Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss

James C. Schnable[a], Nathan M. Springer[b], and Michael Freeling[a,1]

[a]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; and [b]Department of Plant Biology, University of Minnesota, St. Paul, MN 55108

Contributed by Michael Freeling, January 26, 2011 (sent for review November 26, 2010)

Ancient tetraploidies are found throughout the eukaryotes. After duplication, one copy of each duplicate gene pair tends to be lost (fractionate). For all studied tetraploidies, the loss of duplicated genes, known as homeologs, homoeologs, ohnologs, or syntenic paralogs, is uneven between duplicate regions. In maize, a species that experienced a tetraploidy 5–12 million years ago, we show that in addition to uneven ancient gene loss, the two complete genomes contained within maize are differentiated by ongoing fractionation among diverse inbreds as well as by a pattern of overexpression of genes from the genome that has experienced less gene loss. These expression differences are consistent over a range of experiments quantifying RNA abundance in different tissues. We propose that the universal bias in gene loss between the genomes of this ancient tetraploid, and perhaps all tetraploids, is the result of selection against loss of the gene responsible for the majority of total expression for a duplicate gene pair. Although the tetraploidy of maize is ancient, biased gene loss and expression continue today and explain, at least in part, the remarkable genetic diversity found among modern maize cultivars.

genome evolution | paleopolyploidy | synteny

Genomes that have experienced ancient polyploidy show nonequivalence between duplicated genomic regions. The most easily observed aspect of this nonequivalence is that one copy of a duplicated region will retain more genes, whereas the other copy of that same region will lose more genes, a phenomenon known as fractionation bias. This bias in gene loss and retention between duplicated genome segments has been observed in *Arabidopsis* (1) and, more recently, in maize (2, 3) and is probably a general characteristic of posttetraploid eukaryotic genomes (4). Although the proximate mechanism of gene loss following the whole-genome duplication in maize has been shown to be a short deletion mechanism (2), this mechanism does not explain why genes from one genome segment should be more likely to be lost than their homoeolog (homeologs, ohnologs, and syntenic paralogs are synonyms) in the duplicate region of the genome.

A second form of nonequivalence between duplicated regions, in fact between whole genomes, has been shown in studies of more recent allotetraploid species. Wang et al. (5), in the laboratory of Z. J. Chen, used 70-mer oligo microarrays to measure gene expression differences in a synthetic allotetraploid of *Arabidopsis thaliana* and *Arabidopsis arenosa* and compared these results with midpoint values of gene expression in the two parents. They showed that genes originating from *A. arenosa* tend to dominate over homoeologous genes from *A. thaliana* by contributing more to total gene expression in the allotetraploid. The same pattern of genome dominance was observed for the recent natural allotetraploid *Tragopogon miscellus*, a species estimated to have originated less than 80 y ago (6). The laboratories of W. B. Barbazuk and D. E. and P. S. Soltis sequenced leaf RNA from *T. miscellus* and found that the higher expressed members of differential expressed gene pairs were more likely to carry SNPs shared with *Tragopogon dubius* than with the other diploid parental species, *Tragopogon pratensis* (7). Tetraploid cotton species originated in an allotetraploid event between diploid species carrying A and D genomes with an estimated age of 1 and 2 million years (8). Data from these species provide evidence that genome dominance persists over much longer time scales. L. E. Flagel and J. F. Wendel used petal RNA hybridized to microarrays with probes specific to genes originating in the A or D cotton genome to show that although many gene pairs are expressed contrary to the prevailing pattern, genes originating in the D genome are more likely to contribute a majority of total gene expression than their homoeologs from the A genome in five allotetraploid cotton species and a synthetic hybrid between diploid cotton species containing the A and D genomes (9).

Genome dominance has not been observed in studies of any of the more ancient plant tetraploidies. Studies of the expression patterns of homoeologous gene pairs originating from the *Arabidopsis* α-tetraploidy, estimated to have occurred 25–40 million years ago, found no systematic pattern of dominant expression (10). Similarly studies of gene expression patterns across homoeologous regions in rice, originating from a duplication estimated to have occurred 50–70 million years ago (11), report no evidence of genome dominance (12). It appears that homoeologous gene pairs in both rice and *Arabidopsis* are often differentially expressed (10, 13). It should be noted that the ability of these studies to resolve subtle differences was limited by the inability to assign duplicated segments to specific ancestral genomes; thus, analyses were carried out on individual homoeologous segments.

