Two Evolutionarily Distinct Classes of Paleopolyploidy

Olivier Garsmeur,[†,1] James C. Schnable,[†,2] Ana Almeida,[2] Cyril Jourda,[1] Angélique D'Hont,[*,†,1] and Michael Freeling[*,†,2]

[1]Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement (CIRAD), UMR AGAP, Montpellier, France

[2]Department of Plant and Microbial Biology, University of California, Berkeley

[†]These authors contributed equally to this work.

***Corresponding author:** dhont@cirad.fr; freeling@berkeley.edu.

Associate editor: Michael Purugganan

Abstract

Whole genome duplications (WGDs) occurred in the distant evolutionary history of many lineages and are particularly frequent in the flowering plant lineages. Following paleopolyploidization in plants, most duplicated genes are deleted by intrachromosomal recombination, a process referred to as fractionation. In the examples studied so far, genes are disproportionately lost from one of the parental subgenomes (biased fractionation) and the subgenome having lost the lowest number of genes is more expressed (genome dominance). In the present study, we analyzed the pattern of gene deletion and gene expression following the most recent WGD in banana (alpha event) and extended our analyses to seven other sequenced plant genomes: poplar, soybean, medicago, arabidopsis, sorghum, brassica, and maize. We propose a new class of ancient WGD, with *Musa* (alpha), poplar, and soybean as members, where genes are both deleted and expressed to an equal extent (unbiased fractionation and genome equivalence). We suggest that WGDs with genome dominance and biased fractionation (Class I) may result from ancient allotetraploidies, while WGDs without genome dominance or biased fractionation (Class II) may result from ancient autotetraploidies.

Key words: **whole genome duplication, gene fractionation, genome dominance, paleopolyploid, evolution.**

Introduction

Ancient whole genome duplications (WGDs)—or paleopolyploidies—are found throughout the microbial and metazoan tree of life, including protozoa, yeast, fish, frogs, and the 2R event at the base of vertebrates (Semon and Wolfe 2007). However, in all studied nonplant lineages, these events are relatively rare. By contrast, comparative genomic studies in flowering plants have revealed many WGDs in the evolutionary history of different clades (fig. 1 and citations within; up-to-date tree: http://genomevolution.org/wiki/index.php/Plant_paleopolyploidy, last accessed December 11, 2013). For example, after the divergence of poales and zingiberales, two sequential WGDs have been detected in the rice lineage (Paterson et al. 2004; Tang et al. 2010) and three in the recently published genome of *Musa acuminata* (D'Hont et al. 2012). The frequency of WGDs in plants makes it possible to begin to infer some generalized rules about postpolyploidy subgenome evolution to inform the study of WGDs in all eukaryotic lineages.

Comparisons of syntenic regions in *Arabidopsis thaliana* (arabidopsis) resulting from the most recent WGD in that lineage found that duplicate regions (homeologous or homoeologous) did not contain equal numbers of genes (Thomas et al. 2006). This finding of biased gene content was later replicated in a study of the most recent WGD in the maize (*Zea mays*) lineage (Woodhouse et al. 2010) and then generalized to WGDs throughout eukaryotes (Sankoff et al. 2010) with regular enough properties that it could be mathematically modeled (Sankoff et al. 2012).

Differences in gene content between duplicated genomic regions could result from either more gene insertion in the gene-rich region or more gene deletion in the gene-poor region. In maize, comparison to a closely related outgroup species (*Sorghum bicolor*), which lacked the WGD, revealed that the difference in gene content was a result of deletion of duplicate genes to a greater extent from the gene-poor regions (Woodhouse et al. 2010). In maize, it was also possible to show that biased gene loss was consistent across whole reconstructed pairs of ancestral chromosomes and, by extension, was likely a property of whole parental subgenomes (Schnable et al. 2011).

