genic rearrangements. Apparent insertions of genes 1, 4 and 6 were detected in sorghum compared to rice and maize. No gene deletion or insertion was found in the two gene-clustered regions that are separated by a cluster of transposable elements in rice. This observation parallels observations in adh (Tikhonov et al. 1999; Ilic et al. 2003), sh2/a1
(Chen et al. 1997; Li et al. 2002), and php200725 (Song et al. 2002) -orthologous regions, indicating that rice has a relatively stable gene content and order compared with maize, sorghum, or wheat.

Interestingly, all of the non-colinear genes present in the Orp region of sorghum and/or maize were found to have very similar copy numbers in both rice and Arabidopsis, indicating copy number conservation for more than 150 million years of independent evolution (Wolfe et al. 1989). For four non-colinear genes, only single homologs were detected in both rice and Arabidopsis. If one assumes that these single-copy homologs are orthologous to the corresponding genes identified in sorghum, then it is clear that synteny or colinearity is not a perfect indicator of orthology. The relocations of these genes may have occurred in the rice and/or sorghum lineages. Alternatively, these four genes may be paralogous to the corresponding genes detected in rice because deletions removed the actual orthologs. Hence, it is impossible to say whether these genes were deleted, inserted, or relocated in the rice and sorghum genomes based on the current data.

All identified genes in the japonica Orp region were found to have homologs in the homologous region of indica rice. Because most assembled shotgun sequences from the indica genome are relatively small, we did not obtain the complete Orp region of indica, and thus cannot compare order or orientation of these sequence fragments. It is also not clear whether any genes are uniquely present in the indica region. However, complete japonica and indica sequences of the php200725 region show complete conservation of gene order in both subspecies (Song et al. 2002). Furthermore, previous comparison of
~1.1 Mb of orthologous regions between *indica* and *japonica* has demonstrated a lack of gene acquisition or loss from either *indica* or *japonica* (Ma and Bennetzen 2004), supporting the previous observation that the rice genome exhibits relatively stable gene content in contrast to the maize genome (Song *et al.* 2002; Ilic *et al.* 2003).

**A hotspot for gene rearrangement and the insertion of LTR-retrotransposons in rice:** We found a large LTR-retrotransposon-rich segment in the rice genome that contains few genes, and all of the genes within this retrotransposon block were either duplicates or non-colinear inserts relative to sorghum. Our data indicate that this rice region has expended rapidly by insertion of LTR-retrotransposons in the past 2 million years, with most insertions in the few hundred thousand years since the divergence of *indica* and *japonica* ancestors. The ancient insertions (> one million years old) in this region indicate that it has been a hotspot for transposon accumulation for a long time, while the recent insertions suggest that this insertion affinity is still present.

In our dating of relatively intact LTR-retrotransposons in the rice genome, we found that the average age is about 1.3 mya (Ma *et al.* 2004), while in the retrotransposon block of rice Orp region, the average age of all datable LTR-retrotransposons is ~0.7 mya. Moreover, we demonstrated a minimum of eight new transposon insertions within the *japonica* region since the divergence from a common ancestor with *indica*, adding at least 53 kb of new DNA to a target region of 190 kb. This is about 4-fold higher frequency of insertion than observed for 1 Mb of chromosome 4 DNA from our earlier *indica* and *japonica* comparison (Ma and Bennetzen, 2004).
A high percentage of repetitive DNA was observed in the centromeric region of rice chromosome 8. In this region, LTR-retrotransposons account for at least 50% of the DNA (Wu et al. 2004), and more than 80% of these elements were amplified before the divergence of indica and japonica (Ma and Bennetzen, unpub. obs.) In contrast to the centromeric region, 55% of the transposon-rich segment of the rice Orp region is composed of LTR-retrotransposons, and about half of them were amplified after the divergence of these two subspecies (Figure 2). This high rate of transposon insertion, plus the presence of a non-tandem gene triplication and several non-collinear truncated genes in the Orp region, suggests that this block is a hotspot for several different kinds of genome rearrangement. It will be interesting to see if other retrotransposon blocks exhibit this type of exceptional instability when other comparative studies are performed.