We use comparative analysis of the maize and sorghum genomes to examine the differentiation of duplicated genomic regions following the maize tetraploidy. Both grass species are members of the tribe Andropogoneae, and the genomes of both species have been sequenced (3, 14). The lineage leading to maize experienced tetraploidy sometime after the divergence of the two lineages, whereas sorghum remained diploid. An unduplicated outgroup is essential for identifying highly fractionated duplicate genome segments as well as for differentiating between recently transposed genes and genes lost from one duplicated segment but retained in the other (15). The two genomes of maize split from each other ∼12 million years ago, contemporaneous with but following the split between the maize and sorghum lineages, as the result of either autotetraploidy or allotetraploidy (16). The maize tetraploidy, which combined both genomes within one nucleus and began the process of genome fractionation, occurred between 5 and 12 million years ago (3, 16). The genome of maize shows evidence of ongoing gene loss (2), making it an excellent model to study the mechanism of differentiation between duplicated genomic regions.

We show that fractionation bias results from the differentiation of entire ancestral chromosomes and suggest that this chromosomal differentiation reflects differences between the two pa-

Author contributions: J.C.S. and M.F. designed research; J.C.S. performed research; N.M.S. contributed new reagents/analytic tools; J.C.S. analyzed data; and J.C.S. and M.F. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: freeling@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1101368108/-/DCSupplemental.

rental genomes, with one genome being dominant at the level of gene deletion resistance and RNA expression. Biased loss of genes does not appear to be a result of inherent differences in deletion rates between homoeologous regions, because "silent" deletions, deletions in DNA that are usually without specific function, such as those from introns and retrotransposons, show no bias between ancestral chromosomes. Given the correlation observed between the subgenome, which dominates expression in maize, and the ancestral chromosomes, which have experienced less gene deletion, we propose that deletions of duplicate genes from the less frequently expressed subgenome may be less likely to result in reduced fitness. This hypothesis makes sense in light of the gene balance hypothesis, as will be discussed. Following tetraploidy, deletions from one subgenome would be more likely to be removed by purifying selection, whereas deletions from the opposite subgenome would be more likely to be selectively neutral.

Results

Reconstruction of Chromosome Level Organization in the Newly Tetraploid Ancestor of Maize: Defining Two Subgenomes. It was inferred from multiple studies that the ancestral genome of the Andropogoneae consisted of 10 chromosomes. The genome of sorghum is presumed to have approximately retained this ancestral arrangement, whereas the 10 chromosomes of maize represent a reduction from a 20-chromosome tetraploid ancestor by chromosome fusion (17, 18). Given the small total divergence time between maize and sorghum and the fact that tetraploidy can temporarily increase the frequency of genome rearrangements (19), the sorghum genome was treated as representative of the genome organization of both diploid genomes present in the initial tetraploid ancestor of maize.

Using whole-genome dot plots color-coded by synonymous base pair substitution rates (Fig. 1, plotted using CoGe software), it is possible to reconstruct the original duplicate regions within the maize genome on the basis of orthology to the 10 sorghum chromosomes (*SI Appendix*, Table 1). The synonymous substitution rates of individual gene pairs do not permit genes to be unambiguously classified as orthologs or ancient homoeologs. The median synonymous substitution rate of all gene pairs in a syntenic block between maize and sorghum can be used to classify syntenic blocks of 12 or more genes unambiguously as orthologous or homoeologous, however (Fig. 1 and *SI Appendix*, Fig. 1).

Inversions and other intrachromosomal rearrangements are presumed to be more common than translocations between different chromosomes. Therefore, segments of a maize chromosome orthologous to the same sorghum chromosome are assumed to come from the same chromosome copy in the tetraploid ancestor maize. For five sorghum chromosomes at least, both full ancestral copies can be reconstructed in the maize genome using this method. For the remaining five, one full ancestral copy was reconstructed based on all orthologous segments being present on a single maize chromosome and the remaining orthologous segments located on two or, in one case, three maize chromosomes were grouped together by process of elimination (*SI Appendix*, Table 1). There are no cases in which both duplicate copies of the region were located on the same chromosome. Our assumptions and reconstruction are largely concordant with previous ancestral reconstructions of the maize genome (17, 18).

For each pair of reconstructed chromosomes, one copy retained substantially more syntenic genes than the other. Bias in gene loss between pairs of reconstructed chromosomes was consistent across their entire lengths (Fig. 2). For each pair of chromosomes, the copy that possessed a greater number of unique genes retained orthologously in both rice and sorghum was assigned to the maize1 subgenome, whereas the pair with fewer uniquely retained genes was assigned to the maize2 subgenome. Gene counts and the statistical significance of the differences between copies are listed in *SI Appendix*, Table 1. Individual maize genes or gene pairs and their identified orthologs in rice and sorghum are listed in *Dataset S1*. Maize1 and maize2 each constitute a genome orthologous to the entire sorghum genome. The distribution of these two genomes across the 10 modern chromosomes of maize is displayed in *SI Appendix*, Fig. 2.