The model proposed to explain the observed bias in gene loss between subgenomes—called biased fractionation—is that one parental subgenome was more expressed in the initial tetraploid and deletions of the more-expressed gene copy were more likely to reduce fitness and be purged from the population, while deletions of the less-expressed copy were more likely to be selectively neutral (Schnable et al. 2011). This bias in gene expression between parental subgenomes has been previously reported in allotetraploid *A. suecica* (Wang et al. 2006) and in allotetraploid cotton (Flagel and Wendel 2010), although the story there has recently become more complex (Yoo et al. 2013). In *A. suecica*, the less-expressed parental genome—*A. thaliana*—experienced more sequence deletion (Chang et al. 2010). In this article, the bias in expression of duplicate genes between parental subgenomes will be referred to as genome dominance.

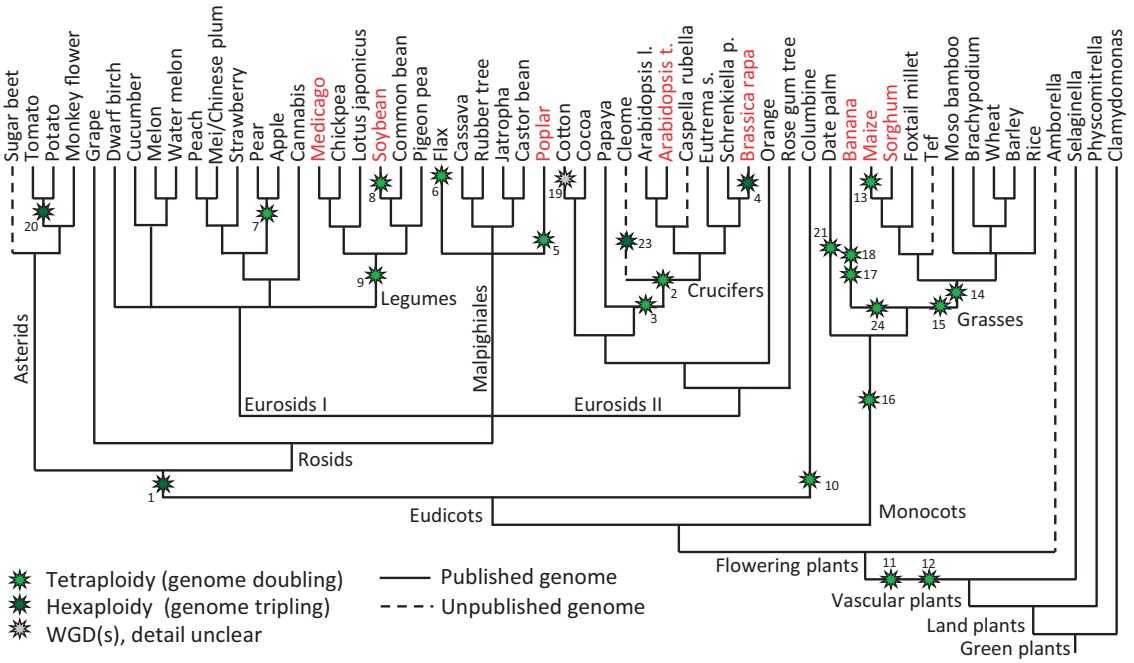

Fig. 1. Phylogenetic tree of plant species with sequenced genomes. Ancient WGDs are represented by colored stars. Full descriptions and references of WGD events are available at http://genomevolution.org/wiki/index.php/Plant_paleopolyploidy (last accessed December 11, 2013). The eight species for which the most recent WGD are analyzed in the present study are indicated in red.

It should be noted that what is almost certainly the same phenomenon has been called "transcriptomic dominance" (Chen 2007) or "unbalanced homoeolog expression bias" (Grover et al. 2012). More recently, studies of a paleohexaploidy in *Brassica rapa*—using arabidopsis as an outgroup—have also revealed biased gene loss across whole subgenomes (Wang et al. 2011) and, like maize, have found that the genome with the most retained genes is enriched in the more expressed copies of duplicated gene pairs (Cheng et al. 2012).

