The evolution of nuclear genome structure in seed plants[1]

Elizabeth A. Kellogg[2,4] and Jeffrey L. Bennetzen[3]

[2]Department of Biology, University of Missouri–St. Louis, One University Boulevard, St. Louis, Missouri 63121 USA; and
[3]Department of Genetics, University of Georgia, Athens, Georgia 30622 USA

Plant nuclear genomes exhibit extensive structural variation in size, chromosome number, number and arrangement of genes, and number of genome copies per nucleus. This variation is the outcome of a set of highly active processes, including gene duplication and deletion, chromosomal duplication followed by gene loss, amplification of retrotransposons separating genes, and genome rearrangement, the latter often following hybridization and/or polyploidy. While these changes occur continuously, it is not surprising that some of them should be fixed evolutionarily and come to mark major clades. Large-scale duplications pre-date the radiation of Brassicaceae and Poaceae and correlate with the origin of many smaller clades as well. Nuclear genomes are largely colinear among closely related species, but more rearrangements are observed with increasing phylogenetic distance; however, the correlation between amount of rearrangement and time since divergence is not perfect. By changing patterns of gene expression and triggering genome rearrangements, novel combinations of genomes (hybrids) may be a driving force in evolution.

Key words: duplication; polyploidy; retrotransposon.

Plant nuclear genomes are enormously variable. Chromosome number, the degree of gene clustering, and chromosome size can all differ by as much as an order of magnitude, even between closely related species. Some variation is generated so rapidly that two different allelic versions of a chromosomal segment (otherwise known as haplotypes) can be dissimilar in gene content and arrangement even within a single plant species like maize (Fu and Dooner, 2002). Hence, plant nuclear genomes vary sufficiently to serve as powerful differentiating factors. Some changes clearly mark particular lineages of seed plants, such as the large inversions and translocations that are found within some clades of grasses (Gale and Devos, 1998). Other changes, such as polyploidy and most gene duplications/deletions, are so frequent that they occur independently in multiple lineages. Recent studies have begun to characterize the natures, rates, and mechanisms of these various types of chromosomal rearrangement, thereby providing our first detailed insights into how these changes contribute to current evolved states and how they may be used in phylogenetic analysis.

This article will describe the standard structural patterns in the nuclear genomes of seed plants and show how these have been conserved over the last 100 million years of angiosperm evolution. The natures, mechanisms, and frequencies of specific chromosomal rearrangements will be described. Genome size variation and genome duplication will be discussed in some detail. Through this presentation, we hope to provide a comprehensive view of the current understanding of plant nuclear genome structure and evolution and indicate future directions of this field of study. Much of this review will focus on grasses (Poaceae), primarily because so much comparative genome structure information is available within this family. Comparative sequence analysis has been undertaken in or-

thologous regions of the barley, maize, rice, sorghum, and wheat genomes (reviewed in Bennetzen and Ramakrishna, 2002), thereby providing a large data set for the characterization of local genome evolution. Comparative recombinational maps have also been generated for these species, as well as for pearl millet, sugarcane, foxtail millet, rye, and a few others (Gale and Devos, 1998). Hence, sufficient data are available only in the grasses for the comprehensive characterization of plant nuclear genome structure and evolution. Comparative data are rapidly accumulating for other families, however, notably for Brassicaceae, and we expect that the next several years will see a tremendous increase in information on how genomes evolve.

Many other important aspects of nuclear genome evolution will not be covered here, although fascinating new data are accumulating. For example, the overall base composition of genomes varies and affects codon usage patterns. Codon usage patterns and percentage GC in onion are more similar to *Arabidopsis* than to rice, suggesting (although hardly proving) that the high GC content of the grasses may be derived (Kuhl et al., 2004). Intron size and number vary among plants, but the correlates of this variation are not well known. In animals, small genome size correlates with smaller introns, but this relationship does not seem to hold in plants (see for example, Dubcovsky et al., 2001; Wendel et al., 2002). Base composition, rates of substitution, and size and number of introns are all important aspects of genome evolution, and we hope that more data and analyses will be forthcoming on all topics in the next few years.

STRUCTURE OF SEED PLANT GENOMES

Gross genome structure—All angiosperms contain relatively complex nuclear genomes with genes scattered across multiple chromosomes. Other plants are less well characterized, but their large genome sizes (see Bennett and Leitch, 2003, for a comprehensive presentation of plant genome sizes) indicate that most nonflowering plant genomes are equally complex. In even the smallest genomes, like *Arabidopsis* (about 140 Mb), more than 20% of the DNA is composed of various repetitive elements (Arabidopsis Genome Initiative, 2000).

These repeats include transposable elements and various types of simple tandem repeats, including satellite DNA and simple sequence repeats (SSRs). In most angiosperms, transposable elements, especially long terminal repeat (LTR) retrotransposons, comprise the vast majority of this repetitive DNA (reviewed in Bennetzen, 2002a).

Chromosome numbers are highly variable in the flowering plants and do not generally relate to overall genome size. For instance, the ~440-Mb haploid rice genome is distributed across 12 chromosomes, whereas the ~4900-Mb haploid barley genome is present on only seven chromosomes. Hence, the average barley chromosome is almost twice as large as the entire rice genome.

Many plant nuclear genomes also contain accessory chromosomes, often known as B chromosomes. These highly condensed chromosomes are usually small and largely or completely devoid of functional genes. In maize, pollen grains that carry B chromosomes exhibit an advantage in fertilization, thereby increasing the chance that these essentially selfish chromosomes will persist in subsequent generations (reviewed in Carlson, 1986). Within a single species, individuals may have anywhere from zero to several of these accessory chromosomes, thereby altering nuclear chromosome number and genome size but having little other effect upon the biology of the organism other than increasing overall rates of nondisjunction.

Despite these dramatic differences in size and number, all seed plants appear to have fairly similar general organizations of their chromosomes. Most angiosperm chromosomes have centromeric regions that are necessary for efficient chromosome segregation. These regions show extensive chromatin condensation and are flanked by large regions of additional heterochromatin that is enriched for tandem repeats and transposable elements. In large genomes like barley and wheat, these heterochromatic pericentromeric regions can make up more than 50% of the physical length of the chromosome. Other heterochromatic regions (often called knobs) are also found in all seed plant genomes, even small ones like *Arabidopsis*, and they are similarly enriched for tandem repeats and transposable elements (Ananiev et al., 1998; Cold Spring Harbor et al., 2000). The ends of all seed plant chromosomes studied to date contain short tandem repeats that are presumably added by the enzyme telomerase (Richards and Ausubel, 1988). In fact, studies by McClintock (1932) in maize first demonstrated the existence of special structures/properties (now called telomeres) at the ends of eukaryotic linear chromosomes that were necessary for their protection from progressive degradation and/or end fusion.

Although centromeric heterochromatin and knobs are shared by all seed plant genomes, their sizes and locations show extreme interspecies and sometimes intraspecies variation. Smaller genomes always have fewer and smaller regions of heterochromatin than larger genomes, both in plants and in animals. The euchromatic component of plant chromosomes can also be highly variable in size and distribution. Euchromatin, less condensed than heterochromatin, appears to contain most of the genes and most of the meiotic recombination in all multicellular eukaryotes, including plants. However, all euchromatin is not likely to be qualitatively or quantitatively consistent between different chromosomal regions, even within a species. For instance, some euchromatic regions in the wheat genome appear to have a significantly higher gene density and similarly higher recombination frequency than other euchromatic regions (Gill et al., 1996a, b).

Plant nuclear genomes exist in an organelle that has significant three-dimensional structure. We know almost nothing about the nature or importance of the three dimensional folding of chromatin in the interphase plant nucleus. Bennett (1987, 1988) and Heslop-Harrison and Bennett (1990) argued that the arrangement of chromosomes in the nucleus is highly structured. They show that in hybrids the two parental chromosomes occupy different regions of the nucleus and that these patterns are consistent for any given pair of parents. However, this work has not been extended to other species beyond the few grasses they investigated. Research in animals indicates that specific and highly dynamic chromatin arrangements are formed and that these arrangements may be unique to a tissue type, developmental time, or stage in the cell cycle (reviewed in Belmont, 2003). These patterns of three-dimensional structure presumably evolved as important components of regulated gene expression, nuclear packaging, and/or chromosomal mechanics. In plants, the conservation of one predicted genome structure component (matrix attachment regions, or MARS) at specific locations in orthologous genes has been observed (Avramova et al., 1998; Tikhonov et al., 2000). These results indicate that the regulated folding of interphase chromatin is an important component of eukaryotic genomes deserving additional investigation. In summary, plant nuclear genomes are conserved in overall structure, all containing multiple chromosomes with centromeres, telomeres, heterochromatic blocks, and euchromatic regions. The locations and sizes of the centromeres and various chromatin types are exceedingly variable between species and can vary within species. Chromosome number, chromosome size, and overall genome size are highly variable between species, sometimes even within the same genus, but tend to be conserved within a species.

Local genome structure—Over the last eight years, genome sequence analysis in plants has shifted from studies of single genes in isolation to detailed studies of larger chromosomal regions, including whole genomes. These studies have shown that plant genes are relatively compact and often clustered, even in large genomes. Plant introns are usually small, averaging less than 200 bp, so that the average transcribed portion of a gene is less than 2.5 kb (Arabidopsis Genome Initiative, 2000). Upstream and downstream regulatory elements are usually small as well, amounting to no more than a few hundred additional bases in most genes. There are exceptions, however, in which a regulatory element can be more than 90 kb upstream of a locus (Stam et al., 2002). However, regulation at a distance (common in many animal genes) appears to be rare in plants, so that the average gene plus its regulatory components will normally occupy only about 1–5 kb of genomic space.

Genes can be tightly juxtaposed in plants and are often within a few hundred base pairs of each other in plants with small genomes like *Arabidopsis* and rice (Arabidopsis Genome Initiative, 2000). Even in plants with large genomes, like maize (~2400 Mb) and barley, gene clusters can be found (Rahman et al., 1997; Llaca and Messing, 1998; Feuillet and Keller, 1999; Fu et al., 2001). Within these gene clusters, gene density approaches one gene per 5 kb, close to the average value of one gene per 4.5 kb for the sequenced portion of the *Arabidopsis* genome (Arabidopsis Genome Initiative, 2000). How-

ever, clusters with numerous adjacent genes appear to be relatively rare in large genome plants. More commonly, ongoing sequencing projects have found mostly clusters of one to three genes in islands surrounded by seas of repetitive DNA. These intergenic repetitive DNA blocks are composed primarily of intact and fragmented LTR-retrotransposons, often arranged in a nested structure of elements inserted within elements (SanMiguel et al., 1996; reviewed in Bennetzen, 2000). In maize, as in all other plants investigated, repetitive DNA blocks are heavily methylated at the cytosines in 5′-CG-3′ and 5′-CNG-3′ sequences (Bennetzen et al., 1994; Kashkush et al., 2002, 2003). Most of these hypermethylated repeat blocks extend for 5–150 kb in maize, making them a significant impediment to physical mapping, genome sequencing, and map-based gene isolation (Bennetzen et al., 1994).

In summary, local genome structure in seed plants exhibits significant general similarity across all studied species. In nuclear genomes with a low repeat content, most genes are found near one another. In larger genomes, repeats are often inserted between genes, although small repeats like miniature inverted repeat transposable elements (MITEs) can commonly be found in the introns, promoters, or 3′ trailer sequences of genes (Wessler et al., 1995). Gene islands in large genome plants often resemble the genic regions of small genome plants, although the large repeat blocks in large genome plants usually separate islands with only a few (one to three) genes.

GENOMIC COLINEARITY IN PLANTS

The use of DNA markers as probes for recombinational mapping allowed the first comprehensive comparisons of gene composition and order between species. The DNA probes used for most early mapping studies were restriction fragment length polymorphism (RFLP) markers, and their required low copy number meant that they were often fragments of genes. The first comparative mapping study in plants, a comparison of tomato and potato by the Tanksley laboratory, indicated excellent conservation of gene content and order across these genomes (Bonierbale et al., 1988). A more distant relative of tomato and potato, pepper, was later investigated by this same group. This study indicated numerous chromosomal rearrangements against a background of conserved gene content (Tanksley et al., 1988). The first comparative genetic mapping project in grasses, a brief study to test the utility of maize DNA probes to map sorghum and other cereal species (Hulbert et al., 1990), was followed by a flood of studies and syntheses showing that grasses could be studied as variants on a single experimental genome (Bennetzen and Freeling, 1993). One important synthesis demonstrated that all cereal species could be represented by a small number of gene linkage blocks, commonly shown in a comparative circle map (Moore et al., 1995; Gale and Devos, 1998). These maps indicated that only a few major chromosomal rearrangements differentiated the nuclear genomes of such distantly related grasses as rice and barley or maize and wheat (but see Gaut, 2002).