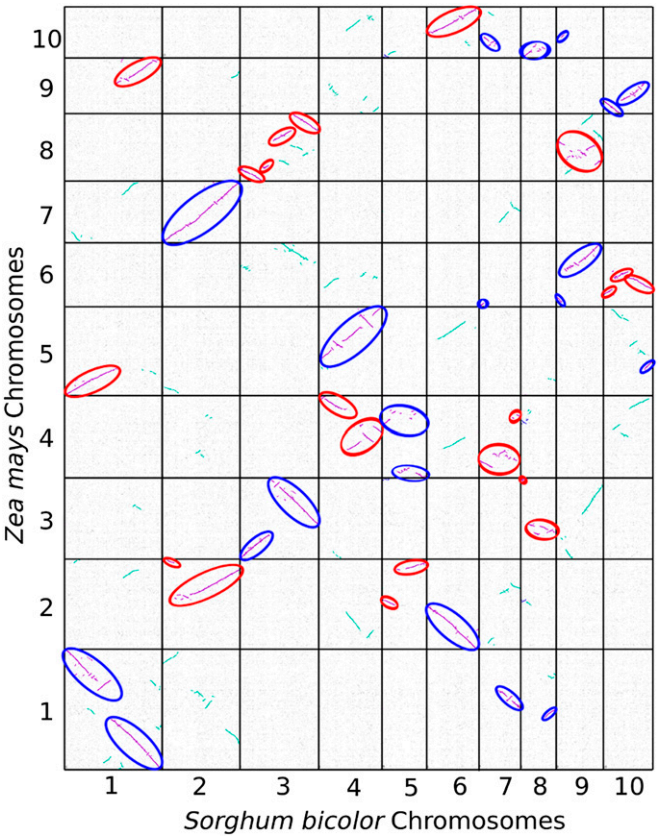

Fig. 1. A dot plot comparison of the maize and sorghum genomes. Each dot marks a pair of genes, one in sorghum and one in maize, identified as homologs in a blast comparison. Genes with conserved syntenic gene order are highlighted in color. Orthologs from the maize sorghum split were distinguished from homoeologs from the pregrass duplication by the synonymous substitution rate (Ks). Orthologs are marked in purple (lower Ks), and pregrass homoeologs are marked in teal (higher Ks). The regions making up one complete ortholog of each sorghum chromosome in the maize genome are circled in blue, and the regions making up the other complete ortholog are circled in red. The original dot plot from which this figure was created was produced using CoGe software and can be regenerated at http://tinyurl.com/2am77tn by clicking "Generate SynMap."

Ongoing Fractionation Among 33 *Zea mays* Accessions Remains Biased. Using only maize genes with retained syntenic orthologs in both sorghum and rice, we constructed two lists of high-confidence genes, the list of retained homoeologs from the maize duplication and the list of genes for which it was possible to say with high confidence that the duplicated copy was lost from the genome (singleton genes). These lists will be referred to as "retained homoeolog" and "lost homoeolog." Each of these gene lists is further subdivided into maize1-specific and maize2-specific lists of genes. A complete description of the criteria used to identify these two high-confidence gene sets is included in *Methods*. There is no significant difference in the annotated length of coding or noncoding sequences between homoeologous copies of genes retained in both maize1 and maize2 subgenomes (*SI Appendix*, Fig. 3).