The mechanism of postpolyploid gene loss has been shown to be deletions mediated by intrachromsosomal recombination (Woodhouse et al. 2010; Tang et al. 2012).

In this article, we take advantage of RNA-seq data from *M. acuminata*, for which a genome sequence was recently published (D'Hont et al. 2012), as well as RNA-seq and comparative genomics data from a number of additional plant species to further test the link between bias in gene loss and bias in gene expression following WGD. We now show that the most recent WGD in the *M.* lineage, that is much older than are the most recent polyploidies in either the maize and *Brassica* lineages, is exceptional: The most recent *Musa* WGD (alpha) lacks both genome dominance and differential gene loss. We suspected the former by simply looking at the angle of the homeologous segments in the original *Musa*–*Musa* gene dot plot (supplementary fig. S2, Supplementary Material online; D'Hont et al. 2012). It turns out that *Musa* is not alone in its exceptionality. Overall, the exceptions prove the rule. In the more common Class I WGD, following WGD, the subgenome that is relatively less deleted expresses its genes to relatively higher mRNA level, and in linked fashion. In Class II WGDs of which *Musa*-alpha is the founding member, gene deletion and gene expression are equal between subgenomes.

Results

Separating Alpha and Beta *Musa* WGDs and Reconstruction of Ancestral Blocks

Two WGD (denoted as alpha and beta) have been reported at a similar period, around 65 Ma, in the *Musa* lineage (D'Hont et al. 2012; fig. 1, WGD 17 and 18). We analyzed the paralogous relationships between the resulting sets of four paralogous chromosome segments to identify the two pairs deriving from the most recent WGD event (alpha). SynMap tool from the CoGe platform (http://genomevolution.org/CoGe/SynMap.pl, last accessed December 11, 2013) was used with a quota align ratio (Tang et al. 2011) of 1 to 1. This quota-align value was used to identify the most conserved pairs of duplicated segments, which we attributed to those deriving from the alpha WGD. The resulting pairs of paralogous segments were manually curated and assembled in 101 alpha ancestral blocks covering 207 Mb of the *Musa* genome (64% of the anchored *Musa* genome sequence) and containing 25,678 genes (76% of the anchored genes; table 1*a* and supplementary table S1, Supplementary Material online).

An all-by-all comparison of the alpha blocks was performed to identify the pairs resulting from the beta WGD (see *Musa* beta dot plot in supplementary fig. S2, Supplementary Material online). This comparison allowed the assembly of 13 beta ancestral blocks for which the four paralogous regions were clearly identified (see e.g., fig. 2). These beta blocks covered 125 Mb (38% of the anchored *Musa* genome sequence and 58% of the alpha blocks) and

Table 1. *Musa* Alpha Ancestral Blocks and Patterns of Fractionation.

a. Description of the alpha 101 ancestral blocks	
Length covered by alpha blocks (Mb)	207 (64%)[a]
Number of genes in alpha blocks	25,678 (76%)[b]
Duplicate genes	6,088
Singlet genes	19,590
b. *Musa* fractionation and dominance analysis	
Regions analyzed	101 (207 Mb)
Regions without biased fractionation	73 (149 Mb)
Regions with biased fractionation	28 (58 Mb)
Regions with differential expression	8
Regions with biased fractionation and genome dominance	2

[a]Percentage of the *Musa*-anchored assembly.
[b]Percentage of genes in the anchored assembly.

comprised 12,583 genes (supplementary table S2, Supplementary Material online).

The alpha versus beta assignments of the paralogous chromosome segments were confirmed using both Ks and phylogenetic approaches. Ks values were calculated from 1,141 genes that were retained in three to four copies among the four paralogous regions. Ks from all possible combinations of paralogous gene pairs were compared and the paralogous regions assembled in alpha blocks revealed to have the lowest Ks median values, supporting that they were correctly assembled (supplementary fig. S2, Supplementary Material online).