Fewer comparative mapping studies have been pursued in eudicots, and these have led to the popular perception that eudicots have more genomic rearrangements than do grasses. Many of the eudicot comparative mapping projects have focused on comparisons to *Arabidopsis*, an organism that has a history of extensive genomic instability (Blanc et al., 2000; Ku et al., 2000); this may give the impression that all eudicot genomes are highly rearranged. Among grasses, some lineages

have also been unstable and would have served as poor reference genomes for repeated comparative studies. For instance, the close wheat relative *Aegilops umbellulata* has more gross chromosomal rearrangements relative to wheat than does barley, a much more distant relative (Zhang et al., 1998). More comparative analyses are needed among much more diverse plant genomes. It is likely that, as in the original eudicot studies (Bonierbale et al., 1988; Tanksley et al., 1988) and in most cereal comparisons (Gale and Devos, 1998), closely related species will usually show the fewest rearrangements. It will be important to determine which lineages show the greatest degrees of conservation in gene arrangement and whether grasses are in any way exceptional in their degree of genome conservation. Gaut (2001) developed a statistical method for assessing colinearity of genomes. In this, he evaluated whether runs of colinear genes could have occurred at random; the method also allowed him to incorporate an estimate of map error into the statistical test. Based on his reanalysis of the maize genome, he estimated 1.3–1.9 rearrangements per million years, a rate comparable to estimates of 1.4–2.8 rearrangements per million years for cotton (Brubaker et al., 1999) and rather higher than rates estimated for the same species by Lagercrantz (1998). Gaut (2002) extended this approach and noted that the rate of rearrangement, measured as the probability of synteny between two markers, is roughly constant among grasses. Similar rates are reported for several Brassicaceae and Solanaceae, although the *Brassica rapa/oleracea* comparison is estimated as 2.5 per million years. This may be an artifact, however. A (relatively) small number of markers is available for cross-species comparisons, and the taxa investigated diverged millions or tens of millions of years ago. Burke et al. (2004) found a much higher rate of rearrangement among closely related *Helianthus* species that diverged less than a million years ago (5.5–7.5 per million years); their genome maps have a high number of comparable markers and thus detection of rearrangements may be easier. The rate of rearrangement in grasses and other plants may simply reflect the number that can be detected over long periods of time.

Comparisons between cereal genomes and the *Arabidopsis* genome were stimulated by the development of a detailed physical map and contiguous DNA sequence for *Arabidopsis*. An early comparative mapping study indicated that significant genetic colinearity had been conserved between *Arabidopsis* and rice over the more than 100 million years since the lineages that gave rise to these two species diverged (Paterson et al., 1996). However, subsequent studies demonstrated that adjacent genes in rice were often not adjacent in *Arabidopsis* (Bennetzen et al., 1998) or linked on the genetic map (Devos et al., 1999). With the completion of the sequences of highly extended regions of the *Arabidopsis* and rice genomes, more comprehensive studies have become feasible, demonstrating "scant collinearity in gene order" (Liu et al., 2001).

Comparative sequence analyses of small chromosomal segments, mostly in grasses, have shown that local gene content, order, and orientation are also conserved in close relatives (reviewed in Bennetzen and Ma, 2003). Even in distant comparisons, like *Arabidopsis* and rice, small segments of colinear genes are apparent (van Dodeweerd et al., 1999; Salse et al., 2002). The sequence conservation observed in comparisons between species that have evolved independently for 10 million years or more appears to be limited to genes. Some conserved noncoding sequences (CNSs) are found in grass genomes, but these short segments of similarity and possible ho-

mology are tightly linked to genes and may have gene regulatory roles (Kaplinsky et al., 2002; Guo and Moose; 2003; Hong et al., 2003). Moreover, the CNSs in plants are nowhere near as abundant or as large as the CNSs discovered by comparative sequence analysis in vertebrates (Thomas et al., 2003).

In summary, closely related plants have similar gene content and stretches of conserved gene order. This colinearity, often called synteny (which actually does not refer to order at all, but only to presence on orthologous chromosomes), diminishes as more distant relatives are compared. Some lineages, like *Arabidopsis*, appear to have exceptionally high frequencies of genomic rearrangement that greatly diminish their colinearity even with close relatives.

CHROMOSOMAL REARRANGEMENT

Cytogenetic studies long ago indicated that major chromosomal rearrangements can mark specific lineages of plants and other eukaryotes. Comparative recombinational maps using DNA markers were able to identify inversions, translocations, chromosome fission/fusion events, and chromosome duplications (both aneuploid and polyploid) that are specific to individual plant families, tribes, or species (Gale and Devos, 1998). Mechanisms of chromosome breakage and repair that might lead to such major changes have been documented in laboratory studies, but we do not know the precise cause of genome rearrangement in any naturally occurring population of plants. Major chromosomal rearrangements can have dramatic biological effects, particularly in the fertility of inversion and translocation heterozygotes, but they do appear to be relatively rare in most plant lineages. For instance, only four major chromosomal rearrangements differentiate the rice and foxtail millet lineages, despite more than 50 million years of independent evolutionary descent (Devos et al., 1998). On the other hand, Rieseberg et al. (1995, 1996, 2003) have shown that genome rearrangements can occur immediately upon hybridization, even at the diploid level. Furthermore, these rearrangements are reproducible in artificial hybrids. They have interpreted this to mean that selection favors particular combinations of genes and genomic arrangements.

In contrast to the rarity of large chromosomal rearrangements, small chromosomal rearrangements are incredibly abundant in plants. In *Arabidopsis*, some analyses have estimated that over 60% of the genes appear to have been rearranged (mostly deleted) since the last polyploidization occurred in this lineage (Vision et al., 2000). In comparisons of rice and sorghum, two highly diploidized species that diverged from a common ancestor 50–70 mya (Grass Phylogeny Working Group, 2001), it appears that approximately 20% of the genes have been rearranged (Bennetzen and Ma, 2003). Hence, small chromosome rearrangements involving genes are orders of magnitude more frequent in plants than are large chromosomal rearrangements. In approximately the same amount of geological time, rearrangement of genes within mammals is less than 1% (Dehal et al., 2001; Cooper et al., 2003; Thomas et al., 2003).

The population genetics of chromosomal rearrangement has been the subject of much theory, but relatively little data (reviewed in Rieseberg, 2001). In general, a chromosomal rearrangement appears to protect a part of the genome from gene flow, fixing a set of alleles simultaneously and preventing their breakup or loss by crossing back to progenitor plants. Individ-
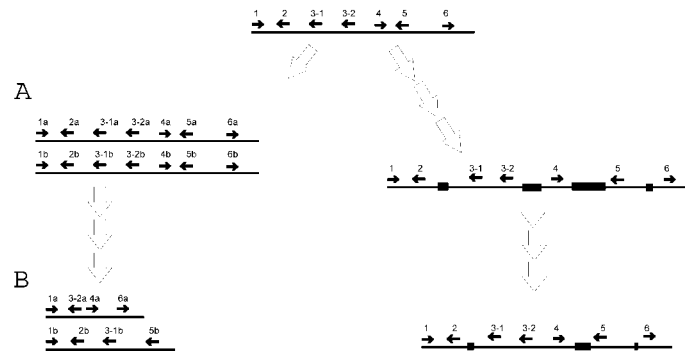

Fig. 1. Mechanisms responsible for genome size variation in plants. The uppermost line indicates an idealized segment of plant nuclear DNA, with horizontal arrows indicating the direction of transcription, location, and size of genes. The genes are numbered sequentially left to right, except for a pair of tandemly duplicated genes that are designated 3–1 and 3–2. A. Genome expansion. The left figure shows the outcome of polyploidy, in which the nucleus now contains twice as many copies of similar genetic regions. The right figure shows multistep growth in genome size by insertion of transposons between genes. B. Genome shrinkage. The left figure shows a progressive (probably random) deletion of genomic DNA, including some genes. These gene losses appear to be tolerated because the organism still retains at least one copy of each gene (Ilic et al., 2003). The right figure shows progressive deletion of nongenic sequences by an accumulation of small deletion events (Bennetzen et al., in press). Of course, in any genome, both the leftward and rightward processes could be ongoing simultaneously, and there will be a dynamic competition between concurrent expansion and contraction processes. Black boxes indicate repetitive DNA inserted between genes, commonly LTR-retrotransposons.

ual chromosomal rearrangements may create only small reductions in plant fitness, at the same time as they contribute to reproductive isolation. Small population sizes and/or metapopulation structure may also permit fixation of mildly deleterious rearrangements.

Investigations of the sequences between genes show that much of this material is derived from transposable elements, even in small genome species like *Arabidopsis* and rice (Devos et al., 2002; Ma et al., 2004). Transposable elements are remarkably active in plants. Bursts of element activity can generate hundreds or thousands of heritable new element copies in a single plant generation. On average, most plant nuclear genomes accumulate several thousand new transposable elements per million years, most of them LTR-retrotransposons and MITEs. Recently, it has been shown that these elements and other nonessential DNA can be relatively quickly removed from plant genomes by illegitimate recombination (Devos et al., 2002; Ma et al., 2004). Deletions caused by illegitimate recombination are usually tiny, the vast majority being less than 100 bp (Bennetzen et al., in press), but their continuous accumulation can rapidly remove a large part of the genome. Ma et al. (2004) conservatively estimated that LTR-retrotransposon sequences had a half life of less than 6 million years in rice and that various deletion processes had removed more than 190 Mb of LTR-retrotransposon sequence from the rice genome in the last 8 million years. Given this rapid dynamic of insertion and deletion, it is not surprising that the unselected sequences between genes are different in species that have diverged for more than a few million years. Figure 1 provides an idealized description of global and local processes that can enlarge or shrink a plant genome.

Gene movement is significantly less rapid than intergenic sequence change, but it is still relatively frequent in plants.
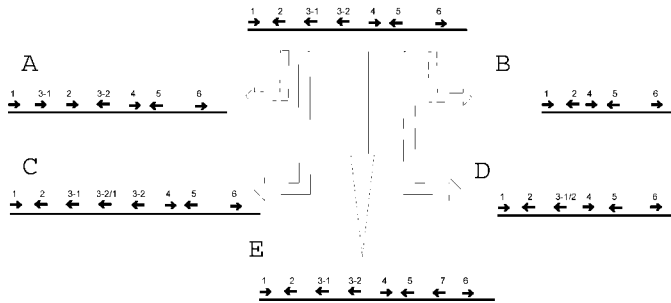
Fig. 2. Types of gene movement observed in plants. The topmost line shows the same idealized segment of plant nuclear DNA as in the same position of Fig. 1. A. Inversion of a two gene segment (genes 2 and 3–1). B. Deletion of a multigene segment (genes 3–1 and 3–2). C. Growth in tandem gene family number by unequal homologous recombination. D. Decrease in tandem gene family number by unequal homologous recombination, creating a single gene from parts of genes 3–1 and 3–2. This is the reciprocal outcome of the event described in panel C. E. Movement of a new gene (gene 7) into the region from an unlinked chromosomal location.

All types of rearrangements are observed, including small inversions, deletions, duplications, and long-distance movements to new chromosomal locations. The high rate of this last class of rearrangement is especially unexpected. In a comparison of orthologous *adh* regions of rice, sorghum, and maize, for instance, four genes were inserted into the *adh* region in the ancestral lineage that gave rise to sorghum. One of these was a two-gene insertion that occurred prior to the divergence of maize and sorghum ancestors. The other two were insertions of a single gene from unlinked chromosomal locations into adjacent positions after the divergence of maize and sorghum lineages (Ilic et al., 2003). A mechanism for the movement of these small gene-bearing DNA fragments has not been documented, but unequal homologous recombination or double-strand break repair (reviewed in Gorbunova and Levy, 1999) are likely candidates. We also do not know the origins of specific events that caused small inversions, gene deletions, or gene duplications in plants, but unequal recombination is a candidate. For a single-gene inversion found near the barley *Vrn1* homologue, Ramakrishna and coworkers observed two highly degenerate flanking transposable elements of the same family in inverted orientation (Ramakrishna et al., 2002). Intrastrand unequal recombination between these elements would have led to just such an inversion. Similarly, unequal recombination between flanking direct repeats or directly repeated gene family members can give rise to the frequent gene duplications and deletions found for all genes, but especially in tandem gene families. Figure 2 provides a summary of the events that can alter gene order and content in specific small regions of a plant genome.

Now that we know that local genome rearrangement is an aggressive and ongoing process in plants, it becomes clear that we need to know much more about it. What are the precise mechanisms that dominate genome rearrangement in any given genome context? What percentage of initial events are fixed? Are there periodicities in the generation of rearrangements, and might these be caused by particular environmental inputs? For all of these questions, are there lineage-specific differences in the answers? Last, and perhaps most important, what effects do these local rearrangements have on gene and genome function? The movement of *adh1* orthologues to different chromosomes and repetitive DNA contexts in rice vs. maize does not seem to have altered the tissue specificity, developmental timing, or inducible properties of this gene (Ilic et al., 2003). This same lack of sensitivity to chromosomal environment has been observed for the majority of other plant genes, indicating that chromosomal rearrangement has surprisingly little influence upon gene expression. Much of this resistance to position effect in plants is probably epigenetic, as shown by the sudden transcriptional alterations that can occur after epigenetic balance is disturbed (Kashkush et al., 2003). Moreover, so many rearrangements occur that it seems likely that even a low frequency of biological significance to individual events will eventually be swamped by rare events that affect gene function. A prime example of this is the insertion of a transposable element into a promoter region, thereby providing the raw material for the evolution of new regulatory properties (reviewed in Wessler et al., 1995).

In summary, local genomic rearrangement is a continuous and highly active process in seed plant genomes. Mechanisms for these events are known or suspected. However, we need much more information to determine the relative frequencies of different classes of rearrangement, whether they differ between plant lineages, and the biological outcomes of these rearrangements.

GENOME SIZE VARIATION

Flowering plants have an impressive range of nuclear genome sizes, including some species with less than 50 Mb of DNA per haploid nucleus and others with more than 85 000 Mb (Bennett and Leitch, 2003). Some of this variation originates in the frequent formation of polyploids in flowering plant lineages (Wendel, 2000), but most appears to be caused by variations in the amount of LTR-retrotransposon DNA in specific genomes (reviewed in Bennetzen, 2002a). In all plant species examined, the vast majority of intact LTR-retrotransposons are relatively young (less than 5 million years since insertion; Bennetzen et al., in press). The youth of these elements is misleading, however, because their ubiquity and the great divergence of active elements indicate that their origin only slightly postdates the origin of eukaryotes. Although recent bursts of LTR-retrotransposon activity are both likely and documented (Vitte and Panaud, 2003), the simplest explanation for the absence of intact ancient elements is that they have been progressively deleted (Devos et al., 2002; Wicker et al., 2003; Ma et al., 2004). These mechanisms can account for the current observed differences in plant genomes without any recourse to a possible selective advantage of particular plant genome sizes, although this conclusion does not negate the possibility that selection on nuclear DNA content can be a major factor in many instances (Bennetzen et al., in press).