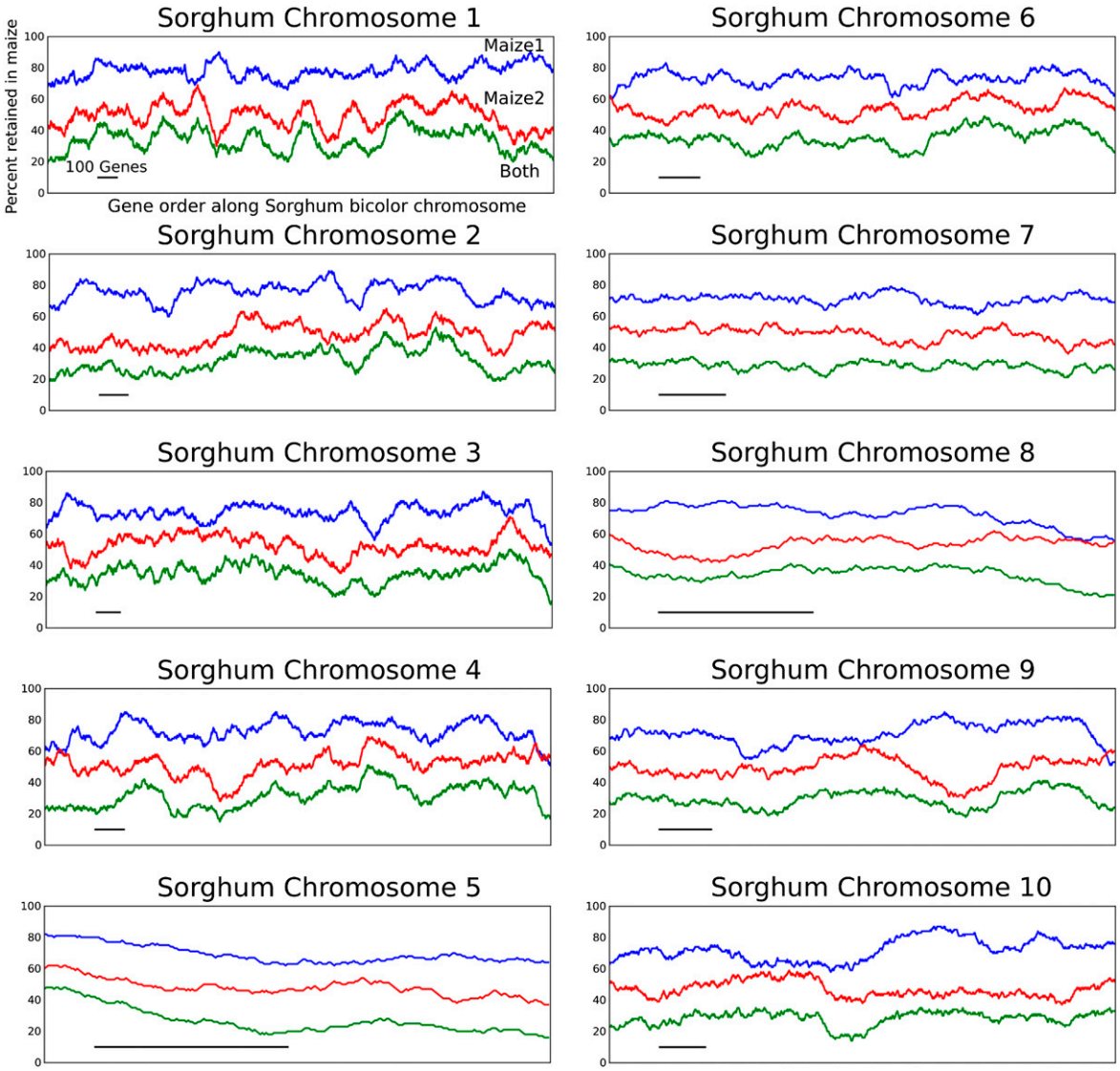

Fig. 2. Biased fractionation is observed for each reconstructed, or "sorghumized," pair of maize ancestral chromosomes. Bias is measured as the number of conserved genes out of 100 in a sliding window (black bars) of genes conserved syntenically between sorghum and rice (*y* axis) and is displayed based on the gene order along sorghum chromosomes (*x* axis). Conservation of genes on reconstructed chromosomes assigned to maize1 is shown in blue. Conservation of genes on reconstructed chromosomes assigned to maize2 is shown in red. The proportion of genes retained on both reconstructed chromosomes is shown in green.

A recently published dataset documents the presence/absence variation (PAV) of genes among 19 diverse maize inbreds and 14 teosinte lines using carefully controlled comparative genomic hybridization (20). Among our high-confidence lost homoeolog gene sets, equal percentages of maize1 and maize2 genes were identified as lost from the genomes of one or more inbreds. Among our high-confidence retained homoeolog gene set, however, significantly more of the genes located on maize2 were identified as lost from one or more inbreds than were the duplicate copies of those same genes located on maize1 ($P = 0.0043$, χ^2 test; df = 2) (Fig. 3*A*). PAV data indicate that ongoing fractionation remains biased in modern maize inbreds.

Maizesequence.org has released at least two sets of gene annotations. The filtered gene set (FGS) contains ~32,000 genes considered to be of higher confidence, whereas the working gene set (WGS) contains over 100,000 genes, including the genes of the FGS as well as many likely pseudogenes, gene fragments, or transposon-related proteins. Genes unique to the WGS have a similar distribution to those genes that show PAV between maize inbreds. Ongoing fractionation by short deletions has been shown to produce truncated gene fragments before their complete removal (2), exactly the sort of sequence that might be annotated as a gene but excluded from the FGS. The distribution of genes found only in the maize WGS supports the conclusion that biased fractionation in the maize genome is ongoing. First, syntenically retained working set genes are more likely to possess a retained homoeolog, which is presumably the undamaged full-length gene copy (Fig. 3*B*). Second, in these cases, the low-confidence gene found only in the WGS is more likely to be the copy located in the maize2 subgenome (Fig. 3*B*). The work described in this section is the only portion of our study in which we did not exclude the low-confidence genes found only in the WGS.