Phylogenetic trees were constructed with genes that were retained in four copies and 83% of trees confirmed alpha assignments (supplementary fig. S2, Supplementary Material online).

Lack of Fractionation Bias and Genome Dominance Following *Musa* Alpha WGD

Gene content was analyzed within the 101 alpha pairs of paralogous regions previously identified. Among the 25,678 genes included in these regions, 24% (6,088 genes) were retained as duplicates, whereas 76% (19,590 genes) were singletons, owing to losses of duplicated copies after WGD events (table 1*a* and supplementary table S1, Supplementary Material online). For the majority (72%) of these duplicated regions, no biased fractionation was observed; both duplicated segments displayed no significant differences in their total number of genes. These regions represented 149 Mb out of 207 Mb covered by the alpha blocks. The 28 remaining pairs of paralogous segments (58 Mb) displayed a significant ($P < 0.05$) bias in the fractionation pattern, of which one of the segments had lost more genes than the other (table 1*b* and supplementary table S1, Supplementary Material online).

Using two sets of *Musa* RNAseq data (supplementary information S1, Supplementary Material online), we tested for genome dominance by comparing the expression level of all duplicated genes (6,088) included in the 101 alpha blocks (see Materials and Methods). For 2,353 gene pairs (4,706 genes), at least one of the genes was expressed in our RNAseq data. Among the 101 pairs of alpha paralogous regions, eight displayed significant differential gene expression levels and two of them displayed both genome dominance and biased fractionation with an overexpression of the segment carrying the higher number of genes (table 1*b* and supplementary table S1, Supplementary Material online).

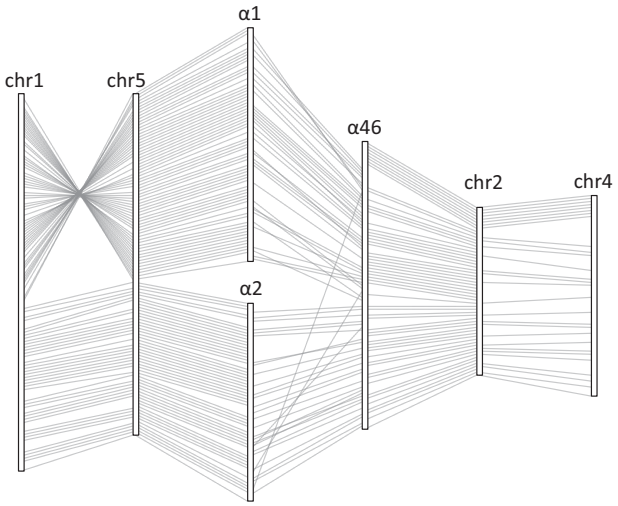

Fig. 2. Example of four paralogous regions identified by all-by-all comparison of *Musa* alpha blocks. Segments from chr1/chr5 were assembled in alpha ancestral blocks α1 and α2, and segments from chr2/chr4 were assembled in alpha ancestral block α46. Horizontal lines link duplicated genes between segments.

Biased Fractionation after the Beta WGD

Gene fractionation that occurred after the beta WGD (but before alpha) was analyzed within the 13 beta blocks identified as described earlier (supplementary fig. S1 and table S2, Supplementary Material online). Among the 12,583 genes included in these beta ancestral blocks, 13.5% (1,710 genes) were retained as duplicates, being present in both paralogous alpha blocks, and 86.5% (10,873 genes) were singlet being present in only one of the paralogous alpha blocks (supplementary table S2, Supplementary Material online, *Musa* beta dot plot in supplementary fig. S2, Supplementary Material online). Among the 13 beta paralogous regions, we found that 11 displayed a bias in the gene-fractionation pattern, one paralogous region containing significantly more genes than the other, representing 85% of the regions compared (supplementary table S2, Supplementary Material online). Interestingly, the overall level of fractionation observed since beta was not very different from the one observed since alpha (86.5% vs. 76%), although much more time was available for fractionation to occur after alpha.

