American Journal of Botany 96(1): 336–348. 2009.

POLYPLOIDY AND ANGIOSPERM DIVERSIFICATION[1]

DOUGLAS E. SOLTIS,[2,11] VICTOR A. ALBERT,[3] JIM LEEBENS-MACK,[4] CHARLES D. BELL,[5] ANDREW H. PATERSON,[4,6] CHUNFANG ZHENG,[7] DAVID SANKOFF,[8] CLAUDE W. DEPAMPHILIS,[9] P. KERR WALL,[9] AND PAMELA S. SOLTIS[10]

[2]Department of Botany, University of Florida, Gainesville, Florida 32611 USA; [3]University at Buffalo (SUNY), 109 Cooke Hall, Buffalo, New York 14260-1300 USA; [4]Department of Plant Biology, University of Georgia, Athens, Georgia 30602 USA; [5]Department of Biological Sciences, University of New Orleans, New Orleans, Louisiana 70148 USA; [6]Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30602 USA; [7]Department of Biology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5; [8]Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5; [9]Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802 USA; and [10]Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611 USA

Polyploidy has long been recognized as a major force in angiosperm evolution. Recent genomic investigations not only indicate that polyploidy is ubiquitous among angiosperms, but also suggest several ancient genome-doubling events. These include ancient whole genome duplication (WGD) events in basal angiosperm lineages, as well as a proposed paleohexaploid event that may have occurred close to the eudicot divergence. However, there is currently no evidence for WGD in *Amborella*, the putative sister species to other extant angiosperms. The question is no longer "What proportion of angiosperms are polyploid?", but "How many episodes of polyploidy characterize any given lineage?" New algorithms provide promise that ancestral genomes can be reconstructed for deep divergences (e.g., it may be possible to reconstruct the ancestral eudicot or even the ancestral angiosperm genome). Comparisons of diversification rates suggest that genome doubling may have led to a dramatic increase in species richness in several angiosperm lineages, including Poaceae, Solanaceae, Fabaceae, and Brassicaceae. However, additional genomic studies are needed to pinpoint the exact phylogenetic placement of the ancient polyploidy events within these lineages and to determine when novel genes resulting from polyploidy have enabled adaptive radiations.

Key words: angiosperm diversification; genome doubling; polyploidy; whole genome duplication.

As the papers in this issue and other recent reviews illustrate (e.g., Kim et al., 2005; Friis et al., 2006; Frohlich and Chase, 2007; Jansen et al., 2007; Tang et al., 2008; Doyle, 2008), through collective progress in paleobotany, phylogeny, and studies of the evolution of plant development, enormous progress has been made in understanding Darwin's "abominable mystery"—the "rapid rise and early diversification" of the angiosperms. For example, recent phylogenetic analyses suggest that following initial, early "experiments," angiosperms radiated rapidly (≤5 million years), yielding the five extant lineages of Mesangiospermae (sensu Cantino et al., 2007; magnoliids and Chloranthaceae as sisters to a clade of monocots and eudicots + Ceratophyllaceae) (Moore et al., 2007; Saarela at al., 2007). This rapid radiation ultimately produced ca. 97% of all angiosperm species. Here we emphasize that polyploidy, often referred to in the genomics literature as whole-genome duplication (WGD), has played a dramatic role in the diversification of most, if not all, eukaryotic lineages, perhaps most impressively within the angiosperms.

[1] Manuscript received 27 February 2008; revision accepted 2 July 2008.

The authors thank M. Wojciechowski, J. Doyle, and D. Olmstead for helpful information regarding phylogeny and genome duplication within Fabaceae and Solanaceae and A. Chanderbali and S. Clifton for use of unpublished *Amborella* EST data. They also thank C. Pires and two anonymous reviewers for helpful comments on the manuscript. This work was supported in part by the Angiosperm Assembling the Tree of Life grant (NSF EF-0431266), the Deep Time RCN (NSF DEB-0090283), and the Ancestral Angiosperm Genome Project (NSF PGR-0638595).