Deletions Within Noncoding Sequences Show No Bias Between Maize1 and Maize2. Maize1 and maize2 subgenomes cover significantly different fractions of the total maize genome: 1.26 gigabases and 0.75 gigabases, respectively. Because coding sequences of annotated genes, including the WGS, account for less than 5% of

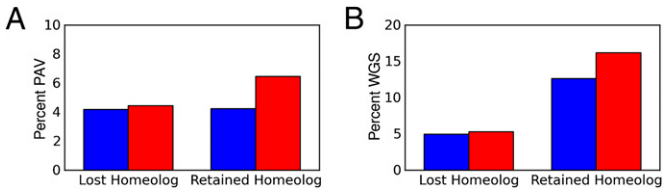

Fig. 3. Multiple measures of ancient and ongoing fractionation. (*A*) Percentage of high-confidence maize genes (*Methods*) that exhibited PAV in a study of maize inbreds and teosinte accessions. (*B*) Percentage of all annotated maize genes conserved syntenically in both rice and sorghum that are excluded from the maize FGS.

the total maize genome and transposons account for 85% (3), this bias in total genomic size would seem to imply that biased fractionation acts on all genomic DNA and not simply on coding sequences. The lengths of both coding sequences and noncoding sequences in high-confidence retained homoeologous pairs on maize1 and maize2 are not significantly different, however (*SI Appendix*, Fig. 3). An analysis of 561 maize1 and maize2 introns that could be completely aligned to the orthologous sorghum intron identified an average of 6.03 deletions per intron in maize1 genes and 6.09 deletions per intron in the homoeologous maize2 genes (*SI Appendix*, Table 2). A similar analysis of deletions within copies of three of the largest families of retrotranposons within the maize genome, Huck, Opie, and Ji, which had inserted into the maize1 or maize2 region of the genome found no difference in deletion frequencies for maize1 vs. maize2 relative to an ancestral sequence for each family created from an alignment of multiple annotated transposon copies (*SI Appendix*, Fig. 4 and Table 3).

Expression Differences Between Maize1 and Maize2 Homoeologous Genes. Gene expression was measured for all genes included in the maize WGS from the sequenced maize inbred B73 (3) and RNA-Seq data from four independent previously published datasets (21–24) (*SI Appendix*, Table 4). Expression data were calculated in units of frequency of aligned reads per kilobase of exon per million reads (RPKM) using the Bowtie (25) and Cufflinks (26) packages. Cufflinks distributes reads that were found to be aligned equally well to multiple gene models proportional to the relative expression rates for those genes calculated from reads with only one best alignment (26) This combination of programs allows us to deal with the ambiguity created by the small fraction of sequences that align equally well to both homoeologs within the maize genome.

The expression of gene pairs included in the high-confidence retained homoeolog set described above was compared using each expression dataset. In each dataset, the number of pairs in which the maize1 homoeolog dominated total gene pair expression outnumbered the number of pairs in which the maize2 homoeolog dominated expression. This bias was robust, appearing whether we defined dominance as any measurable difference in expression (*SI Appendix*, Fig. 5), at least a twofold difference in homoeolog expression (Fig. 4), or a fourfold difference in homoeolog expression (*SI Appendix*, Fig. 6). The bias toward gene pairs dominated by expression of the maize1 copy remains consistent across a range of cutoffs for the expression of the nondominant homoeolog. At cutoffs as high as 30 RPKM for the less frequently expressed gene copy, maize1 homoeologs continued to disproportionately dominate expression in all parts of the maize plant examined (*SI Appendix*, Fig. 7). Biased expression is also observed when examining individual pairs of reconstructed chromosomes (*SI Appendix*, Fig. 8), which are effectively independent replicates of our experiment. The median difference in expression between homoeologs ranges from 1.8- to 2.8-fold in different expression datasets. In every expression dataset, the median difference between homoeologs in which the maize1 gene is expressed at a higher level is marginally higher than the median difference for the pair in which maize2 is expressed at a higher level (*SI Appendix*, Table 5).