Fractionation Bias and Genome Dominance in Other Species

The level of bias in fractionation of the most recent WGD was analyzed for seven other sequenced species (table 2 and fig. 3 and supplementary fig. S2, Supplementary Material online). Maize and *B. rapa* have been previously demonstrated to exhibit both genome dominance (a tendency toward higher

Table 2. Fractionation Pattern and Genome Dominance in Eight Species.

Species	WGD Class	Substitution Rate (Ks)	Bias Ratio between Duplicate Regions	Fractionation Pattern	Genome Dominance	Expression Data from
Medicago	I	0.87	1.23	Biased	No data	
Sorghum	I	0.95	1.24	Biased (Schnable et al. 2012)	Yes	Dugas et al. (2011)
Arabidopsis	I	0.76	1.17	Biased (Thomas et al. 2006)	Yes	Gan et al. (2011)
Brassica	I	0.34	1.47	Biased (Wang et al. 2011)	Yes (Cheng et al. 2012)	
Maize	I	0.17	1.46	Biased (Woodhouse et al. 2010)	Yes (Schnable et al. 2011)	
Poplar	II	0.23	1.05	Unbiased	No data	
Soybean	II	0.15	1.03	Unbiased	No	Schmidt et al. (2011)
Banana	II	0.39	1.06	Unbiased	No	D'Hont et al. (2012) and supplementary table S4, Supplementary Material online

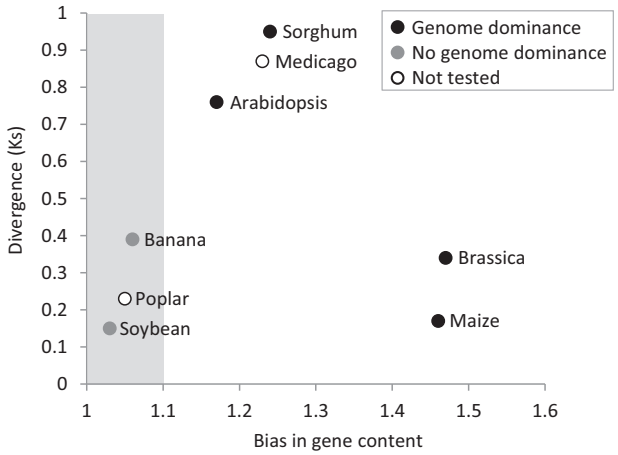

Fig. 3. Distribution of biased gene loss, divergence since duplication, and observable genome dominance among eight ancient alpha WGDs. For example, the sorghum data point is the *Rho* pregrass tetraploidy (fig. 1, WGD number 14). The *x*-axis shows the median bias in gene content and the *y*-axis the modal synonymous substation rate between homeologous genes resulting from the most recent WGD in the named species. The shaded area that includes banana, comprises exceptional WGD (Class II) displaying no biased fractionation and no genome dominance between duplicated regions.

expression of the subgenome with less gene deletions; Schnable et al. 2011; Cheng et al. 2012) in addition to biased fractionation (Woodhouse et al. 2010; Wang et al. 2011). Although high-quality RNA-seq data sets were not available for all species in our data set, it was possible to identify statistically significant genome dominance in sorghum and some genome dominance in arabidopsis (supplementary table S3, Supplementary Material online); both species have been previously reported to exhibit biased fractionation (Thomas et al. 2006; Schnable et al. 2012). In soybean, we found no biased fractionation (table 2 and supplementary fig. S2, Supplementary Material online) and no evidence of biased expression between duplicate regions (supplementary table S3, Supplementary Material online). Poplar showed no evidence of biased fractionation while some bias could be observed in medicago (table 2 and supplementary fig. S2, Supplementary Material online), but no RNA-seq data were available for these two species. The oldest WGD in our data set is the pregrass paleotetraploidy (modal Ks between sorghum subgenomes 0.9, fig. 3) and it is Class I (biased fractionation), so Class I includes older WGDs as well as younger ones.