[11] Author for correspondence (e-mail: dsoltis@botany.ufl.edu)

doi:10.3732/ajb.0800079

Researchers have long recognized that polyploidy is an inseparable part of angiosperm biology. Polyploidy in angiosperms has been studied for a century, dating to the work of De Vries (see Lutz, 1907; Gates, 1909) and to early interest in a putative chromosome duplication in maize (*Zea mays*) (Kuwada, 1911). Early reviews of polyploidy in plants included Müntzing (1936), Darlington (1937), Clausen et al. (1945), Löve and Löve (1949), and Stebbins (1947, 1950, 1971, 1985). Following the work of Stebbins (1940, 1947, 1950) in particular, polyploidy became a major focus of biosystematic research. As a result, plant scientists have long recognized that polyploid lineages may have complex relationships with each other and their diploid ancestors, making application of species concepts problematic (reviewed in Rieseberg and Willis, 2007; D. Soltis et al., 2007).

Fueled in part by evidence for ancient polyploidy in genome sequences of *Arabidopsis* (e.g., Blanc et al., 2000, 2003; Paterson et al., 2000; Vision et al., 2000; Simillion et al., 2002; Bowers et al., 2003), and more recently published angiosperm genome sequences (discussed later), the past decade has seen a dramatic resurgence in the study of polyploidy (e.g., see volume edited by Leitch et al., 2004; also reviewed in D. Soltis et al., 2003; Adams and Wendel, 2005; Tate et al., 2005; Wendel and Doyle, 2005). There has been renewed interest in the mechanisms of polyploid formation and establishment (Ramsey and Schemske, 1998, 2002; Husband, 2004), the frequency of recurrent polyploidization (e.g., D. Soltis and Soltis, 1999; P. Soltis and Soltis, 2000), the ecological effects of plant polyploidy (reviewed in Thompson et al., 1997, 2004), and the genetic, epigenetic, chromosomal, and genomic consequences of polyploidization (e.g., Liu and Wendel, 2003; Bowers et al., 2003; Osborn et al., 2003; Leitch and Bennett, 2004; Adams and

Wendel, 2005; Rapp and Wendel, 2005; Paterson et al., 2006; Leitch and Leitch, 2008; Tang et al., 2008). This research has resulted in major modifications to many of the traditional tenets of polyploid evolution.

Despite numerous attempts over the past 70 years to estimate the frequency of polyploidy in plants, it has proved difficult to determine the actual frequency of the process in various plant lineages. Angiosperms, in particular, have received much attention. Using chromosome numbers and hypotheses for the presumed dividing line between "diploid" and "polyploid" chromosome numbers, many estimates of the frequency of polyploidy in flowering plants have been made. However, these estimates have varied depending on the base chromosome number used, as well as on the sample of taxa considered. Müntzing (1936) and Darlington (1937) speculated that about one half of all angiosperm species were polyploid, while Stebbins (1950) estimated the frequency of polyploidy to be 30–35% in angiosperms. Grant (1963, 1981), basing his estimate of the frequency of polyploidy in angiosperms on chromosome numbers for 17 138 species available in 1955, hypothesized that flowering plants with haploid chromosome numbers of $n = 14$ or higher were of polyploid origin. Using this cut-off point, Grant inferred that 47% of all flowering plants were of polyploid origin and proposed that 58% of monocots and 43% of "dicots" were polyploid. Goldblatt (1980) suggested that Grant's (1963) estimate was too conservative; he thought that taxa with chromosome numbers above $n = 9$ or 10 would have had polyploidy in their evolutionary history. Using these lower numbers, he calculated that at least 70%, and perhaps 80%, of monocots are of polyploid origin. Lewis (1980) applied a similar approach to "dicots" and estimated that 70–80% were polyploid. However, one problem with these sorts of approaches is that they do not accommodate reductions in chromosome number that occurred rapidly on an evolutionary scale. For example, species of *Sorghum* with $n = 5$ are recently derived from taxa with $n = 10$ (Spangler et al., 1999), and *Arabidopsis thaliana* with $n = 5$ appears to have resulted from three chromosomal condensations since its divergence from its common ancestor with *A. lyrata* and the more distantly related *Capsella rubella* (Kuittinen et al., 2004; Koch and Kiefer, 2005; Yogeeswaran et al., 2005).