Discussion

Biased gene loss is clearly not a transient phenomenon that occurred only in the early generations following tetraploidy in maize. Rather, biased gene loss is a reflection of a significant differentiation of two complete subgenomes within a tetraploid lineage, and these differences are stably inherited over millions of generations. The link we observe between the biased gene loss and biased expression is likely not unique to the maize tetraploidy. A recent study of a 1-megabase region of the common bean (*Phaseolus vulgaris*) and the two co-orthologous regions of the soybean genome also found that the homoeologous region with more syntenically retained genes tended to be expressed at higher levels (27). Although we have shown that bias in the loss of duplicate gene copies continues in the maize lineage, as it presumably has for the past several million years, evidence from deletions in introns and retrotransposons suggests that this bias is not the result of fundamentally different frequencies of sequence deletion between maize1 and maize2 chromosomal segments. The equivalent deletion rates we observe for both subgenomes is concordant with our finding that single-copy genes on either subgenome are equally likely to be identified as showing PAV between inbreds.

Our data suggest a model in which deletions in both maize genomes occur at the same overall rate, but purifying selection is

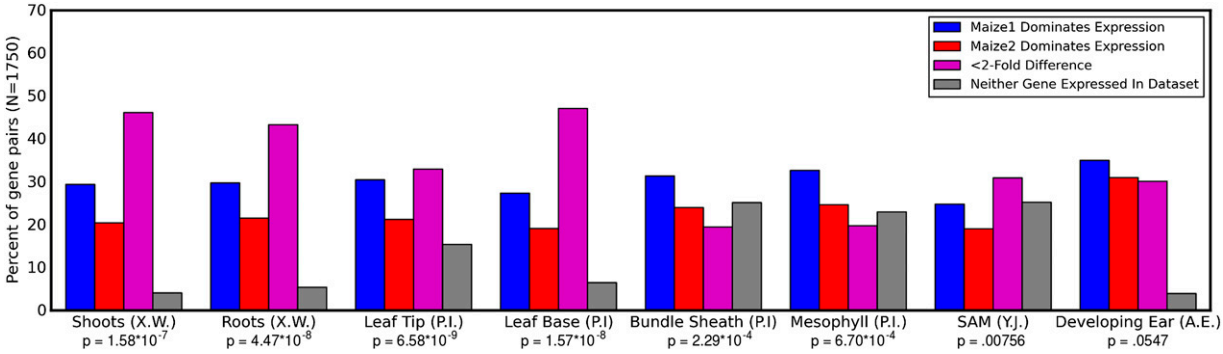

Fig. 4. Patterns of expression for the 1,750 best-confidence (*Methods*) pairs of maize homoeologs in eight organ systems, organs, or cell types. Homoeologs were considered to be differentially expressed if the expression of one homoeolog was at least twice the expression of the other. RNA-Seq data were from Wang et al. (21), Li et al. (22), Jia et al. (23), and Eveland et al. (24). All *P* values were calculated using cumulative binomial distributions assuming an equal chance of gene copies on maize1 or maize2 dominating total expression for the gene pair.

more likely to remove deletion alleles of higher expressed duplicate copies from the population, whereas the loss of less frequently expressed homoeologs is more likely to be selectively neutral or near-neutral when the higher expressed copy remains present in the genome. This model is consistent with selection against changes in the balance of gene products, as reviewed elsewhere (28–31). Our model states that smaller changes in total gene pair expression (maize1 transcript + maize2 transcript) are more likely to be tolerated than larger changes. The removal of a singleton gene, whether it is located in maize1 or maize2, involves the complete loss of that gene product. Because the effect of the loss of a singleton gene would be the same regardless of genomic location, no bias would be predicted for these genes, and we detected no bias. Our control experiments showing that deletions within transposons and most deletions within introns are unbiased between maize1 and maize2 demonstrate that maize1 and maize2 have no inherent difference in mutability. This result is consistent with our model that biased fractionation is a result of purifying selection acting preferentially against deletion alleles of gene copies that contribute more to total gene pair expression.

There are precedents for the idea that changes in total gene product dosage often lower fitness. Genes encoding proteins with more interaction partners, such as protein kinases and phosphatases, or subunits of complex machines, such as ribosomes, proteasomes, and motors, are predicted (32) to be more dosage-sensitive, and these are precisely the classes of genes that are more likely to be retained as homoeologous pairs following tetraploidy (1, 10, 33, 34). Greater changes in total gene product dosage have also been shown to be more likely to have a negative impact on fitness in the absence of tetraploidies. For example, the loss of highly expressed gene copies in yeast has been shown to be more likely to have a significant impact on fitness than the loss of their less frequently expressed paralogs (35). Knockouts of duplicate genes in yeast with similar levels and patterns of expression, those presumed to be the most dose-sensitive, have been shown to share similar patterns of epistatic relations, demonstrating that the loss of either equally expressed duplicate gene has an impact on function in a similar way (36).