Discussion

We have shown that ancient WGDs in plants fall into two classes. Class I, represented in the evolutionary histories of the first two plant species to be sequenced (rice and arabidopsis), can be defined as having bias in gene loss between duplicate regions and bias in expression between duplicate regions. Class II shows equal rates of gene loss between duplicate regions and no bias in expression between duplicate regions.

Given the model that bias in gene expression is responsible for bias in gene loss (Schnable and Freeling 2011; Schnable et al. 2011), if a WGD should lack either biased expression or biased gene loss, then both should be absent. Similarly, gene pairs with equally balanced expression are predicted to be more difficult to fractionate without causing fitness reducing changes to gene dosage. In contrast, in gene pairs with biased expression (common in Class I WGD), the less-expressed gene copy can be removed while retaining the majority of total gene pair expression. In an equally balanced gene pair, the loss of either copy results in the loss of 50% of total gene pair expression. Therefore, if some WGDs result in genome dominance and others result in genome equivalence, the two classes of WGD we have observed make perfect sense.

Perhaps the most important features of a scientific model regard its ability to explain a larger set of evidence, while providing more predictive power going forward. Interestingly, the generalized fractionation-dominance model proposed here is in line with such characteristics. As demonstrated in figure 3, our model predicts little or no genome dominance in poplar because of its low bias in gene content, and it also predicts genome dominance in *Medicago* because the subgenomes show a stronger fractionation bias.

What could explain the creation of these two classes of WGD? One tempting and obvious answer is that WGDs with genome dominance and biased fractionation result from

ancient allotetraploidies (arising from hybridization between species) while WGDs without genome dominance or biased fractionation result from ancient autotetraploidies (arising from hybridization within species or from a self of somatically tetraploid perfect flower). One published model to explain bias in gene expression (Freeling et al. 2012) was based on the observation that alleles of genes with a nearby methylated promoter (due to transposon insertions) tend to be expressed at lower levels than alleles of the same genes that lack transposon insertions (Hollister and Gaut 2009; Hollister et al. 2011; Freeling et al. 2012). Even closely related species will tend to exhibit differences in methylated transposon density, and when the genomes of the two species are combined within a single nucleus by allotetraploidy, the gene copy with a methylated transposon closer to its transcription start site will be expressed at lower levels. However, in a true autotetraploid (in this scenario, Class II), transposon content of the original subgenomes would be identical, suggesting that WGDs resulting from autotetraploidy should more likely lack genome dominance. There is no causal evidence either for or against this hypothetical suggestion.

The mode of chromosome pairing during meiosis may also explain why subgenomes of autopolyploids evolve in a different manner than those from allopolyploids. In autopolyploids, by contrast to allopolyploids, the recurrent random assortment of chromosomes may select against deletions of duplicated genes because it may give rise to individuals (and gametes) lacking a complete gene set. Thus, the lack of biased fractionation as well as the equivalence of gene expression within duplicate regions that we observed in Class II WGD could allow maintaining functional genes whatever the chromosome pair involved during meiosis and formation of autotetraploids.

Regardless of the true reason for the lack of genome dominance we observed in some ancient WGDs, these exceptions provided a useful test for the predicted link between genome dominance and biased gene loss (biased fractionation). Genome dominance may serve a beneficial role in some plant lineages by reducing the length of time a lineage suffers from "genomic obesity" following WGD (Bennetzen and Kellogg 1997). Although the test of our hypothesis—allopolypoidy causes genome dominance and biased fractionation—may well involve synthetic polyploids, we caution against interpretations of data taken before several polyploid generations have elapsed, before the epigenetic consequences of being a unique, permanent hybrid have equilibrated.