Future studies will need to determine whether differences in genome size in any particular lineage are caused by an unusually low or high rate of transposable element amplification or by differences in the mechanisms of transposable element removal. Petrov and coworkers demonstrated that the frequency and size of DNA deletions in LINE retroelements was greater in the small genome of the insect *Drosophila melanogaster* than in large genomes of the insect genus *Laupala* (Petrov et al., 2000). Similarly, Kirik and colleagues (2000) have shown that double-strand breaks are more commonly resolved with insertions and less commonly with deletions in tobacco, a plant with a relatively large genome, than in *Arabidopsis*. Hence, it is likely that different organismal lineages

will exhibit different rates and modes of DNA removal, thus providing one contribution to differences in current genome sizes.

Lynch and Conery (2003), working with a large sample of prokaryotes and eukaryotes, showed clearly that genome size is negatively correlated with the parameter $N_e\mu$, which is the product of the effective population size and the mutation rate per nucleotide. Because the range of mutation rates is rather narrow, variation in this parameter generally reflects variation in population size. Small population sizes correlate with large genomes. Vinogradov (2003) found a similar result in angiosperms, by comparing genome size in rare angiosperms with that of more common species. On average, rare plants had larger genomes. There was no correlation, however, with life history, a result also found for species of *Hordeum* (Poaceae; Jakob et al., 2004). In some clades of *Hordeum*, annual species had smaller genomes than perennials, whereas the relationship was reversed in other clades.

Soltis et al. (2003) have examined variation in genome size among the angiosperms and have found that, as expected, genome size varies appreciably when mapped on the phylogenetic tree. They estimated the ancestral size as "very small" (<1.4 pg per 1C nucleus) and noted that very large genomes appear in only a few clades (Santalales, Asparagales, and Liliales). They also observed increases and decreases of genome size over evolutionary time. This inference of fluctuation is expected in part because they used both parsimony and squared-change parsimony to estimate ancestral states. Both methods assume that increasing and decreasing values are equally likely, the latter giving the equivalent of a Bayesian reconstruction that assumes a Brownian motion model of evolution of the character (Maddison, 1991) and that Brownian motion is "infinitely jiggly" (Felsenstein, 2004, p. 392). Other models of character change (e.g., those listed by Felsenstein, 1988) might give a different result. For example, Bennetzen and Kellogg (1997) reconstructed ancestral genome sizes under both the Brownian motion model and a model in which genomes could only get bigger. Not surprisingly, the reconstructions were sensitive to the underlying model of character state change.

Figure 3 illustrates the evolution of genome size among diploid grasses. Sampling is more comprehensive than in our previous study (Bennetzen and Kellogg, 1997) but remains heavily biased toward subfamily Pooideae (all taxa derived from the common ancestor of *Nardus* and *Aegilops*). We retrieved 1C values for all diploid grasses and outgroups from Cyperaceae and Juncaceae from the Angiosperm C-values Database (Bennett and Leitch, 2003). *Lygeum spartum* ($2n = 40$) and *Deschampsia antarctica* ($2n = 26$) are listed as diploids but almost certainly are polyploid based on their chromosome numbers; they were therefore excluded. For monophyletic genera with more than one species, we calculated the average 1C value. If, however, more than one chromosome base number was present, we calculated the average value for each base number separately. This assumes that each chromosomal group within a genus is monophyletic, which is in fact unlikely. However, phylogenetic trees are not available for many of the individual genera in the tree. Averaging chromosome numbers within a genus therefore seemed a reasonable compromise between illustrating variation vs. losing information. Because *Festuca* is paraphyletic and contains *Lolium*, we used an average value for *Lolium*, but then treated each clade of *Festuca*

separately following the molecular cladogram presented by Catalán et al. (2004).

We optimized 1C value for diploid genomes using squared change parsimony in MacClade 4.0 (Maddison and Maddison, 2000). Under this optimization, we estimated an ancestral 1C genome size about 1.3 pg of DNA, slightly smaller than our previous estimate (ca. 1.9) and just below the boundary between "very small" and "small" as defined by Soltis et al. (2003). Genome size is inferred to increase at the base of Pooideae and also at the base of core Pooideae (after the divergence of *Nardus*). This latter increase correlates with the origin of a chromosome base number of $x = 7$, which requires the combination of several chromosomes and loss of multiple centromeres. Consistent with this genomic reorganization, comparative maps find novel genome arrangements (see Kellogg, 1998). The largest genomes in grasses are found in Triticeae, notably in *Psathyrostachys* (average 8.38) and *Secale* (average 7.8).

Under this simple model of evolution, however, genome size apparently decreases in multiple lineages. The genus *Phleum* seems to have an unusually small genome, as do *Corynephorus*, *Holcus*, the two annual species of *Poa* (*infirma* and *supina*), and the $x = 7$ species of *Phalaris*. *Zingeria biebersteiniana*, with $x = 2$, also has a very small genome, but it has not been placed phylogenetically so is not included in this tree. Morphological similarity would place it near *Avena*.

A more sophisticated model of evolutionary change might lead to different conclusions about the exact size of the ancestral genome and the relative frequency of genome expansion or contraction. Such optimization is also sensitive to taxon sampling; because there are no genome size estimates for any grass lineages that diverged before the common ancestor of maize and rice, nor for any of the immediate outgroups, we expect that the details of our estimates here are subject to change. However, the current data and optimization are sufficient to demonstrate that genome size is labile.

It is also interesting to note that base chromosome number does not correlate precisely with size. A base number of 7 is ancestral and synapomorphic for core Pooideae, but there have been reductions to $x = 4$ in *Milium vernale*, 6 in some species of *Phalaris*, and 5 in *Briza minor*. Whereas *Briza minor* has a smaller genome than its $x = 7$ congeners (2.9 vs. an average of 5.7), the estimate for *Milium* is higher than that for *Phleum*, and genomes of the $x = 6$ *Phalaris* are on average bigger than for the $x = 7$ species. Similarly, *Sarga versicolor* (Andropogoneae; $x = 5$) has a 1C value almost three times that of *Sorghum bicolor* and twice that of *Vetiveria*.

In Fig. 4, we present a more detailed view of tribe Triticeae, which includes the largest known genomes in the grasses. We also include the polyploid species for which genome sizes are known. The phylogenetic relationships of the diploid Triticeae are not clear; every gene investigated appears to have had a different history (Kellogg et al., 1996; Mason-Gamer et al., 1998; Mason-Gamer, 2004, and references therein). Accordingly, we have used the cladogram for the plastid genome (Mason-Gamer et al., 2002) as one of several possible inferences of the history of the group. The figure shows that even at this level of analysis, genome size is variable. Also, the sizes of the genomes of the polyploids are often, but not always, smaller than the sum of those of their diploid progenitors. A similar result was found by Jakob et al. (2004) in their detailed analysis of the evolution of genome size in *Hordeum*. The occasional reduction of genome size in polyploids is con-
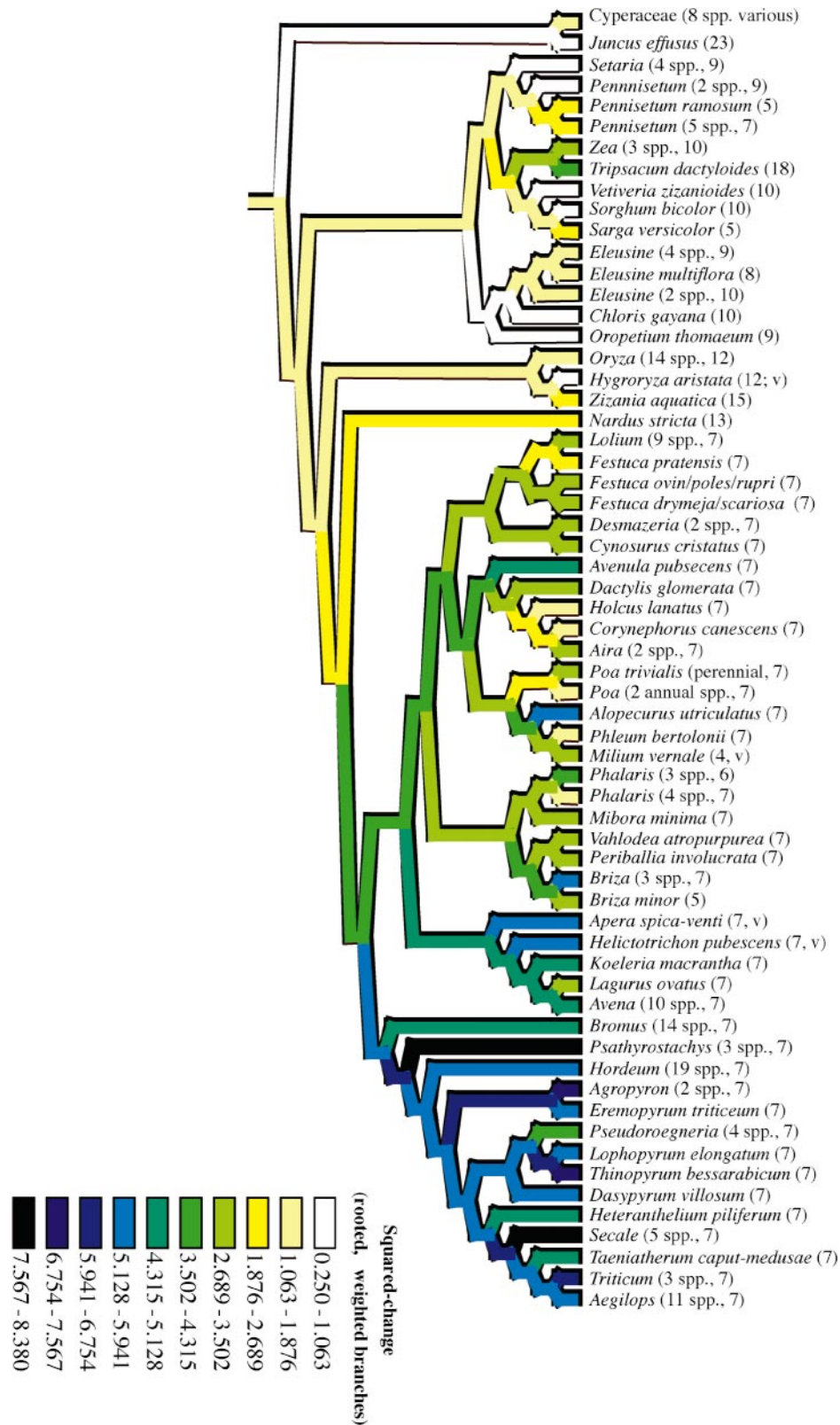
Fig. 3. A hypothesis of the history of diploid grasses showing the evolution of genome size (1C). Data on genome size are from the Angiosperm C-values Database (Bennett and Leitch, 2003). Relationships among the major clades are based on the Grass Phylogeny Working Group (2001) tree. The cladogram for subfamily Pooideae follows Soreng and Davis (2000); cladogram of *Festuca* and *Lolium* follows Torrecilla and Catalán (2002) and Catalán et al. (2004); cladogram of Triticeae follows Kellogg et al. (1996). Base chromosome numbers are indicated in parentheses. Number of species = the number of species for which genome size estimates were included in the average generic value; v = voucher available.
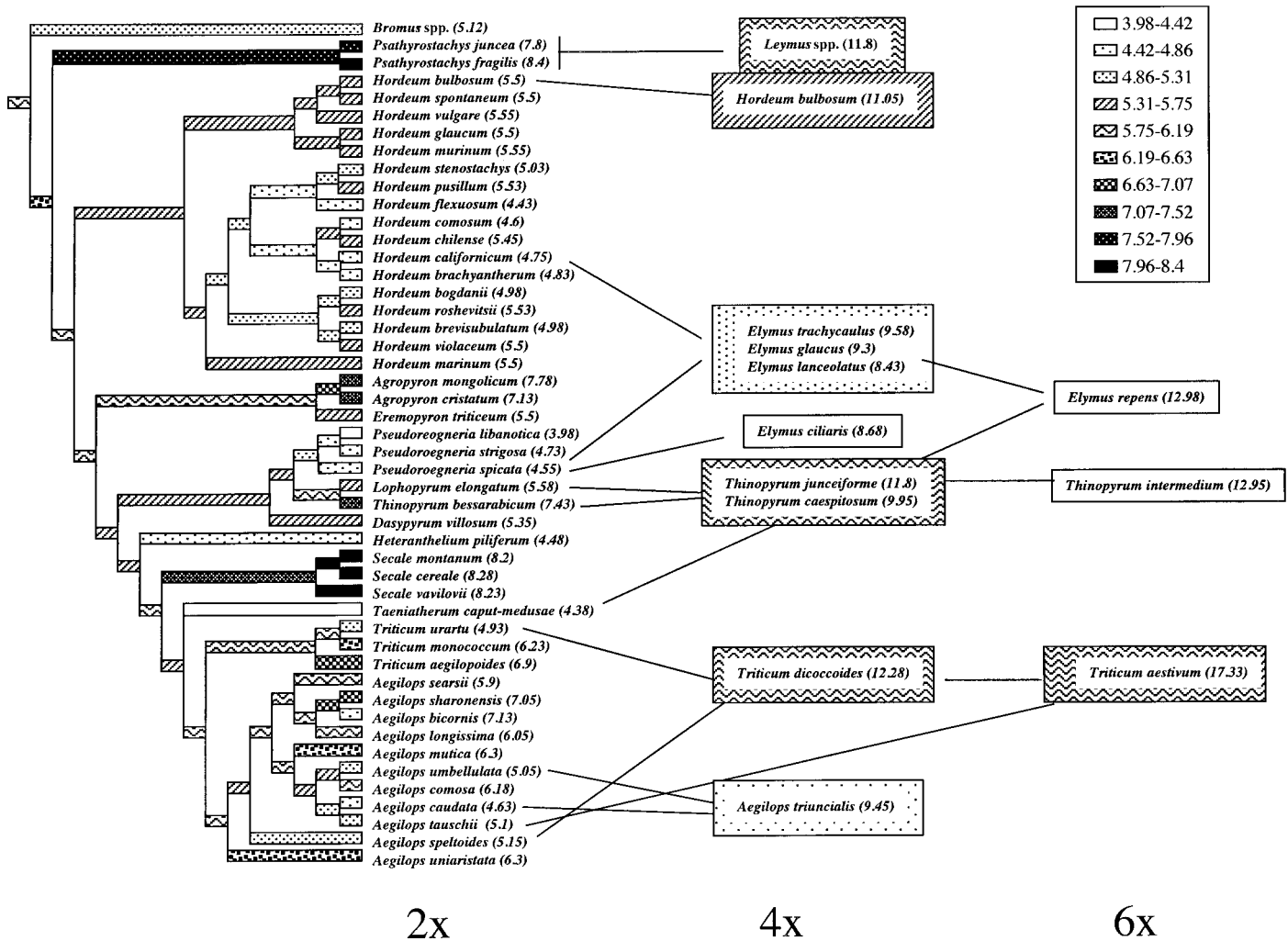
Fig. 4. One estimate of the phylogeny of Triticeae showing the evolution of genome size (1C) among diploid, tetraploid, and hexaploid members of the tribe. Relationships among species of *Hordeum* are based on data from Jakob et al. (2004), which should be consulted for a much more comprehensive study of genome size evolution in that genus; relationships among major clades of Triticeae are based on the plastid tree, as estimated by Mason-Gamer et al. (2002); these relationships are different from those shown by 5S RNA spacers, the ITS, and *waxy*, which are all different from each other. Several polytomies have been resolved in a way that is compatible with one or more gene trees, but results should be considered illustrative.

sistent with the hypothesis that polyploidy may be followed by loss of genetic material (discussed later). This also extends the observation of Levy and Feldman (2002) on grasses in general, in which they observed that the average genome size of polyploids was less then twice the average for the diploids.