An intra-element retrotransposon conversion event. While maize retrotransposons are frequently intact at their termini—including the presence of short flanking host site duplications, there is no shortage of more tattered elements present in any maize BAC sequence. Unequal recombination is frequently invoked to explain the presence of solo-LTRs (Devos et al. 2002; Ma et al. 2004). Here we suggest that this phenomenon may explain a larger group of rearrangements simply by positing that an unequal recombination event initiating inside LTRs might migrate outside a terminus of these LTRs. While our diagram (Figure 4) depicts a recombination event with symmetric exchange of strands, non-symmetric events should also occur. These would yield the same outcome. Repair of heteroduplex DNA will also play a role in these sorts of
recombination events and this could result in more complex rearrangements than depicted if the repair was non-continuous over the recombination tract.

One other model could be proposed to explain the structure we found. Two grande elements (most likely proximate to one another) on the same chromosome in the same orientation could recombine unequally to create a double element, sharing an LTR. But a second event would be required to explain the deletion of the 5’ end of the 5’ element. This second model is also unlikely because the duplicated region resulting from this mechanism would be likely to have a greater percent of mismatched bases over the duplication than the 3% that is observed. Comparison of a grande element from the 22-kD alpha zein gene family (Song et al. 2001) and this grande element yields a 15% mismatch frequency over aligned bases. Rarely are retrotransposons (even from the same family) more than 90% similar over greater than 2 kb regions.

This work was supported by the National Science Foundation Plant Genome Program (grant no. 9975618). We wish to thank Katrien M. Devos and Zuzana Swigoňová for useful discussions, and two anonymous reviewers for their valuable comments.

LITERATURE CITED


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.


Li, W., and B. S. Gill, 2002 The colinearity of the Sh2/A1 orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the triticeae. Genetics 160: 1153-1162.


Ramakrishna, W., J. Dubcovsky, Y.-J. Park, C. Busso, J. Emberton et al., 2002a Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 162:1389-1400.

Ramakrishna, W., J. Emberton, P. SanMiguel, M. Ogden, V. Llaca et al., 2002b Comparative sequence analysis of the sorghum Rph region and the maize Rp1 resistance gene complex. Plant Physiol. 130: 1728-1738.


Yan, L., A. Loukoianov, A. Blechl, G. Tranquilli, W. Ramakrishna et al., 2004 The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science.303: 1640-1644.