Materials and Methods

Genome Sequences

Genome-scale analyses were performed using Synmap, in which sequenced genomes of *M. acuminata* DH-pahang, *S. bicolor* BTx623, *Populus trichocarpa*, *Glycine max*, *B. rapa* Chiifu-401, *Medicago truncatula*, *A. thaliana* Col-0, and *Z. mays* are implemented (genome versions are detailed in supplementary information S1, Supplementary Material online).

Construction of *Musa* Ancestral Blocks

Paralogous relationships between the 11 *Musa* chromosomes were analyzed using SynMap tool from the CoGe platform (http://genomevolution.org/CoGe/SynMap.pl, last accessed December 11, 2013; Lyons et al. 2008) with a quota align ratio (Tang et al. 2011) of 1 to 1. Clusters of paralogous chromosome segments containing at least ten genes with a maximal distance of 30 genes between duplicate genes were retained. The SynMap output was checked and manually corrected because some redundancy was observed within the clusters. Some clusters were included or overlapped with some others, certainly due to differences of hits that could be retained when comparing either gene A with B or gene B with A. A total of 101 alpha blocks were assembled by merging the paralogous chromosome segments. Genes were ordered in the alpha blocks based on their orders in the paralogous segments. Singlet genes that were localized between duplicated genes in the paralogous segments were equitably distributed between the duplicated genes in the blocks. These steps were automated using in-house python scripts.

The 13 beta regions were identified by performing an all-by-all comparison of the alpha blocks using BlastP and an e-value cutoff of 1e−05 (fig. 2 and *Musa* beta dot plot in supplementary fig. S2, Supplementary Material online).

Fractionation Bias

For all species, paralogous regions were identified using SynMap (Lyons et al. 2008) with QuotaAlign (Tang et al. 2011) enabled. For the *Musa*-specific analyses, bias of gene fractionation was measured by calculating the ratio of the total number of genes between the paralogous regions.

For interspecies comparisons presented in figure 3, bias was estimated by manually measuring the angle of syntenic diagonals in SynMap graphics set to scale images by gene count and taking the median ratio of longer-of-length-or-width/smaller-of-length-or-width (supplementary fig. S2, Supplementary Material online). A simulation of unbiased fractionation was performed by randomly removing the genes from duplicate regions (equivalent in number and gene content to the *Musa* alpha reconstructed regions) until the total gene loss equaled that observed in *Musa*. Observed bias was calculated as the median bias observed among all simulated segments (genes in more gene rich copy of a segment/genes in less gene rich copy of a segment). After 1,000 iterations, the 95th percentile of bias in gene content was taken as the upper bound of the apparent bias expected to be observed in species with no difference in propensity for gene loss between duplicated segments. A χ^2 test was used to statistically support the measured bias and paralogous regions showing a difference of at least 10% in gene count (bias ratio >1.1) were classified as resulting from biased fractionation.

Patterns of Gene Expression and Genome Dominance

Gene expression quantifications were measured using RNA-seq data and performed as described in supplementary information S1, Supplementary Material online. For *Musa*, two sets

of mRNA-seq data were used (supplementary table S4, Supplementary Material online) and for all other species without previously published descriptions of genome dominance, mRNA reads here downloaded from NCBI's (National Center for Biotechnology Information) Sequence Read Archive. For all species, genome dominance was tested by comparing the expression levels within each pair of duplicated genes along pairs of paralogous regions. Two methods were used: first, a "horserace experiment" was performed, in which we considered that within a duplicate gene pair, a gene was more expressed than its duplicated copy if it had a higher expression level. Second, a "2-fold change method" was used in which a gene was considered more expressed than its duplicated copy when it was expressed at least 2-fold higher. For *Musa*, genome dominance was measured for all alpha blocks by comparing the expression levels of duplicate genes shared between paralogous regions (supplementary table S1, Supplementary Material online). A binomial test (*P* value, 0.05) was used to test whether the differential expression patterns were significant. For other species, the analysis was done using an in-house python script, allowing the identification of duplicated regions containing different total gene counts. For gene pairs present in both regions the total number of cases where the more expressed gene was present in a region with fewer total genes was compared with the total number of cases where the less expressed gene was present in a region with fewer total genes (supplementary table S3, Supplementary Material online). A binomial test was performed to test whether those total numbers of cases were significantly different from what would be expected under zero dominance.