In summary, we now know that size variation in plant nuclear genomes is mechanistically driven by polyploidy, transposable element amplification, and recurrent DNA deletion. Because closely related plant lineages can differ dramatically in both genome size and trends in genome size (see later), it is likely that some or all of these mechanisms vary in their intensity in different lineages.

The role of natural selection in determining genome size is unknown. As we come to understand better the mechanisms that control genome size, we may be able to develop clear testable hypotheses. Previous authors (e.g., Cavalier-Smith, 1985) have suggested that selection operates on nuclear volume and that this indirectly affects genome size. Data bearing on this hypothesis are mostly correlational rather than experimental, but the correlations are far from perfect (see multiple

examples given by John and Miklos, 1988), and attempts to link genome size and phenotypic characters have not always been successful (e.g., Bachmann et al., 1985). Recent data on genome structure point to many more levels at which selection might act. For example, selection on the ability of the plant to limit transposon activity might keep genomes from expanding indefinitely. Selection could also act on the epigenetic mechanisms that silence genes and genomes in polyploids. Until we know more details of how such mechanisms work, it is difficult to devise a test (experimental or statistical) of the hypothesis of selection.

POLYPLOIDY

A major and well-known form of genomic change is the duplication of entire genomes or polyploidy. Polyploidy has been the subject of several recent reviews (Soltis and Soltis, 2000; Wendel, 2000; Levy and Feldman, 2002; Cronn and Wendel, 2003), and interested readers should consult those papers for a full exploration of the subject. These papers together

cited evidence that polyploids occur in all major plant groups and that repeated origins of polyploidy are common (Soltis and Soltis, 1993, 1999; Soltis et al., 1992).

Genome sizes of allopolyploids are not necessarily arithmetic sums of the sizes of the parental genomes. Genetic material is lost and genomes are rearranged. For example, *Triticum dicoccoides* is an amphidiploid resulting from a cross between a progenitor similar to *Triticum urartu* (with the A genome) and one similar to *Aegilops speltoides* (with a genome similar to the B genome). Belyayev et al. (2000) showed that genomic probes made from *Aegilops speltoides* hybridize strongly to large portions of the B genome of *T. dicoccoides*, but also to dispersed locations in the A genome, suggesting that repetitive sequences from the B genome have been preferentially amplified in the A genome or have moved from one location in the genome to another. Other examples are cited by Wendel (2000).

Like most other eukaryotes, plants undergo cycles of polyploidization, followed by diploidization, the latter characterized by gene loss and/or pseudogene formation. Even such "good" diploids as *Arabidopsis* and rice are now known to be ancient polyploids or at least to have undergone extensive segmental duplication (discussed see next). Thus previous attempts to estimate the percentage of polyploid species among angiosperms (Stebbins, 1971; Masterson, 1994) are now seen to be oversimplified; many "diploid" species are paleopolyploid. It appears that 100% of flowering plants are current polyploids or have a polyploid history. In the next few paragraphs, we cite examples from major plant families, in which well-known and well-documented duplication events are superimposed on more ancient duplications.

Polyploidy in Poaceae—*Duplication at the base of the family*—The grasses all apparently share a number of ancient duplications of at least parts of their genomes. Draft genome sequences have been published for rice (Goff et al., 2002; Yu et al., 2002), but the genome has only recently been assembled into a pseudomolecule and is still being annotated. It has thus been difficult to check on earlier suggestions of genome duplications. Recently, Vandepoele et al. (2003) have produced genomic scaffolds for the rice genome and used these to estimate the number and extent of duplicated blocks. Only 15% of the rice genome falls into identifiable duplicated blocks, which is appreciably less than the 60% estimate for *Arabidopsis*. Furthermore, these duplications involve only a few of the chromosomes. Finally, a plot of the percentage of duplicated genes against the number of substitutions per silent site did not produce an obvious peak, as would be expected if many of the genes in the genome had been duplicated around the same time. Nonetheless, most of the duplications preceded the diversification of the cereal crops, indicating that they occurred before the origin of the grasses.

The maize duplication—Polyploidy of maize was suggested by Edgar Anderson in 1945 and has been documented by Rhoades (1951), Helentjaris et al. (1988), and Wendel et al. (1989), among others. Anderson speculated that the ancestors might be five-chromosome species of *Sorghum* and *Coix*, ignoring the obvious difficulty that five-chromosome *Sorghum* (now placed in the genus *Sarga*; Spangler, 2003) is native to Australia and Africa and *Coix* to India (Clayton and Renvoize, 1986). Anderson's hypothesis predicts that genes from maize should be sister to genes from one or more of the $x = 5$

Andropogoneae, but instead all analyses to date placed *Zea* sister to *Tripsacum* (e.g., *waxy*, Mason-Gamer et al., 1998; phytochrome B, Mathews et al., 2002; *teosinte branched 1*, Lukens and Doebley, 2001; *ndhF*, Spangler et al., 1999; S. Kleweis, S. Malcomber, and E. A. Kellogg, University of Missouri-St. Louis, unpublished data). The sister group of these two is not well supported in any estimate of phylogeny to date, but no study has linked the *Zea-Tripsacum* clade with *Sarga* (= Australian "sorghum") or with *Coix* (Mason-Gamer et al., 1998; Spangler et al., 1999; Giussani et al., 2001; Mathews et al., 2002; Aliscioni et al., 2003; S. Kleweis, S. Malcomber, and E. A. Kellogg, University of Missouri-St. Louis, unpublished data).

Gaut and Doebley (1997), in a much cited paper, inferred that maize is a segmental allotetraploid. Their hypothesis predicts that trees of orthologous genes should produce one of two alternative patterns—either (maize [maize, sorghum]) or ([maize, maize] sorghum) ([M {M, S}] or [{M, M} S])—and that linked genes should share the same pattern. Gaut and Doebley (1997) could not undertake such a comparison because of the lack of corresponding sequence from sorghum and rice. Recently, however, Swigonova et al. (in press) investigated six pairs of maize genes and their orthologues from sorghum and rice. Phylogenetic analyses are equivocal. Only two of the genes (*r1/b1*; *grf1/grf2*) strongly supported the (M [M, S]) tree, and one (*orp1/orp2*) strongly supported the alternative. The remaining genes produced equivocal trees that were not significantly different from a trichotomy, indicating that the ancestor of the two maize genomes arose about the same time as their divergence from the ancestor of sorghum (about 11 million years ago), consistent with the rapid radiation of the tribe. Wilson et al. (1999) also argued for the ([M M] S) tree, but hypothesized that the maize ancestors had $x = 8$, a number that would be highly unusual for a panicoid grass.

Fertilization independent endosperm (*fie*) genes in maize showed that *fie2* was more closely related to sorghum than either gene is to *fie1* supporting the (M [M, S]) tree (Danilevskaya et al., 2003). However, Swigonova et al. (in press) have shown that *fie1* and *fie2* in maize are not orthologous, despite being on duplicated segments of the genome. Instead, each region originally contained two paralogous genes; one copy was lost from chromosome 4 and the other from chromosome 10. The lack of orthology may also help explain the unusually long branch leading to *Zmfie1* in the figure in Danilevskaya et al. (2003).

Duplications in other groups of grasses—Grass chromosome numbers have been studied in detail since the synthesis provided by Avdulov (1931). Comparative genome mapping efforts have expanded on his observations. Core Pooideae are marked by having their genes arranged in seven large chromosomes; in addition, all members studied so far have one chromosome that corresponds to a combination of rice linkage groups 5 and 10 and another that corresponds to a novel combination of parts of rice 4 and 7 (Kellogg, 1998). Panicoideae are divided into three major clades (Giussani et al., 2001), corresponding to chromosome base numbers of $x = 10$ (Andropogoneae and Paspaleae) and $x = 9$ (Paniceae s.s.). All studied panicoids share a chromosome that corresponds to rice chromosomes 3 and 10 and another that corresponds to rice 7 and 9.

The woody bamboos (tribe Bambuseae) are almost all polyploid, with the exception of *Chusquea talamancensis* and pos-

sibly *C. subtesselata* (Judziewicz et al., 1999). It is tempting to speculate that the many morphological novelties of the group were generated by major changes in genome structure and gene expression following polyploidy. This hypothesis predicts that the two diploid species are derived rather than ancestral and that the woody bamboos all have two full copies of a rice-like genome. Because of the enormous difficulty of doing genetic studies on woody bamboos, such investigations are a long way off.

Zizania, in tribe Oryzeae, shows that duplications need not encompass the entire genome. *Zizania aquatica* (North American wild-rice), with 15 chromosomes, has 14 chromosomes that are colinear with 11 of the 12 chromosomes of *Oryza sativa* and three chromosomes that are apparently duplicates of rice chromosomes 1, 4, and 9 (Kennard et al., 1999). *Zizania* is more closely related to *Oryza* than to any other mapped cereal (Ge et al., 2002). Nonetheless, there have been a number of rearrangements, all of which appear to involve duplicated loci.

Polyploidy in Brassicaceae

—Recent studies in Brassicaceae have been ably reviewed by Koch et al. (2003). Here we summarize a few of the major new findings.

Duplication at the base of the family—The whole genome sequence of *Arabidopsis* made it possible to analyze the patterns of gene duplication across this supposedly compact diploid genome. Surprisingly, it became clear that the genome contained extensive duplicated blocks of sequence (Arabidopsis Genome Initiative, 2000; Blanc et al., 2000; Paterson et al., 2000). Initial attempts to date this duplication using molecular-clock estimates indicated four rounds of genome duplication (Vision et al., 2000). Subsequent analyses have verified the most recent duplication event, although the estimated date is now thought to be more recent (Blanc et al., 2003). At least one and possibly two more ancient duplications can be demonstrated (Simillion et al., 2002). However, a more powerful approach uses phylogenetic events to provide relative dates (Bowers et al., 2003; Ermolaeva et al., 2003). The phylogenetic analyses show that the duplication preceded the divergence of *Brassica* from *Arabidopsis*. Because the *Brassica* and *Arabidopsis* lineages diverged soon after the origin of Brassicaceae (M. Beilstein, University of Missouri, St. Louis, unpublished data), we infer that the ''*Arabidopsis*'' duplication is actually a Brassicaceae duplication.

Comparisons of the degenerated ''homoeologous'' genomes in *Arabidopsis thaliana* have been informative (Blanc et al., 2000; Ku et al., 2000; Arabidopsis Genome Initiative, 2000). However, these homoeologous genomes are highly rearranged within *Arabidopsis*, partly because the multiple rounds of polyploidy were all ancient events and partly because the polyploid state apparently removed constraints against extensive gene loss. Hence, the current status of the *Arabidopsis* genome indicates rearrangements within rearrangements within rearrangements, making it difficult to sort out the nature, timing, and mechanisms of individual events.

Arabidopsis—*Arabidopsis*, as currently delimited, includes nine species and five subspecies (O'Kane and Al-Shehbaz, 1997). One species, *Arabidopsis suecica*, is clearly a recently formed polyploid, the product of a naturally occurring cross between *A. thaliana* and *A. arenosa* (O'Kane et al., 1996). In addition, *A. thaliana* has been crossed with the related *A. lyr-*

ata ssp. *petraea*, and the resulting amphidiploid shows some promise as a tool for understanding the mechanisms and immediate results of amphidiploidy (Nasrallah et al., 2000).

Phylogenetic studies (e.g., Koch et al., 2001) have shown that *Arabidopsis* falls into the same major clade as shepherd's purse, *Capsella*, a much closer relationship than previously thought (although the relationship was suggested by Brummitt, 1992). Consistent with this close phylogenetic relationship, Rossberg et al. (2001) found that a 27-kb region of the *Capsella rubella* genome is perfectly colinear with a 31.5-kb region of the *Arabidopsis thaliana* genome. The region includes five genes in the same orientation in both species.