Although the maize lineage tetraploidy occurred 5–12 million years ago, the latest transposon blooms in maize occurred only in the past few million years (3, 37). It is conceivable that the gene contents of maize1 and maize2 genomes were already significantly different at the time of this most recent transposon bloom. Opie and Ji have both been preferentially shown to insert into heterochromatin near genes (38), suggesting that transposons insertions will tend to track total gene content over time. We hypothesize that transposons inserted into maize1 and maize2 in approximate proportion to the gene content of these regions. If this were indeed the case, the difference in mobile dispensable DNA between the two genomes is simply an artifact of preexisting differences in gene content. Further experiments are necessary to evaluate fully the degree to which selection can explain the many differences between the two maize genomes, but it is remarkable that selection frequently differentiates between relatively minor levels of gene expression. The general concept of expression thresholds, so common in discussions of allelic dominance and recessiveness, has not proven useful in interpreting our data.

The explanation of biased fractionation by genome dominance leaves unanswered the question of the mechanism behind the origin and maintenance of genome dominance. The most likely candidate remains differential epigenetic marking of genomes within an allotetraploid. Allotetraploidy has been shown to produce epigenetically inherited differentiation of parental genomes (5, 39, 40). There is no conclusive evidence to support either an auto- or allotetraploid origin for maize, although one study found that ZFL2 may be more closely related to orthologs in the Andropogoneae genera *Coelorachis* and *Elionurus* than to the duplicate homoeolog in maize ZFL1 (41). Although there is currently a dearth of high-quality epigenetic data for maize available in published literature, ongoing research projects are likely to remedy this situation in the near future, thereby illuminating the mechanism responsible for differentiation of maize1 and maize2 gene copies.

Whatever the mechanism, an event occurred early in the process of tetraploidy that differentiated the two parental genomes of maize, maize1 and maize2. We have shown that these differences have persisted through millions of generations and continue to have an impact on both gene expression and the pattern of ongoing gene loss in maize. Ongoing fractionation by the mechanism we describe here provides an explanation as to why *Zea mays* is particularly genetically diverse.

Methods

Identification of Orthologous and Homoeologous Genes. Syntenic blocks were identified between and within grass genomes using the SynMap application within CoGe, an online comparative genomic toolbox (42). Syntenic blocks were assigned to specific evolutionary events, either speciation (orthology) or whole-genome duplication (homoeology) based on the median synonymous substitution rates of genes within a syntenic block. Maize genes scored as orthologous to sorghum genes were assigned to reconstructed ancestral chromosomes according to the arrangement shown in *SI Appendix*, Table 1.

Identification of High-Confidence Retained Homoeolog and Lost Homoeolog Genes. High-confidence genes were considered to be the subset of the maize FGS with annotated start and stop codons whose gene models were supported by expression data (cDNA and/or EST; 27,313 of the 32,540 genes in the maize FGS satisfied these criteria). We further required that it be possible to identify a retained syntenic ortholog in both the rice and sorghum genomes (14,855 of the 27,313 genes) and a recognizable homoeologous location within the maize genome (13,844 of 14,855 genes). Genes with a history of tandem duplication in rice, sorghum, maize1, or maize2 were eliminated from the analysis because these genes are expected to show greater rates of copy number variation, create problems for comparative expression studies, and confuse all arguments involving selection (9,536 of 13,844 genes). Finally, two high-confidence sublists were created. High-confidence retained homoeologous pairs are those pairs in which there are genes that satisfy all the above criteria and are present at both locations in the genome (1,750 genes in both maize1 and maize2). High-confidence no-homoeolog genes are those that satisfy all the above criteria, excluding those genes in which a homoeologous working set or other low-confidence gene is present at the homoeologous location in the genome as well as those genes in which an unannotated syntenic blast hit was detected as the homoeologous location in the genome (3,617 genes located in maize1 and 1,577 genes located in maize2). A total of 842 genes satisfying all these criteria were disqualified from inclusion in either the high-confidence retained homoeolog or high-confidence no-homoeolog list because of a homoeolog made ambiguous by being either a low-confidence gene or unannotated syntenic blast hit.

Calculation of Gene Expression Levels. Gene expression data were calculated from mRNA-Seq data published by four different laboratories (laboratories of X. W. Deng, T. P. Brutnell, P. S. Schnable, and D. Jackson) (21–24) (*SI Appendix*, Table 4). For all expression sets except immature ears, reads were aligned to the maize genome using Bowtie, allowing one mismatch per read and disregarding reads with more than two best alignments (25). Expression values were calculated in units of RPKM with Cufflinks (26), using the published annotations of the B73 refgen_v1 working gene list (3). Immature ear expression data were generated using a digital gene expression technique. For this expression dataset, collapsed reads were aligned to the genome using Bowtie, disregarding all alignments with one or more mismatches and all alignments with more than one unique alignment in the genome. Expression values for each gene were calculated as the sum of the number of reads represented by each collapsed read mapping within a window starting 300 bp upstream of the start of the gene model and extending 300 bp downstream of the gene model. Final gene expression values were calculated in units of reads per million reads.