Synonymous Substitution Rates (Ks) and "Dating" of Tetraploidy Events

To identify relative sequence divergence (a proxy for age) among multiple WGDs examined, self–self comparisons of genomes were run using SynMap with default settings (Lyons et al. 2008) and the optional QuotaAlign filter (Tang et al. 2011). A quota of 1:1 was used for tetraploidies (*Me. trunculata*, sorghum, arabidopsis, maize, poplar, soybean, banana) or 2:2 for hexaploidies (*B. rapa*).

Synonymous substitution rates were taken from SynMap, where values are calculated by aligning the protein sequences of a gene pair using the package nwalign (http://pypi.python.org/pypi/nwalign/, last accessed December 11, 2013), translating that alignment back to DNA codons, and calculating Ks using CodeML (Yang 2007).

Supplementary Material

Supplementary information S1, figures 1 and 2, and tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank N. Yahiaoui for providing early access to *Musa* mRNAseq data and for critical reading of the manuscript, J.F. Dufayard for his help in phylogenetic analyses, and G. Martin for his help in the development of bioinformatics scripts. We thank the SouthGreen Bioinformatics Platform (http://southgreen.cirad.fr, http://banana-genome.cirad.fr/, last accessed December 11, 2013) for providing us with computational resources. This work was supported by funding from NSF IOS1248106 to M.F. and from CIRAD (Centre de cooperation Internationale en Recherche Agronomique pour le Developpement).

References

Bennetzen JL, Kellogg EA. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9:1509–1514.

Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV. 2010. Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. *Genome Biol.* 11:R125.

Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol.* 58:377–406.

Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X. 2012. Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 7:e36442.

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, et al. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488:213–217.

Dugas DV, Monaco MK, Olsen A, Klein RR, Kumari S, Ware D, Klein PE. 2011. Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. *BMC Genomics* 12:514.

Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* 186:184–193.

Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC. 2012. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr Opin Plant Biol.* 15:131–139.

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477:419–423.

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196:966–971.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19:1419–1428.

Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A.* 108:2322–2327.

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, et al. 2008. Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol.* 148:1772–1781.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci U S A.* 101:9903–9908.

Sankoff D, Zheng C, Wang B. 2012. A model for biased fractionation after whole genome duplication. *BMC Genomics* 13(Suppl 1), S8.

Sankoff D, Zheng C, Zhu Q. 2010. The collapse of gene complement following whole genome duplication. *BMC Genomics* 11:313.

Schmidt MA, Barbazuk WB, Sandford M, May G, Song Z, Zhou W, Nikolau BJ, Herman EM. 2011. Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. *Plant Physiol.* 156:330–345.

Schnable JC, Freeling M. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS One* 6:e17855.
Schnable JC, Freeling M, Lyons E. 2012. Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol Evol.* 4:265–277.
Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci U S A.* 108:4069–4074.
Semon M, Wolfe KH. 2007. Rearrangement rate following the whole-genome duplication in teleosts. *Mol Biol Evol.* 24:860–867.
Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci U S A.* 107:472–477.
Tang H, Lyons E, Pedersen B, Schnable JC, Paterson AH, Freeling M. 2011. Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinformatics* 12:102.
Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC. 2012. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190:1563–1574.
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.
Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, et al. 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172: 507–517.
Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, et al. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet.* 43:1035–1039.
Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol.* 8:e1000409.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
Yoo M-J, Szadkowski EP, Wendel JF. 2013. Homeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110: 171–180.