Brassica—The genome of *Brassica oleracea* is extensively rearranged relative to that of *Arabidopsis*, complicating efforts to compare the order of genes and infer ancient patterns of genome duplication (Lukens et al., 2003, and references therein). Even diploid brassicas appear to be duplicated and/or even triplicated, although the evidence for the latter is not clear.

Neo-polyploidy among species of *Brassica* is well documented, and most introductory students learn about the triangle of U (1935). Mapping studies in *Brassica napus*, a naturally occurring amphidiploid of *B. oleracea* × *B. rapa*, have indicated that the genome of the polyploid is colinear with that of the ancestral diploids; no evidence was found for genome rearrangement (Parkin and Lydiate, 1997). A similar result was found for *Brassica juncea*, the amphiploid product of *B. rapa* and *B. nigra* (Axelsson et al., 2000). These results indicate that polyploidy does not necessarily have a destabilizing effect on genome structure in all plant species.

Polyploidy in Rosaceae (Maloideae)

—The rose family is divided into four subfamilies, distinguished by their floral morphology and chromosome number. Maloideae include such economically important species as pears and apples, as well as many other less familiar fruits and ornamentals (e.g., shadbushes, hawthorns). Most species of Maloideae have a base chromosome number of 17, which was suggested (Sax, 1931, 1932, 1933) to be an ancient polyploid based on a cross between a member of subfamily Amygdaloideae (cherries and apricots; $x = 8$) and subfamily Spiraeoideae (bridalwreath; most with $x = 9$). The argument was principally arithmetic: $8 + 9 = 17$. Molecular phylogenetic studies using maternally inherited plastid genes showed that Amygdaloideae are not sister to Maloideae but could not rule out the possibility of wide hybridization.

Using DNA sequences of a low copy nuclear gene (granule bound starch synthase I, GBSSI) Evans and Campbell (2002) have now shown convincingly that Maloideae originated from an ancestor with $x = 9$. They found two copies of GBSSI in most Rosaceae and four in Maloideae, consistent with the allopolyploid hypothesis. However, sequences from the diploid genus *Gillenia* ($x = 9$) were sister to the GBSSI clades of Maloideae, indicating that the ancestor of the maloids was probably spiraeoid. From the phylogenetic data, plus additional morphological similarities among the early-divergent maloids, Evans and Campbell concluded that the ancestral maloid must have had $x = 18$. Chromosome number would have reduced from $x = 18$ to $x = 17$ via dysploidy, with two chromosomes fusing to become one. As genome maps become available for more Rosaceae, it will be interesting to see if indeed one of the chromosomes of apple corresponds to two chromosomes in *Gillenia*.

The older of the two GBSSI duplications preceded divergence of Rosaceae and Rhamnaceae (Evans and Campbell, 2002). It is not known whether this corresponds to a whole genome duplication or just duplication of this single gene, but this could be tested. Because Rosaceae are sister to all other Rosales (Savolainen et al., 2000), the duplication may mark the origin of the order.

Polyploidy in cotton and soybean—*Glycine* is presumed to be a paleopolyploid, based on its chromosome number of $2n = 40$ (Goldblatt, 1981). This is supported by data from genomic mapping studies (Shoemaker et al., 1996) and phylogenetic data from the nuclear gene glutamine synthetase (Doyle et al., 2003). The latter are consistent with a duplication occurring just before the origin of *Glycine*, although it is not clear whether the duplication corresponded to the whole genome or just the gene being studied.

Cultivated cotton (*Gossypium hirsutum*) is also the result of an allotetraploid event (Beasley, 1940; Brubaker et al., 1999), although presumably much more recent than the one that created the genus *Glycine*. *Gossypium hirsutum*, along with four other tetraploid species, is the result of a cross between an A genome species, similar to *G. arboreum* or *G. herbaceum*, and a D genome species, similar to *G. raimondii*. The implications of this for genome evolution and expression of duplicated genes have been explored extensively (reviewed in Cronn and Wendel, 2003; Wendel and Cronn, 2003).

Endoreduplication—In some cell types in some plants, chromosomal duplication proceeds without cytokinesis, leading to individual cells with multiple copies of the genome; this process is known as endoreduplication. Stebbins (1971) cited multiple examples of this, including water storage cells in the interior of leaves of some succulents and in the suspensor of the embryo (Nagl, 1962). Endoreduplication is also well known in tapetal cells and antipodal cells of some taxa (Stebbins, 1971).

Endoreduplication has been documented in trichomes of a number of taxa including members of Caryophyllaceae, Cucurbitaceae, Geraniaceae, Tropaeolaceae, and Urticaceae (Tschermak-Woess and Hasitschka, 1954), but has been studied most extensively in trichomes of *Arabidopsis thaliana*. In the last species, the amount of DNA in the trichome nucleus correlates with the extent of trichome branching (reviewed by Szymanski et al., 2000). Because trichome branching is an important taxonomic character in Brassicaceae, it is possible that these genome duplications thus directly influence morphological change. Nonetheless, endoreduplicated nuclei are not known to contribute to gametes and hence do not affect subsequent generations.

GENE FAMILIES AND GENE CONTENT

All plant genome projects to date seem to find a remarkably high number of novel genes. Novel in this context means genes that cannot be assigned to any known gene family and without any significant match in another genome. When *Arabidopsis* was the only plant genome for which extensive data were available, the high rate of novelty was ascribed to the lack of other plant genomes. Presumably, a sequence that appeared in *Arabidopsis* but not in animals or bacteria would be characteristic of all plants, but not organisms in other kingdoms. Accumulating sequence data in rice, however, have led

some researchers to propose that close to 50% of the genes in rice have no clear homologue in *Arabidopsis*, although *Arabidopsis* appeared to have far fewer novel genes relative to rice (Goff et al., 2002; Yu et al., 2002). The evolutionary distance between the species is sufficient that rapidly evolving genes might not be detected even if they were truly orthologous. Moreover, more detailed (for instance, manual) annotation of rice genomic regions yielded significantly fewer genes than found in automated annotations (Swigonova et al., in press). Most of the candidate genes found by automated annotation of the rice genome that are re-assigned in subsequent manual annotation turn out to be the "novel" genes predicted in each species (J. Ma, R. Liu, and J. Bennetzen, University of Georgia, unpublished observations). Although some verifiable genes do differ in presence/absence between rice and *Arabidopsis*, it appears that the large majority of those originally proposed as novel are actually artifacts of annotation (Bennetzen, 2002b). Much further research is needed to investigate these questions, not least because the differences in gene content between different species should contribute significantly to their current biology and evolutionary potential.

Because of differences in ploidy, recent and ancient, and in the fates of duplicated gene families, it is expected that gene numbers will differ between plant species. However, because of the limitations of current approaches to gene annotation, especially automated annotation, we do not have precise gene numbers for any plant species. The most accurate estimate to date is for *Arabidopsis*, which has approximately 25 000 genes (Arabidopsis Genome Initiative, 2000). The far greater number (32 000–62 000) predicted for rice (reviewed in Delseny, 2003) is likely to be somewhat inflated by annotation artifacts, as noted earlier.

Many new genes appear, of course, via gene duplication and the formation of gene families. Space does not permit detailed discussion of the evolution of gene families here, although this is an important component of genome evolution. As mentioned in the earlier section on genome rearrangement, unequal recombination and other processes can rapidly duplicate genes to create a gene family and also rapidly remove gene copies (e.g., TCP genes, Reeves and Olmstead, 2003; MADS box genes, Becker and Theissen, 2003). Lynch and Conery (2000) and Moore and Purugganan (2003) estimated that genes duplicate frequently among eukaryotes, with an average rate of fixed duplications of about 0.002 to 0.02 duplication event per gene per million years. Just as genome duplications appear to mark some well-known groups recognized on the basis of morphology, amplification of gene families appears to correlate with the production of evolutionary novelty. Examples include the extensive duplication of R2R3 Myb genes in plants (Rabinowicz et al., 1999), AP3-like MADS box genes in Ranunculaceae (Kramer et al., 2003), and SEP-like genes in grasses (S. Malcomber and E. Kellogg, unpublished data), among many others. We speculate that such gene family amplification is not just a correlate of morphological diversity, but may actually cause it.

PHYLOGENETIC IMPLICATIONS

How do the foregoing data affect our view of phylogenetic history? First, we consider our ability to detect genomic change. Recent polyploidy is one of the easiest genomic modifications to detect, particularly among closely related species. Examples are numerous, beginning with the early observation

of *Primula* × *kewensis* by Newton and Pellew (1929). Many cultivated species are known to be polyploid. Examples include *Nicotiana tabacum* (Goodspeed and Clausen, 1928) and cultivated blueberries (Darrow et al., 1944). Many more examples are cited by Soltis and Soltis (1993, 1999, 2000). The last few years have seen studies of the origins of polyploids in *Dactylorhiza* (Orchidaceae; Hedrén et al., 2001), *Erythronium* (Liliaceae; Allen, 2001), *Nicotiana* (Solanaceae; Chase et al., 2003), *Sorbus* (Rosaceae; Robertson et al., 2004), *Spartina* (Poaceae; Baumel et al., 2001), and *Tarasa* (Malvaceae; Tate and Simpson, 2003), to cite just a few published since the most recent reviews.

Allopolyploidy is also known among species in closely related genera. This has been particularly well documented in grasses. Bread wheat (*Triticum aestivum*) is an allohexaploid derived from three diploid progenitors similar to current *Triticum monococcum* ssp. *urartu* (A genome donor), *Aegilops tauschii* (D genome donor), and *A. speltoides* (similar but not identical to the B genome; McFadden and Sears, 1946; Lilienfield, 1951; Kihara, 1954). Despite their ability to form allopolyploids, *Aegilops* and *Triticum* are conventionally treated as separate genera (see for example, van Slageren, 1994). Other similar cases have been demonstrated. For example, the tetraploid species of *Elymus* are clearly derived from species of *Pseudoroegneria* and *Hordeum*, as shown by cytogenetic studies (e.g., Dewey, 1984 and references therein) and DNA sequence data (Mason-Gamer, 2001; Mason-Gamer et al., 1998, 2002). *Zingeria trichopoda* is thought to be an allopolyploid derivative of *Z. beibersteiniana* and another genus, possibly *Colpodium* (Kotseruba et al., 2003).

We expect that the taxonomic origins of many more polyploids will be elucidated in coming years. Relatively few plant taxonomists have used nuclear introns for phylogenetic studies (but see Koch et al., 2001; Cronn et al., 2002; Malcomber, 2002; Doyle et al., 2003), but these markers are uniquely powerful in documenting the origins of polyploid genomes (e.g., Mason-Gamer, 2001; Evans and Campbell, 2002; Ingram and Doyle, 2003). Plastid genes, although widely used, are maternally inherited, so they will only show one parent of a polyploid; in addition, they are often not variable enough to resolve relationships (e.g., Cronn et al., 2002; Malcomber, 2002). Ribosomal genes, such as the internal transcribed spacer, exhibit intricate patterns of paralogy, are subject to gene conversion, and require a major investment of time to sort out the complexities of their molecular evolution (Álvarez and Wendel, 2003; Bailey et al., 2003; Chase et al., 2003; Razafimandimbison et al., 2004). As suggested by Álvarez and Wendel (2003), low-copy nuclear genes will be most informative for studies of the evolution of genomes.

In contrast to the ease of detecting recent polyploidy, detecting more complex genome rearrangements requires comparative mapping studies, such as have been done in grasses. Mapping populations need to be developed for each species studied, and then these need to be mapped with markers that can be compared across the entire group of species. Highly repetitive markers (SSRs and microsatellites) can be used in some groups (e.g., *Helianthus*; see Rieseberg et al., 2003; Burke et al., 2004), but it is really low-copy sequences—notably genes—that have permitted the broad comparisons among grasses. These markers have been mostly RFLPs, requiring laborious gel blot hybridization procedures to screen populations. As they are converted to PCR-based markers, comparative mapping will become easier. Nonetheless, the production of the original cross and mapping population will continue to be the limiting factor.

The highest level of resolution comes from genomic sequencing, and indeed many of the results described have been the direct outcome of the *Arabidopsis* and rice genome projects. As broad tracts of genomic sequence become available for more and more species, we anticipate that even smaller and rarer genomic changes will be detectable.

Any character that varies among organisms can be used for phylogeny reconstruction, but its value will depend in part on the frequency of change and the likelihood of convergent or parallel evolution. Major genomic rearrangements appear to be rare and also appear to correlate with phylogeny, such as the observation of particular linkage groups that characterize major lineages of grasses (Kellogg, 1998; see earlier comments). On the other hand, many genomic changes are known to have occurred in parallel. For example, both natural and artificial *Helianthus* hybrids show similar genome rearrangements (Rieseberg et al., 1996, 2003), even though the origins of the novel arrangements are known to be independent. In Triticeae, a translocation has occurred between chromosomes 4L and 5L in the A genome of bread wheat, and in *Aegilops umbellulata*, *Thinopyrum bessarabicum*, and *Secale cereale*, but the untranslocated arrangement is observed in the B and D genomes of bread wheat and in their diploid relatives (*A. tauschii* and *A. speltoides*; Devos et al., 1993). The species with the translocation are not monophyletic in any known gene tree, and certainly no biologist today would consider polyploidy rare or unique; there are many examples of repeated origins of similar allopolyploids (see Soltis and Soltis, 1993 and references therein).

Similarities in underlying genome organization can be detected only if relatively few rearrangements have occurred. After multiple inversions, translocations, etc., later events obscure the patterns of the earlier ones, just as multiple hits may confound attempts at comparing sequences. The problem of multiple rearrangements is compounded by problems created by gene loss and duplication, both of which occur continually.