ACKNOWLEDGMENTS. We are grateful to Damon Lisch for advice during the course of this research and critical reading of the manuscript. This research was supported by National Science Foundation Grants DBI-0701871 (to M.F.) and IOS-0922095 (to N.M.S.).

1. Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res* 16:934–946.
2. Woodhouse MR, et al. (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol* 8: e1000409.
3. Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112–1115.
4. Sankoff D, Zheng C, Zhu Q (2010) The collapse of gene complement following whole genome duplication. *BMC Genomics*, 10.1186/1471-2164-11-313.
5. Wang J, et al. (2006) Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. *Genetics* 172:507–517.
6. Ownbey M (1950) Natural hybridization and amphiploidy in the genus Tragopogon. *Am J Bot* 37:487–499.
7. Buggs RJA, et al. (2010) Characterization of duplicate gene evolution in the recent natural allopolyploid Tragopogon miscellus by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol Ecol* 19(Suppl 1):132–146.
8. Senchina DS, et al. (2003) Rate variation among nuclear genes and the age of polyploidy in Gossypium. *Mol Biol Evol* 20:633–643.
9. Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol* 186:184–193.
10. Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* 16:1679–1691.
11. Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 101:9903–9908.
12. Li L, et al. (2006) Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* 38:124–129.
13. Throude M, et al. (2009) Structure and expression analysis of rice paleo duplications. *Nucleic Acids Res* 37:1248–1259.
14. Paterson AH, et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556.
15. Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature* 428:617–624.
16. Swigonová Z, et al. (2004) Close split of sorghum and maize genome progenitors. *Genome Res* 14:1916–1923.
17. Wei F, et al. (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genet* 3:e123.
18. Salse J, et al. (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell* 20:11–24.
19. Kasahara M, et al. (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719.
20. Swanson-Wagner RA, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res* 20:1689–1699.
21. Wang X, et al. (2009) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* 21:1053–1069.
22. Li P, et al. (2010) The developmental dynamics of the maize leaf transcriptome. *Nat Genet* 42:1060–1067.
23. Jia Y, et al. (2009) Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PLoS Genet* 5:e1000737.
24. Eveland AL, et al. (2010) Digital gene expression signatures for maize development. *Plant Physiol* 154:1024–1039.
25. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10.1186/gb-2009-10-3-r25.
26. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515.
27. Lin JY, Stupar RM, Hans C, Hyten DL, Jackson SA (2010) Structural and functional divergence of a 1-Mb duplicated region in the soybean (Glycine max) genome and comparison to an orthologous region from Phaseolus vulgaris. *Plant Cell* 22:2545–2561.
28. Freeling M (2009) Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol* 60: 433–453.
29. Birchler JA, Veitia RA (2010) The gene balance hypothesis: Implications for gene regulation, quantitative traits and evolution. *New Phytol* 186:54–62.
30. Edger PP, Pires JC (2009) Gene and genome duplications: The impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res* 17:699–717.
31. Sémon M, Wolfe KH (2007) Consequences of genome duplication. *Curr Opin Genet Dev* 17:505–512.
32. Birchler JA, Yao H, Chudalayandi S (2007) Biological consequences of dosage dependent gene regulatory systems. *Biochim Biophys Acta* 1769:422–428.
33. Maere S, et al. (2005) Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* 102:5454–5459.
34. Seoighe C, Gehring C (2004) Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. *Trends Genet* 20:461–464.
35. Gu Z, et al. (2003) Role of duplicate genes in genetic robustness against null mutations. *Nature* 421:63–66.
36. VanderSluis B, et al. Genetic interactions reveal the evolutionary trajectories of duplicate genes. *Mol Syst Biol*, 10.1038/msb.2010.82.
37. Bennetzen JL (2007) Patterns in grass genome evolution. *Curr Opin Plant Biol* 10: 176–181.
38. Baucom RS, et al. (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genet* 5:e1000732.
39. Lee HS, Chen ZJ (2001) Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. *Proc Natl Acad Sci USA* 98:6753–6758.
40. Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* 58:377–406.
41. Bomblies K, Doebley JF (2005) Molecular evolution of FLORICAULA/LEAFY orthologs in the Andropogoneae (Poaceae). *Mol Biol Evol* 22:1082–1094.
42. Lyons E, Pedersen B, Kane J, Freeling M (2008) The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Tropical Plant Biol* 1:181–190.