### TABLE 1

Identified genes in *Orp* regions of maize, sorghum and rice

<table>
<thead>
<tr>
<th>Gene</th>
<th>Rice</th>
<th>Sorghum</th>
<th>Maize</th>
<th>Maize</th>
<th>Homology(^a)</th>
<th>Protein products</th>
<th>Acc. No.</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Expressed protein (<em>Arabidopsis</em>)</td>
<td>NP_188808</td>
<td>2e-6</td>
</tr>
<tr>
<td>2</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
<td>Homeobox protein (<em>Arabidopsis</em>)</td>
<td>NP_193906</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>Tryptophan synthase beta-subunit (<em>Arabidopsis</em>)</td>
<td>NP_194437</td>
<td>2e-87</td>
</tr>
<tr>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td>F box protein (<em>Arabidopsis</em>)</td>
<td>NP_191482</td>
<td>2e-6</td>
</tr>
<tr>
<td>5</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>Hypothetical protein (<em>Arabidopsis</em>)</td>
<td>NP_188494</td>
<td>2e-14</td>
</tr>
<tr>
<td>6</td>
<td>√</td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td>Hypothetical protein (<em>Arabidopsis</em>)</td>
<td>NP_200360</td>
<td>2e-24</td>
</tr>
<tr>
<td>7</td>
<td>√</td>
<td></td>
<td>√</td>
<td>√</td>
<td></td>
<td>Endonuclease/exonuclease/phosphatase family (<em>Arabidopsis</em>)</td>
<td>NP_566904</td>
<td>2e-24</td>
</tr>
<tr>
<td>8</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
<td>Expressed protein (<em>Arabidopsis</em>)</td>
<td>NP_192387</td>
<td>1e-9</td>
</tr>
<tr>
<td>9</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
<td>Glycosyl hydrolase family 17 (<em>Arabidopsis</em>)</td>
<td>NP_181895</td>
<td>5e-21</td>
</tr>
<tr>
<td>10</td>
<td>√</td>
<td></td>
<td>√</td>
<td>√</td>
<td></td>
<td>Copine-related protein (<em>Arabidopsis</em>)</td>
<td>NP_565206</td>
<td>2e-98</td>
</tr>
<tr>
<td>11</td>
<td>√</td>
<td></td>
<td>√</td>
<td>√</td>
<td></td>
<td>Hypothetical protein (<em>Arabidopsis</em>)</td>
<td>NP_187362</td>
<td>6e-19</td>
</tr>
<tr>
<td>12(^b)</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>Transporter related protein (<em>Arabidopsis</em>)</td>
<td>NP_566487</td>
<td>1e-70</td>
</tr>
<tr>
<td>13</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hypothetical protein (<em>Arabidopsis</em>)</td>
<td>NP_201419</td>
<td>6e-13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----------------------------------------------------------------</td>
<td>-------</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>√</td>
<td></td>
<td>phospholipid/glycerol acyltransferase family (<em>Arabidopsis</em>)</td>
<td>NP_181346</td>
<td>4e-89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>√</td>
<td></td>
<td>Protein phosphatase 2C (<em>Arabidopsis</em>)</td>
<td>NP_194903</td>
<td>5e-20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>√</td>
<td></td>
<td>Hypothetical protein (<em>Arabidopsis</em>)</td>
<td>NP_173847</td>
<td>4e-47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>√</td>
<td>√</td>
<td>Galactosyltransferase family (<em>Arabidopsis</em>)</td>
<td>NP_177618</td>
<td>9e-56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>√</td>
<td>√</td>
<td>Hypothetical protein (<em>Arabidopsis</em>)</td>
<td>NP_175765</td>
<td>e-155</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>√</td>
<td>√</td>
<td>Cytochrome P450 protein (<em>Arabidopsis</em>)</td>
<td>NP_171635</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>√</td>
<td>√</td>
<td>Putative lip transfer protein precursor (<em>Arabidopsis</em>)</td>
<td>NP_179109</td>
<td>1e-13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>√</td>
<td>√</td>
<td>Phototropic response protein family (<em>Arabidopsis</em>)</td>
<td>NP_174332</td>
<td>e-151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>√</td>
<td>√</td>
<td>Expressed protein (<em>Arabidopsis</em>)</td>
<td>NP_193039</td>
<td>8e-17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>√</td>
<td></td>
<td>Expressed protein (<em>Arabidopsis</em>)</td>
<td>NP_564055</td>
<td>1e-24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>√</td>
<td></td>
<td>Copine BONZAI1 (BON1) (<em>Arabidopsis</em>)</td>
<td>NP_568944</td>
<td>3e-11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>√</td>
<td></td>
<td>Putative protein (<em>Arabidopsis</em>)</td>
<td>NP_198805</td>
<td>5e-19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>√</td>
<td>Tubby-like protein (<em>Arabidopsis</em>)</td>
<td>NP_849975</td>
<td>2e-11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The “√” indicates the genes listed are present in the corresponding regions of maize, rice, and/or sorghum.
### TABLE 2

Chromosomal distribution of homologs of investigated genes in rice and *Arabidopsis*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Orp region</th>
<th>Copy no.</th>
<th>Chr.</th>
<th>Locus</th>
<th>E value</th>
<th>Chr.</th>
<th>Locus</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sorghum</td>
<td>1</td>
<td>7</td>
<td>2009.t00001</td>
<td>4e-4</td>
<td>1</td>
<td>3</td>
<td>At3g21710.1</td>
</tr>
<tr>
<td>4</td>
<td>Sorghum</td>
<td>&gt;2</td>
<td>6</td>
<td>3460t000003</td>
<td>3e-120</td>
<td>&gt;2</td>
<td>3</td>
<td>At3g59230.1</td>
</tr>
<tr>
<td>6</td>
<td>Sorghum</td>
<td>2</td>
<td>9</td>
<td>5014t000010</td>
<td>8e-104</td>
<td>1</td>
<td>5</td>
<td>At5g55490.1</td>
</tr>
<tr>
<td>7</td>
<td>Sorghum</td>
<td>1</td>
<td>12</td>
<td>5048.t00006</td>
<td>8e-125</td>
<td>1</td>
<td>3</td>
<td>At3g48425.1</td>
</tr>
<tr>
<td>13</td>
<td>Sorghum</td>
<td>1</td>
<td>2</td>
<td>5875t000005</td>
<td>8e-127</td>
<td>1</td>
<td>5</td>
<td>At5g66180.1</td>
</tr>
<tr>
<td>15</td>
<td>Sorghum</td>
<td>&gt;2</td>
<td>2</td>
<td>5021t000003</td>
<td>2e-46</td>
<td>&gt;2</td>
<td>4</td>
<td>At4g31750.1</td>
</tr>
<tr>
<td>16</td>
<td>Sorghum</td>
<td>&gt;2</td>
<td>4</td>
<td>5483t00011</td>
<td>3e-87</td>
<td>&gt;2</td>
<td>1</td>
<td>At1g24370.1</td>
</tr>
<tr>
<td>22</td>
<td>Sorghum</td>
<td>1</td>
<td>5</td>
<td>6505t00025</td>
<td>7e-74</td>
<td>1</td>
<td>4</td>
<td>At4g13030.1</td>
</tr>
<tr>
<td>d maize</td>
<td>&gt;2</td>
<td>2</td>
<td>4877t00020</td>
<td>1e-15</td>
<td>&gt;2</td>
<td>2</td>
<td>At2g18280.1</td>
<td>6e-13</td>
</tr>
</tbody>
</table>
### TABLE 3