If the taxonomic distribution of a genome rearrangement conflicted with a set of congruent and well-supported gene trees, it seems unlikely that the genome character would override the accumulated evidence of nucleotide characters. More likely, this would be interpreted as evidence that some aspect of genome structure created a predisposition for particular rearrangements. Major genome rearrangements are few in number relative to point mutations; more importantly, their distribution across a large number of taxa is difficult to assess. For logistical reasons, therefore, phylogenetic inference will likely continue to rest on large amounts of sequence data. Changes in genome structure, however, are likely to be important as corroboration of an estimate of phylogeny and as support for otherwise weakly supported groupings.

DYNAMIC GENOMES

Genome size, chromosome number, and ploidy are all snapshots of the current state of a set of continuous processes, like stills from a reel of film. Genome evolution is a continuous and dynamic process in which genomes expand and contract, duplicate in whole or in part, and gain and lose genes at the same time as transposons amplify and are excised. Genome expansion, genome rearrangement, hybridization, and allopolyploidy all may be causally related to each other. Liu and

Wendel (2000) investigated the effect of introgressing a small amount of genetic material from *Zizania latifolia* into *Oryza sativa*. They found an increased number of copia-like and gypsy-like retrotransposons, indicating that these elements had not only been transcriptionally activated but also copied themselves extensively.

Studies of gene expression provide a link among the processes of genome evolution. Duplicated genes—whether in polyploids or in tandem duplications—may be silenced or activated (see review by Liu and Wendel, 2003). Adams et al. (2003) have shown changes in gene expression in natural and synthetic polyploids of cotton relative to their diploid progenitors. Kashkush et al. (2002, 2003), in their work on wheat, provided a possible explanation for the phenomenon seen in *Gossypium*. They show that retrotransposons in wheat become actively transcribed in synthetic allopolyploids. In some of these retroelements, the LTR of the transposon provides a promoter or enhancer sequence such that adjacent genes are transcribed. Thus some genes become more highly expressed than in the diploid progenitors, whereas expression levels of others are reduced, possibly by anti-sense transcripts. Perhaps surprisingly, they did not find new insertions of transposons in their limited study, but just more active transcription.

Not all genome combinations and rearrangements occur, however. Limits to hybridization ability have been extensively documented (see for example many references in Stebbins [1950] and Grant [1981]). Limits to allopolyploid formation also exist; polyploid formation is not indiscriminate. Once again, some of the best examples come from the grasses. Allopolyploidy is amply documented and common among members of tribe Triticeae, but until recently there was no evidence for involvement of taxa outside the tribe. In fact, *Psathyrostachys*, which in some phylogenetic reconstructions is sister to the rest of the tribe, produces polyploids among its species (which are classified in the genus *Leymus*), but the N genome of *Psathyrostachys* does not seem to be involved in any of the intricate amphiploids that are so well documented elsewhere in the tribe (Bödvarsdóttir and Anamthawat-Jónsson, 2003; Wu et al., 2003). Similarly, the high complexity of Triticeae does not appear to extend to species in the sister clade, Bromeae, even though *Bromus* species will form polyploids among themselves. Similar ''limited'' allopolyploidy is documented in the clade containing foxtail millet and its relatives; polyploids involve a number of genera closely related to *Setaria italica* but do not appear to involve the sister clade which includes *Pennisetum/Cenchrus* (A. Penly and E. A. Kellogg, unpublished data).

The one exception to the foregoing is the allohexaploid *Elymus repens* (Triticeae), recently documented in detail by Mason-Gamer (2004). The species contains genes similar to those of diploid species of *Pseudoroegneria* and *Hordeum*; this was expected based on chromosome pairing studies. However, nuclear gene trees also showed that species related to *Taeniatherum* (also Triticeae) contributed to the genome of *E. repens*, which was unexpected. Even more surprising, some genes appear similar to distantly related species in the tribe Poeae. As more nuclear genes are investigated in allopolyploids, it will be interesting to know whether this species is exceptional or whether truly wide hybridization really is common.

Duplication of genes and genomes can produce gene trees that are often incongruent with each other, and evolution appears reticulate. Between such groups, however, gene trees are largely in agreement, and evolution appears divergent. Thus

groups of species that can form hybrids and polyploids (''combinable groups'') are reminiscent of coalescent species (Baum and Shaw, 1995), although obviously much larger than what is commonly named as a species.

The multiple examples cited show that major genomic changes correlate with the origin of major clades. The maloid duplication, the Brassicaceae duplication, the Poaceae duplication (segmental or otherwise), and the genome rearrangement in the ancestor of the pooid grasses all indicate that genome rearrangements might be phylogenetically informative. More importantly, we hypothesize that groups that we now call families or major clades correspond to major genome rearrangements or polyploidization events. Far from being a dead end, rearrangement, polyploidy, and gene duplication are all apparently processes of genome regeneration. Transposon amplification, gene loss, and changes in gene expression (the latter perhaps induced by epigenetic changes such as alterations in DNA methylation and histone modification) alter both the structure of the genome and the function of genes. We are only beginning to glimpse how these processes and events correlate with patterns of diversification and how they may provide the substrate for evolution.

LITERATURE CITED

ADAMS, K. L., R. CRONN, R. PERCIFIELD, AND J. F. WENDEL. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences, USA* 100: 4649–4654.

ALISCIONI, S. S., L. M. GIUSSANI, F. O. ZULOAGA, AND E. A. KELLOGG. 2003. A molecular phylogeny of *Panicum* (Poaceae: Paniceae): test of monophyly and phylogenetic placement within the Panicoideae. *American Journal of Botany* 90: 796–821.

ALLEN, G. A. 2001. Hybrid speciation in *Erythronium* (Liliaceae): a new allotetraploid species from Washington State. *Systematic Botany* 26: 263–272.

ÁLVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

ANANIEV, E. V., R. L. PHILLIPS, AND H. W. RINES. 1998. Complex structure of knob DNA on maize chromosome 9: Retrotransposon invasion into heterochromatin. *Genetics* 149: 2025–2037.

ANDERSON, E. 1945. What is *Zea mays*? A report of progress. *Chronica Botanica* 9: 88–92.

ARABIDOPSIS GENOME INITIATIVE. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.

AVDULOV, N. P. 1931. Kario-sistematicheskoye issledovaniye semeystva zlakov. *Trudy po Prikladnoj Botanike i Prilozheniye* 44: 1–352 [Karyosystematic studies in the grass family. Supplement 44 to *Bulletin of Applied Botany, Genetics, and Plant-Breeding, Leningrad*. Russian 1–352; German summary, 353–425; index 426–428. English translation (mislabeled as supplement 43) published by the Smithsonian Institution and the National Scientific Documentation Centre, New Delhi. 1975. TT 70–53085.].

AVRAMOVA, Z., A. TIKHONOV, M. CHEN, AND J. L. BENNETZEN. 1998. Matrix-attachment regions and structural colinearity in the genomes of two grass species. *Nucleic Acids Research* 26: 761–767.

AXELSSON, T., C. M. BOWMAN, A. G. SHARPE, D. J. LYDIATE, AND U. LAGERCRANTZ. 2000. Amphidiploid *Brassica juncea* contains conserved progenitor genomes. *Genome* 43: 679–688.

BACHMANN, K., K. L. CHAMBERS, AND H. J. PRICE. 1985. Genome size and natural selection: observations and experiments in plants. *In* T. Cavalier-Smith [ed.], The evolution of genome size, 267–276. John Wiley and Sons, Chichester, UK.

BAILEY, C. D., T. G. CARR, S. A. HARRIS, AND C. E. HUGHES. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29: 435–455.

BAUM, D. A., AND K. L. SHAW. 1995. Genealogical perspectives on the species problem. *In* P. C. Hoch and A. G. Stephenson [eds.], Experi-

mental and molecular approaches to plant biosystematics, 289–303. Missouri Botanical Garden, St. Louis, Missouri, USA.

BAUMEL, A., M. L. AINOUCHE, AND J. E. LEVASSEUR. 2001. Molecular investigations in populations of *Spartina anglica* C.E. Hubbard (Poaceae) invading coastal Brittany (France). *Molecular Ecology* 10: 1689–1701.

BEASLEY, J. O. 1940. The origin of American tetraploid *Gossypium* species. *American Naturalist* 74: 285–286.

BECKER, A., AND G. THEISSEN. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution* 29: 464–489.

BELMONT, A. S. 2003. Dynamics of chromatin, proteins, and bodies within the cell nucleus. *Current Opinion in Cell Biology* 15: 1–7.

BELYAYEV, A., O. RASKINA, A. KOROL, ET AL. 2000. Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. *Genome* 43: 1021–1026.

BENNETT, M. D. 1987. Ordered disposition of parental genomes and individual chromosomes in reconstructed plant nuclei, and their implications. *Somatic Cell and Molecular Genetics* 13: 463–466.

BENNETT, M. D. 1988. Parental genome separation in F1 hybrids between grass species. *Kew Chromosome Conference* III: 195–208.

BENNETT, M. D., AND I. J. LEITCH. 2003. Angiosperm DNA C-values database, release 4.0, available at, website: http://www.rbgkew.org.uk/cval/homepage.html.

BENNETZEN, J. L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology* 42: 251–269.

BENNETZEN, J. L. 2002a. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 115: 29–36.

BENNETZEN, J. L. 2002b. Opening the door to comparative plant biology. *Science* 296: 60–63.

BENNETZEN, J. L., AND M. FREELING. 1993. Grasses as a single genetic system: genome composition, collinearity and complementarity. *Trends in Genetics* 9: 259–261.

BENNETZEN, J., AND E. A. KELLOGG. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9: 1509–1514.

BENNETZEN, J. L., AND J. MA. 2003. The genetic colinearity of rice and other cereals based on genomic sequence analysis. *Current Opinion in Plant Biology* 6: 128–133.

BENNETZEN, J. L., J. MA, AND K. M. DEVOS. In press. Mechanisms of recent genome size variation in flowering plants. *Annals of Botany.*

BENNETZEN, J. L., AND W. RAMAKRISHNA. 2002. Exceptional haplotype variation in maize. *Proceedings of the National Academy of Sciences, USA* 99: 9093–9095.

BENNETZEN, J. L., P. SANMIGUEL, M. CHEN, A. TIKHONOV, M. FRANCKI, AND Z. AVRAMOVA. 1998. Grass genomes. *Proceedings of the National Academy of Sciences, USA* 95: 1975–1978.

BENNETZEN, J. L., K. SCHRICK, P. S. SPRINGER, W. E. BROWN, AND P. SANMIGUEL. 1994. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* 37: 565–576.

BLANC, G., A. BARAKAT, R. GUYOT, R. COOKE, AND M. DELSENY. 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* 12: 1093–1101.

BLANC, G., K. HOKAMP, AND K. H. WOLFE. 2003. A recent polyploidy superimposed on older large-scale duplication in the *Arabidopsis* genome. *Genome Research* 13: 137–144.

BÖDVARSDÓTTIR, S. K., AND K. ANAMTHAWAT-JÓNSSON. 2003. Isolation, characterization, and analysis of *Leymus*-specific DNA sequences. *Genome* 46: 673–682.

BONIERBALE, M. W., R. L. PLAISTED, AND S. D. TANKSLEY. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120: 1095–1103.

BOWERS, J. E., C. ABBEY, S. ANDERSON, C. CHANG, X. DRAYE, A. H. HOPPE, R. JESSUP, C. LEMKE, J. LENNINGTON, Z. LI, Y.-R. LIN, S.-C. LIU, L. LUO, B. S. MARLER, R. MING, S. E. MITCHELL, D. QIANG, K. REISCHMANN, S. R. SCHULZE, D. N. SKINNER, Y.-W. WANG, S. KRESOVICH, K. F. SCHERTZ, AND A. H. PATERSON. 2003. A high-density genetic recombination map of sequence-tagged sites for *Sorghum*, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. *Genetics* 165: 367–386.

BRUBAKER, C. L., A. H. PATERSON, AND J. F. WENDEL. 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42: 184–203.

BRUMMITT, R. K. 1992. Vascular plant families and genera. Royal Botanic Gardens, Kew, UK.

BURKE, J. M., Z. LAI, M. SALMASO, T. NAKAZATO, S. TANG, A. HEESACKER, S. J. KNAPP, AND L. H. RIESEBERG. 2004. Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* 167: 459–470.

CARLSON, W. R. 1986. The B chromosome of maize. *Critical Reviews in Plant Sciences* 3: 201–226.

CATALÁN, P., P. TORRECILLA, J. A. LÓPEZ RODRIGUEZ, AND R. G. OLMSTEAD. 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31: 517–541.

CAVALIER-SMITH, T. 1985. Cell volume and the evolution of eukaryotic genome size. *In* T. Cavalier-Smith [ed.], The evolution of genome size, 105–184. John Wiley and Sons, Chichester, UK.

CHASE, M. W., S. KNAPP, A. V. COX, J. J. CLARKSON, Y. BUTSKO, J. JOSEPH, V. SAVOLAINEN, AND A. S. PAROKONNY. 2003. Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Annals of Botany* 92: 107–127.

CLAYTON, W. D., AND S. A. RENVOIZE. 1986. Genera graminum. Her Majesty's Stationery Office, London, UK.

COLD SPRING HARBOR LABORATORY, WASHINGTON UNIVERSITY GENOME SEQUENCING CENTER, AND PE BIOSYSTEMS ARABIDOPSIS SEQUENCING CONSORTIUM. 2000. The complete sequence of a heterochromatic island from a higher eukaryote. *Cell* 100: 377–386.

COOPER, G. M., M. BRUDNO, NISC COMPARATIVE SEQUENCING PROGRAM, E. D. GREEN, S. BATZOGLOU, AND A. SIDOW. 2003. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Research* 13: 813–820.

CRONN, R., R. L. SMALL, T. HASELKORN, AND J. F. WENDEL. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89: 707–725.

CRONN, R. R., AND J. F. WENDEL. 2003. Cryptic trysts, genomic mergers, and plant speciation. *New Phytologist* 161: 133–142.

DANILEVSKAYA, P. HERMON, S. HANTKE, M. G. MUSZYNSKI, K. KOLLIPARA, AND E. V. ANANIEV. 2003. Duplicated *fie* genes in maize: expression pattern and imprinting suggest distinct functions. *Plant Cell* 15: 425–438.

DARROW, G. M., W. H. CAMP, H. E. FISCHER, AND H. DERMEN. 1944. Chromosome numbers in *Vaccinium* and related groups. *Bulletin of the Torrey Botanical Club* 71: 498–506.

DELSENY, M. 2003. Towards an accurate sequence of the rice genome. *Current Opinion in Plant Biology* 6: 101–105.

DEHAL, P., P. PREDKI, A. S. OLSEN, A. KOBAYASHI, P. FOLTA, S. LUCAS, M. LAND, A. TERRY, C. L. ECALE ZHOU, S. RASH, Q. ZHANG, L. GORDON, J. KIM, C. ELKIN, M. J. POLLARD, P. RICHARDSON, D. ROKHSAR, E. UBERBACHER, T. HAWKINS, E. BRANSCOMB, AND L. STUBBS. 2001. Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. *Science* 293: 104–111.

DEVOS, K. M., M. D. ATKINSON, C. N. CHINOY, H. A. FRANCIS, R. L. HARCOURT, R. M. D. KOEBNER, C. J. LIU, P. MASOJC, D. X. XIE, AND M. D. GALE. 1993. Chromosomal rearrangements in the rye genome relative to that of wheat. *Theoretical and Applied Genetics* 85: 673–680.

DEVOS, K. M., J. BEALES, Y. NAGAMURA, AND T. SASAKI. 1999. *Arabidopsis*—Rice: will colinearity allow gene prediction across the eudicot–monocot divide? *Genome Research* 9: 825–829.

DEVOS, K. M., J. M. BROWN, AND J. L. BENNETZEN. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* 12: 1075–1079.

DEVOS, K. M., Z. M. WANG, J. BEALES, T. SASAKI, AND M. D. GALE. 1998. Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). *Theoretical and Applied Genetics* 96: 63–68.

DEWEY, D. R. 1984. The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. *In* D. R. Dewey [ed.], Gene manipulation in plant improvement, 209–279. Plenum Press, New York, New York, USA.

DOYLE, J. J., J. L. DOYLE, AND C. HARBISON. 2003. Chloroplast-expressed glutamine synthetase in *Glycine* and related Leguminosae: phylogeny, gene duplication, and ancient polyploidy. *Systematic Botany* 28: 567–577.

DUBCOVSKY, J., W. RAMAKRISHNA, P. SANMIGUEL, C. S. BUSSO, L. YAN, B.

A. SHILOFF, AND J. L. BENNETZEN. 2001. Comparative sequence analysis of colinear barley and rice BACs. *Plant Physiology* 125: 1342–1353.

ERMOLAEVA, M. D., M. WU, J. A. EISEN, AND S. L. SALZBERG. 2003. The age of the *Arabidopsis thaliana* genome duplication. *Plant Molecular Biology* 51: 859–866.

EVANS, R. C., AND C. S. CAMPBELL. 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *American Journal of Botany* 89: 1478–1484.

FELSENSTEIN, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* 19: 445–471.

FELSENSTEIN, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts, USA.

FEUILLET, C., AND B. KELLER. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. *Proceedings of the National Academy of Sciences, USA* 96: 8265–8270.

FU, H., AND H. K. DOONER. 2002. Intraspecific violation of genetic colinearity and its implications in maize. *Proceedings of the National Academy of Sciences, USA* 99: 9573–9578.

FU, H., W. PARK, X. YAN, Z. ZHENG, B. SHEN, AND H. K. DOONER. 2001. The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. *Proceedings of the National Academy of Sciences, USA* 98: 8903–8908.

GALE, M. D., AND K. M. DEVOS. 1998. Plant comparative genetics after 10 years. *Science* 282: 656–659.

GAUT, B. S. 2001. Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Genome Research* 11: 55–66.

GAUT, B. S. 2002. Evolutionary dynamics of grass genomes. *New Phytologist* 154: 15–28.

GAUT, B. S., AND J. F. DOEBLEY. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences, USA* 94: 6809–6814.

GE, S., A. LI, B.-R. LU, S.-Z. ZHANG, AND D.-Y. HUANG. 2002. A phylogeny of the rice tribe Oryzeae (Poaceae) based on *matK* sequence data. *American Journal of Botany* 89: 1967–1972.

GILL, K. S., B. S. GILL, T. R. ENDO, AND E. BOYKO. 1996a. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics* 143: 1001–1012.

GILL, K. S., B. S. GILL, T. R. ENDO, AND T. TAYLOR. 1996b. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics* 144: 1883–1991.

GIUSSANI, L. M., J. H. COTA-SÁNCHEZ, F. O. ZULOAGA, AND E. A. KELLOGG. 2001. A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C₄ photosynthesis. *American Journal of Botany* 88: 1993–2012.

GOFF, S. A., D. RICKE, T. H. LAN, G. PRESTING, AND R. L. WANG, ET AL. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92–100.

GOLDBLATT, P. 1981. Cytology and the phylogeny of Leguminosae. *In* R. M. Polhill and P. H. Raven [eds.], Advances in legume systematics, part 2, 427–463. Royal Botanic Gardens, Kew, UK.

GOODSPEED, T. H., AND R. E. CLAUSEN. 1928. Interspecific hybridization in *Nicotiana*. VIII. The *sylvestris-tomentosa-tabacum* triangle and its bearing on the origin of *Tabacum*. *University of California Publications in Botany* 11: 245–256.

GORBUNOVA, V., AND A. A. LEVY. 1999. How plants make ends meet: DNA double-strand break repair. *Trends in Plant Science* 4: 263–269.

GRANT, V. 1981. *Plant speciation*. Columbia University Press, New York, New York, USA.

GRASS PHYLOGENY WORKING GROUP. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88: 373–457.

GUO, H., AND S. P. MOOSE. 2003. Conserved non-coding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. *Plant Cell* 15: 1143–1158.

HEDRÉN, M., M. F. FAY, AND M. W. CHASE. 2001. Amplified fragment length polymorphisms (AFLP) reveal details of polyploid evolution in *Dactylorhiza* (Orchidaceae). *American Journal of Botany* 88: 1868–1880.

HELENTJARIS, T., D. F. WEBER, AND S. WRIGHT. 1988. Duplicate sequences in maize and identification of their genomic locations through restriction fragment length polymorphisms. *Genetics* 118: 353–363.

HESLOP-HARRISON, J. S., AND M. D. BENNETT. 1990. Nuclear architecture in plants. *Trends in Genetics* 6: 401–405.

HONG, R. L., L. HAMAGUCHI, M. A. BUSCH, AND D. WEIGEL. 2003. Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell* 15: 1296–1309.

HULBERT, S. H., T. E. RICHTER, J. D. AXTELL, AND J. L. BENNETZEN. 1990. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proceedings of the National Academy of Sciences, USA* 87: 4251–4255.

ILIC, K., P. J. SANMIGUEL, AND J. L. BENNETZEN. 2003. A complex history of rearrangement in an orthologous region of the maize, sorghum and rice genomes. *Proceedings of the National Academy of Sciences, USA* 100: 12265–12270.

INGRAM, A. L., AND J. J. DOYLE. 2003. The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear *waxy* and plastid *rps16*. *American Journal of Botany* 90: 116–122.

JAKOB, S. S., A. MEISTER, AND F. R. BLATTNER. 2004. The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Molecular Biology and Evolution* 21: 860–869.

JOHN, B., AND G. L. G. MIKLOS. 1988. The eukaryote genome in development and evolution. Allen and Unwin, London, UK.

JUDZIEWICZ, E. J., L. G. CLARK, X. LONDOÑO, AND M. J. STERN. 1999. American bamboos. Smithsonian Institution Press, Washington, D.C., USA.

KAPLINSKY, N. J., D. M. BRAUN, J. PENTERMAN, S. A. GOFF, AND M. FREELING. 2002. Utility and distribution of conserved noncoding sequences in the grasses. *Proceedings of the National Academy of Sciences, USA* 99: 6147–6151.

KASHKUSH, K., M. FELDMAN, AND A. A. LEVY. 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160: 1651–1659.

KASHKUSH, K., M. FELDMAN, AND A. A. LEVY. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* 33: 102–106.

KELLOGG, E. A. 1998. Relationships of cereal crops and other grasses. *Proceedings of the National Academy of Sciences, USA* 95: 2005–2010.

KELLOGG, E. A., R. APPELS, AND R. J. MASON-GAMER. 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). *Systematic Botany* 21: 321–347.

KENNARD, W., R. PHILLIPS, R. PORTER, AND A. GROMBACHER. 1999. A comparative map of wild rice (*Zizania palustris* L. 2n = 2x = 30). *Theoretical and Applied Genetics* 99: 793–799.

KIHARA, H. 1954. Consideration on the evolution and distribution of *Aegilops* species based on the analyzer method. *Cytologia* 19: 336–357.

KIRIK, A., S. SALOMON, AND H. PUCHTA. 2000. Species-specific double-strand break repair and genome evolution in plants. *EMBO Journal* 19: 5562–5566.

KOCH, M., I. A. AL-SHEHBAZ, AND K. MUMMENHOFF. 2003. Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). *Annals of the Missouri Botanical Garden* 90: 151–171.

KOCH, M., B. HAUBOLD, AND T. MITCHELL-OLDS. 2001. Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. *American Journal of Botany* 88: 534–544.

KOTSERUBA, V., D. GERNAND, A. MEISTER, AND A. HOUBEN. 2003. Uniparental loss of ribosomal DNA in the allotetraploid grass *Zingeria trichopoda* (2n = 8). *Genome* 46: 156–163.

KRAMER, E. M., V. S. DISTILIO, AND P. M. SCHLÜTER. 2003. Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. *International Journal of Plant Sciences* 164: 1–11.

KU, H. M., T. VISION, J. P. LIU, AND S. D. TANKSLEY. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences, USA* 97: 9121–9126.

KUHL, J. C., F. CHEUNG, Q. YUAN, W. MARTIN, Y. ZEWDIE, J. MCCALLUM, A. CATANACH, P. RUTHERFORD, K. C. SINK, M. JENDEREK, J. P. PRINCE, C. D. TOWN, AND M. J. HAVEY. 2004. A unique set of 11,008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders Asparagales and Poales. *Plant Cell* 16: 114–125.

LAGERCRANTZ, U. 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150: 1217–1228.

LEVY, A. A., AND M. FELDMAN. 2002. The impact of polyploidy on grass genome evolution. *Plant Physiology* 130: 1587–1593.

LILIENFIELD, F. A. 1951. Genome-analysis in *Triticum* and *Aegilops*. Concluding review. *Cytologia* 16: 101–123.

LIU, B., AND J. F. WENDEL. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome* 43: 874–880.

LIU, B., AND J. F. WENDEL. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution* 29: 365–379.

LIU, H., R. SACHIDANANDAM, AND L. STEIN. 2001. Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order. *Genome Research* 11: 2020–2026.

LLACA, V., AND J. MESSING. 1998. Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. *Plant Journal* 15: 211–220.

LUKENS, L., AND J. DOEBLEY. 2001. Molecular evolution of the *teosinte branched* gene among maize and related grasses. *Molecular Biology and Evolution* 18: 627–638.

LUKENS, L., F. ZOU, D. LYDIATE, I. PARKIN, AND T. OSBORN. 2003. Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* 164: 359–372.

LYNCH, M., AND J. S. CONERY. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.

LYNCH, M., AND J. S. CONERY. 2003. The origins of genome complexity. *Science* 302: 1401–1404.

MA, J., K. M. DEVOS, AND J. L. BENNETZEN. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research* 14: 860–869.

MADDISON, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Zoology* 40: 304–314.

MADDISON, D. R., AND W. P. MADDISON. 2000. MacClade 4.0. Sinauer Associates, Sunderland, Massachusetts, USA.

MALCOMBER, S. T. 2002. Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of a rapid radiation in a widespread, morphologically diverse genus. *Evolution* 56: 42–57.

MASON-GAMER, R. J. 2001. Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Systematic Botany* 26: 757–768.

MASON-GAMER, R. J. 2004. Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Systematic Biology* 53: 25–37.

MASON-GAMER, R. J., N. L. ORME, AND C. M. ANDERSON. 2002. Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. *Genome* 45: 991–1002.

MASON-GAMER, R. J., C. F. WEIL, AND E. A. KELLOGG. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molecular Biology and Evolution* 15: 1658–1673.

MASTERSON, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264: 421–424.

MATHEWS, S., R. J. MASON-GAMER, R. E. SPANGLER, AND E. A. KELLOGG. 2002. Phylogeny of the Andropogoneae inferred from phytochrome B, GBSSI, and *ndhF*. *International Journal of Plant Sciences* 163: 441–450.

MCCLINTOCK, B. 1932. A correlation of ring-shaped chromosomes with variegation in *Zea mays*. *Proceedings of the National Academy of Sciences, USA* 18: 677–681.

MCFADDEN, E. S., AND E. R. SEARS. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *Journal of Heredity* 37: 81–89.

MOORE, G., K. M. DEVOS, Z. WANG, AND M. D. GALE. 1995. Cereal genome evolution. Grasses, line up and form a circle. *Current Biology* 5: 737–739.

MOORE, R. C., AND M. D. PURUGGANAN. 2003. The early stages of duplicate gene evolution. *Proceedings of the National Academy of Sciences, USA* 100: 15682–15687.

NAGL, W. 1962. 4096-Ploidie und ''Riesenchromosomen'' im Suspensor von *Phaseolus coccineus*. *Naturwissenschaften* 49: 261–262.

NASRALLAH, M. E., K. YOGEESWARAN, S. SNYDER, AND J. B. NASARALLAH. 2000. *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiology* 124: 1605–1614.

NEWTON, W. C. F., AND C. PELLEW. 1929. *Primula kewensis* and its derivatives. *Journal of Genetics* 20: 405–467.