**Assessments of creditability of gene prediction methods**

<table>
<thead>
<tr>
<th>Region</th>
<th>Transposable element&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Identified gene&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Unknown sequence&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shared</td>
<td>Unshared</td>
<td>Shared</td>
<td>Unshared</td>
</tr>
<tr>
<td>Rice</td>
<td>28</td>
<td>14</td>
<td>3+2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Sorghum</td>
<td>2</td>
<td>15</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Maize Orp2</td>
<td>15</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maize Orp1</td>
<td>27</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup>Predicted genes belonging to transposable elements identified in this study or matching to that deposited in the public databases.

<sup>b</sup>All of these genes were found to have homologs in the *Arabidopsis* whole genome database.

<sup>c</sup>Genes were only predicted by gene finding programs.

<sup>d</sup>Two duplicated genes that were predicted to be derived from gene 12-1, a possible ortholog of sorghum gene 12.
FIGURE LEGENDS

Figure 1. Sequence organization and comparison of orthologous Orp regions of sorghum, maize and rice. Arrows represent identified genes and their transcriptional orientations. The orthologous or homoeologous genes shared by different regions are connected by gray lines. The organizations of retrotransposons and transposons are illustrated by gray shaded boxes and cross-hatched boxes, respectively. Dotted lines in rice connect contiguous sequences. –f, truncated transposon fragment; -s, solo-LTR.

Figure 2. Comparison of orthologous regions of indica and japonica. Arrows represent predicted genes. Gray shaded boxes and cross-hatched boxes represent retrotransposons and DNA transposons, respectively. The shaded regions connecting the japonica and indica sequences outline the conserved regions between these two subspecies. Because the indica sequence is fragmentary, we cannot compare the order or orientation of the sequences on different indica fragments with the ordered sequence for japonica.

Figure 3. Estimated insertion times of identified LTR-retrotransposons in the Orp regions of maize and rice. The dates of LTR-retrotransposon insertions are shown in brackets. Mya, million years ago.

Figure 4. Unequal homologous recombination (without exchange of flanking markers) may have replaced the 5’ flanking region of grande_573F08-1 with 2145 bases of its internal sequence. This figure is not drawn to scale. Both strands of unequally
paired regions from homeologous chromosomes are shown in panels A and B. Hollow arrows denote LTRs. The element is arbitrarily divided into a green, horizontally striped 5’ region and a purple, vertically lined 3’ region. Upstream and downstream flanking DNA is depicted as solid and diagonally striped lines respectively. A. Putative initial structure of the grande element prior to recombination. The 3’ LTR of the top homologue unequally pairs with the 5’ LTR of the bottom homologue. The red “x” shows the site of initiation of recombination and the red arrow shows the direction of the hypothetical strand transfer. B. Branch migration progresses beyond the 5’ termini of the LTRs in which the recombination initiated. C. One of many possible outcomes of this recombination event is shown, the one that is the structure present in 573F08. Two landmark sequences associated with LTR-retrotransposon transposition, PBS (primer binding site) and PPT (polypurine tract), are indicated.
Figure 2
Figure 3

![Graph showing the number of LTR-retrotransposons inserted at different times in Maize (Orp1), Maize (Orp2), and Rice (Orp).]
Figure 4

A

B

C