O'KANE, S. L., B. A. SCHAAL, AND I. A. AL-SHEHBAZ. 1996. The origins of

Arabidopsis suecica (Brassicaceae) as indicated by nuclear rDNA sequences. *Systematic Botany* 21: 559–566.

O'KANE, S. L., AND I. A. AL-SHEHBAZ. 1997. A synopsis of *Arabidopsis* (Brassicaceae). *Novon* 7: 323–327.

PARKIN, I. A. P., AND D. J. LYDIATE. 1997. Conserved patterns of chromosome pairing and recombination in *Brassica napus* crosses. *Genome* 40: 496–504.

PATERSON, A. H., J. E. BOWERS, M. D. BUROW, X. DRAYE, C. G. ELSIK, C. X. JIANG, C. S. KATSAR, T. H. LAN, Y. R. LIN, AND R. MING. 2000. Comparative genomics of plant chromosomes. *Plant Cell* 12: 1523–1540.

PATERSON, A. H., T. H. LAN, K. P. REISCHMANN, C. CHANG, Y. R. LIN, S. C. LIU, M. D. BUROW, S. P. KOWALSKI, C. S. KATSAR, T. A. DELMONTE, K. A. FELDMANN, K. F. SCHERTZ, AND J. F. WENDEL. 1996. Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genetics* 14: 380–382.

PETROV, D. A., T. A. SANGSTER, J. S. JOHNSTON, D. L. HARTL, AND K. L. SHAW. 2000. Evidence for DNA loss as a determinant of genome size. *Science* 287: 1060–1062.

RABINOWICZ, P. D., E. L. BRAUN, A. D. WOLFE, B. BOWEN, AND E. GROTEWOLD. 1999. Maize *R2R3 Myb* genes: sequence analysis reveals amplification in higher plants. *Genetics* 153: 427–444.

RAHMAN, S., S. ABRAHAMS, D. ABBOTT, Y. MUKAI, M. SAMUEL, M. MORELL, AND R. APPELS. 1997. A complex arrangement of genes at a starch branching enzyme I locus in the D-genome donor of wheat. *Genome* 40: 465–474.

RAMAKRISHNA, W., J. DUBCOVSKY, Y.-J. PARK, C. BUSSO, J. EMBERTON, P. SANMIGUEL, AND J. L. BENNETZEN. 2002. Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. *Genetics* 162: 1389–1400.

RAZAFIMANDIMBISON, S. G., E. A. KELLOGG, AND B. BREMER. 2004. Recent origin and phylogenetic implications of high-level ITS polymorphisms: a case study from Naucleae (Rubiaceae). *Systematic Biology* 53: 177–192.

REEVES, P. A., AND R. G. OLMSTEAD. 2003. Evolution of the TCP gene family in Asteridae: cladistic and network approaches to understanding regulatory gene family diversification and its impact on morphological evolution. *Molecular Biology and Evolution* 20: 1997–2009.

RHOADES, M. M. 1951. Duplicate genes in maize. *American Naturalist* 85: 105–110.

RICHARDS, E. J., AND F. M. AUSUBEL. 1988. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* 53: 127–136.

RIESEBERG, L. H. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* 16: 351–358.

RIESEBERG, L. H., O. RAYMOND, D. M. ROSENTHAL, Z. LAI, K. LIVINGSTONE, T. NAKAZATO, J. L. DURPHY, A. E. SCHWARZBACH, L. A. DONOVAN, AND C. LEXER. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211–1216.

RIESEBERG, L. H., B. SINERVO, C. R. LINDER, M. C. UNGERER, AND D. M. ARIAS. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 272: 741–745.

RIESEBERG, L. H., C. VAN FOSSEN, AND A. DESROCHERS. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375: 313–316.

ROBERTSON, A., A. C. NEWTON, AND R. A. ENNOS. 2004. Multiple hybrid origins, genetic diversity and population genetic structure or two endemic *Sorbus* taxa on the Isle of Arran. *Molecular Ecology* 13: 123–134.

ROSSBERG, M., K. THERES, A. ACARKAN, R. HERRERO, T. SCHMITT, K. SCHUMACHER, G. SCHMITZ, AND R. SCHMIDT. 2001. Comparative sequence analysis reveals extensive microcolinearity in the *Lateral Suppressor* regions of the tomato, *Arabidopsis*, and *Capsella* genomes. *Plant Cell* 13: 979–988.

SALSE, J., B. PIEGU, R. COOKE, AND M. DELSENY. 2002. Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Research* 30: 2316–2328.

SANMIGUEL, P., A. TIKHONOV, Y.-K. JIN, N. MOTCHOULSKAIA, D. ZAKHAROV, A. MELAKE-BERHAN, P. S. SPRINGER, K. J. EDWARDS, M. LEE, Z. AVRAMOVA, AND J. L. BENNETZEN. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765–768.

SAVOLAINEN, V., M. W. CHASE, S. B. HOOT, C. M. MORTON, D. E. SOLTIS, C. BAYER, M. F. FAY, A. Y. DE BRUIJN, S. SULLIVAN, AND Y.-L. QIU. 2000. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* sequences. *Systematic Biology* 49: 306–362.

SAX, K. 1931. The origin and relationships of the Pomoideae. *Journal of the Arnold Arboretum* 12: 3–22.

SAX, K. 1932. Chromosome relationships in Pomoideae. *Journal of the Arnold Arboretum* 13: 363–367.

SAX, K. 1933. The origin of the Pomoideae. *Proceedings of the American Society of Horticultural Science* 30: 147–150.

SHOEMAKER, R. C., K. POLZIN, J. LABATE, J. SPECHT, E. D. BRUMMER, T. OLSON, N. YOUNG, V. CONCIBIDO, J. WILCOX, J. P. TAMULONIS, G. KOCHERT, AND H. R. BOERMA. 1996. Genome duplication in soybean (*Glycine* subgenus *Soja*). *Genetics* 144: 329–388.

SIMILLION, C., K. VANDEPOELE, M. C. E. VAN MONTAGU, M. ZABEAU, AND Y. VAN DE PEER. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* 99: 13627–13632.

SOLTIS, D. E., AND P. S. SOLTIS. 1993. Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Science* 12: 243–273.

SOLTIS, P. S., AND D. E. SOLTIS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Science, USA* 97: 7051–7057.

SOLTIS, D. E., AND P. S. SOLTIS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution* 14: 348–352.

SOLTIS, P. S., J. J. DOYLE, AND D. E. SOLTIS. 1992. Molecular data and polyploid evolution in plants. *In* P. S. Doltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants, 177–201. Chapman & Hall, New York, New York, USA.

SOLTIS, D. E., P. S. SOLTIS, M. D. BENNETT, AND I. J. LEITCH. 2003. Evolution of genome size in the angiosperms. *American Journal of Botany* 90: 1596–1603.

SORENG, R. J., AND J. I. DAVIS. 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification versus reticulation. *In* S. W. L. Jacobs and J. E. Everett [eds.], Grasses: systematics and evolution, 61–74. CSIRO, Collingwood, Victoria, Australia.

SPANGLER, R. E. 2003. Taxonomy of *Sarga, Sorghum* and *Vacoparis* (Poaceae: Andropogoneae). *Australian Systematic Botany* 16: 279–299.

SPANGLER, R., B. ZAITCHIK, E. RUSSO, AND E. A. KELLOGG. 1999. Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using *ndhF* sequences. *Systematic Botany* 24: 267–281.

STAM, M., C. BELELE, W. RAMAKRISHNA, J. DORWEILER, J. L. BENNETZEN, AND V. L. CHANDLER. 2002. The regulatory regions required for *B'* paramutation and expression are located far upstream of the maize *b1* transcribed sequences. *Genetics* 162: 917–930.

STEBBINS, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York, New York, USA.

STEBBINS, G. L. 1971. Chromosome evolution in higher plants. Addison-Wesley, Reading, Massachusetts, USA.

SWIGONOVA, Z., J. LAI, J. MA, W. RAMAKRISHNA, V. LLACA, J. L. BENNETZEN, AND J. MESSING. In press. Close split of sorghum and maize genome progenitors. *Genome Research.*

SZYMANSKI, D. B., A. M. LLOYD, AND M. D. MARKS. 2000. Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. *Trends in Plant Science* 5: 214–219.

TANKSLEY, S. D., R. BERNATZKY, N. L. LAPITAN, AND J. P. PRINCE. 1988. Conservation of gene repertoire but not gene order in pepper and tomato. *Proceedings of the National Academy of Science, USA* 85: 6419–6423.

TATE, J. A., AND B. B. SIMPSON. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28: 723–737.

THOMAS, J. W., J. W. TOUCHMAN, R. W. BLAKESLEY, G. G. BOUFFARD, S. M. BECKSTROM-STERNBERG, E. H. MARGULIES, M. BLANCHETTE, A. C. SIEPEL, P. J. THOMAS, J. C. MCDOWELL, B. MASKERI, N. F. HANSEN, M. S. SCHWARTZ, R. J. WEBER, W. J. KENT, D. KAROLCHIK, T. C. BRUEN, R. BEVAN, D. J. CUTLER, S. SCHWARTZ, L. ELNITSKI, J. R. IDOL, A. B. PRASAD, S.-Q. LEE-LIN, V. V. B. MADURO, T. J. SUMMERS, M. E. PORTNOY, N. L. DIETRICH, N. AKHTER, K. AYELE, B. BENJAMIN, K. CARIAGA, C. P. BRINKLEY, S. Y. BROOKS, S. GRANITE, X. GUAN, J. GUPTA, P. HAGHIGHI, S.-L. HO, M. C. HUANG, E. KARLINS, P. L. LARIC, R. LEGASPI, M. J. LIM, Q. L. MADURO, C. A. MASIELLO, S. D. MAS-

TRIAN, J. C. MCCLOSKEY, R. PEARSON, S. STANTRIPOP, E. E. TIONGSON, J. T. TRAN, C. TSURGEON, J. L. VOGT, M. A. WALKER, K. D. WETHERBY, L. S. WIGGINS, A. C. YOUNG, L.-H. ZHANG, K. OSOEGAWA, B. ZHU, B. ZHAO, C. L. SHU, P. J. DE JONG, C. E. LAWRENCE, A. F. SMIT, A. CHAKRAVARTI, C. HAUSSLER, P. GREEN, W. MILLER, AND E. D. GREEN. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424: 788–793.

TIKHONOV, A. P., J. L. BENNETZEN, AND Z. AVRAMOVA. 2000. Structural domains and matrix attachment regions along colinear chromosomal segments of maize and sorghum. *Plant Cell* 12: 249–264.

TORRECILLA, P., AND P. CATALÁN. 2002. Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (Poaceae) based on nuclear ITS sequences. *Systematic Botany* 27: 241–251.

TSCHERMAK-WOESS, E., AND G. HASITSCHKA. 1954. Über die endomitotische Polyploidisierung im Zuge der Differenzierung von Trichomen und Trichozyten bei Angiospermen. *Österreichische Botanische Zeitschrift* 101: 79–117.

U, N. 1935. Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and its peculiar mode of fertilization. *Japanese Journal of Botany* 7: 389–452.

VAN DODEWEERD, A. M., C. R. HALL, E. G. BENT, S. J. JOHNSON, M. W. BEVAN, AND I. BANCROFT. 1999. Identification and analysis of homoeologous segments of the geneomes of rice and *Arabidopsis thaliana*. *Genome* 42: 887–892.

VANDEPOELE, K., C. SIMILLION, AND Y. VAN DE PEER. 2003. Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* 15: 2192–2202.

VAN SLAGEREN, M. W. 1994. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig. *Wageningen Agricultural University Papers* 94–7.

VINOGRADOV, A. E. 2003. Selfish DNA is maladaptive: evidence from the plant Red List. *Trends in Genetics* 19: 609–614.

VISION, T. J., D. G. BROWN, AND S. D. TANKSLEY. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290: 2114–2117.

VITTE, C., AND O. PANAUD. 2003. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Molecular Biology and Evolution* 20: 528–540.

WENDEL, J. F. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.

WENDEL, J. F., AND R. C. CRONN. 2003. Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* 78: 139–186.

WENDEL, J. F., R. C. CRONN, I. ALVAREZ, B. LIU, R. L. SMALL, AND D. S. SENCHINA. 2002. Intron size and genome size in plants. *Molecular Biology and Evolution* 19: 2346–2352.

WENDEL, J. F., C. W. STUBER, M. M. GOODMAN, AND J. B. BECKETT. 1989. Duplicated plastid and triplicated cytosolic isozymes of triosephosphate isomerase in maize (*Zea mays* L.). *Journal of Heredity* 80: 218–228.

WESSLER, S. R., T. E. BUREAU, AND S. E. WHITE. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Current Opinion in Genetics and Development* 5: 814–821.

WICKER, T., N. YAHIAOUI, R. GUYOT, E. SCHLAGENHAUF, Z. LIU, J. DUBCOVSKY, AND B. KELLER. 2003. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. *Plant Cell* 15: 1186–1197.

WILSON, W. A., S. E. HARRINGTON, W. L. WOODMAN, M. LEE, M. E. SORRELLS, AND S. R. MCCOUCH. 1999. Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. *Genetics* 153: 453–473.

WU, X., S. R. LARSON, Z. HU, A. J. PALAZZO, T. A. JONES, R. R.-C. WANG, K. B. JENSEN, AND N. J. CHATTERTON. 2003. Molecular genetic linkage maps for allotetraploid *Leymus* wildryes (Gramineae: Triticeae). *Genome* 46: 627–646.

YU, J., S. N. HU, J. WANG, G. K. S. WONG, AND S. G. LI, ET AL. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79–92.

ZHANG, H., J. JIA, M. D. GALE, AND K. M. DEVOS. 1998. Relationship between the chromosomes of *Aegilops umbellulata* and wheat. *Theoretical and Applied Genetics* 96: 69–75.