Masterson (1994) compared leaf guard cell size in fossil and extant taxa from a few angiosperm families (Platanaceae, Lauraceae, Magnoliaceae) to estimate polyploid occurrence through time. Because guard cell size is often much larger in polyploids than in diploids, this provided a gross estimate of whether the fossil taxa were diploid (smaller guard cells than extant taxa) or polyploid (the same or larger guard cell sizes vs. extant species). From these comparisons, Masterson (1994) estimated that 70% of all angiosperms had experienced one or more episodes of polyploidy in their ancestry. Using an innovative approach for estimating the incidence of polyploidy based on the distribution of haploid chromosome numbers, Otto and Whitton (2000, p. 427) suggested that roughly 2–4% of speciation events in angiosperms may have involved polyploidy. As a result, they suggested that "polyploidization may be the single most common mechanism of sympatric speciation in plants."

The older estimates of the frequency of polyploidy discussed tend to confound the actual process of polyploidization with the result (i.e., polyploid lineages). Actual WGD events should be far less frequent than particular chromosome numbers suggestive of polyploidy in any given angiosperm species. As reviewed next, through the interplay of genomic and phylogenetic approaches, we are on the verge of determining the frequency of ancient polyploidy events throughout angiosperm history.

A GENOMICS APPROACH TO POLYPLOIDY

Summary of genomes sequenced to date—Investigations of completely sequenced nuclear genomes have dramatically altered the polyploidy paradigm. They have revealed that flowering plants, and perhaps all eukaryotes, possess genomes with considerable gene redundancy, much of which is the result of (ancient) WGDs. Those entire angiosperm genomes that have been completely sequenced to date—*Oryza sativa* (rice, Poaceae; Paterson et al., 2004), *Arabidopsis thaliana* (Brassicaceae; Blanc et al., 2000; Paterson et al., 2000; Vision et al., 2000; Simillion et al., 2002; Bowers et al., 2003), *Populus trichocarpa* (poplar, Salicaceae; Tuskan et al., 2006), *Vitis vinifera* (grape, Vitaceae; Jaillon et al., 2007; Velasco et al., 2007), and most recently *Carica papaya* (papaya, Caricaceae; Ming et al., 2008)—all show evidence of WGD events.

Complete sequencing of the very small genome (for angiosperms, 157 Mb; Bennett et al., 2003) of *Arabidopsis thaliana* (Brassicacae; *Arabidopsis* Genome Initiative, 2000; IRGSP, 2005) revealed numerous duplicate genes and suggested two or three rounds of genome-wide duplication (Paterson et al., 2000; Blanc et al., 2000; *Arabidopsis* Genome Initiative, 2000; Vision et al., 2000; Bowers et al., 2003; Simillion et al., 2002; reviewed in Van de Peer and Meyer, 2005), corroborating early suspicions based on genetic mapping (McGrath et al., 1993; Kowalski et al., 1994). Analyses of the complete genome sequence of *A. thaliana* suggested three ancient polyploidy events (Blanc et al., 2003; Bowers et al., 2003). Bowers et al. (2003) proposed that one of these events (termed α) may have occurred within Brassicales and suggested the other two (termed β and γ) could be considerably older (but this view has now been revised; see Fig. 1; Sampedro et al., 2005; Jaillon et al., 2007; and discussed later).

Sequencing of the nuclear genome of rice (*Oryza sativa*; Poaceae) again suggested ancient polyploidy in a plant with a chromosome number ($n = 12$) considered by many to be a model "diploid" (Paterson et al., 2004;Yu et al., 2005), although secondary associations (loose pairing at meiosis) among homologous and nonhomologous chromosomes have long been known (Lawrence, 1931). The exact phylogenetic placement of rice's WGD event is unclear; Paterson et al. (2004) estimated that the WGD occurred after the divergence of Poales from other monocot orders but before the divergence of the major cereals from one another (see later). However, there is also equivocal evidence of a more ancient genome-doubling event in rice that might characterize all monocots (Velasco et al., 2007). The possibility of an ancient genome doubling in monocots is gaining support from analysis of recently sequenced genomes and new analytical methods (H. Tang, J. Bowers, X. Wang, University of Georgia; A. Paterson, unpublished data).

Sequencing of the *Populus* (Salicaceae) genome also revealed an independent WGD event after the divergence of the eurosid I and eurosid II lineages (Fabidae and Malvidae, Cantino et al., 2007), but before the divergence of *Salix* and *Populus* (Tuskan et al., 2006). Ancient polyploidy in Salicaceae (both *Populus* and *Salix*) was suggested by Stebbins (1950) and has been supported by isozyme data (D. Soltis and Soltis, 1990). It remains to be determined if this genome doubling event will be evident in other lineages of the now expanded Salicaceae

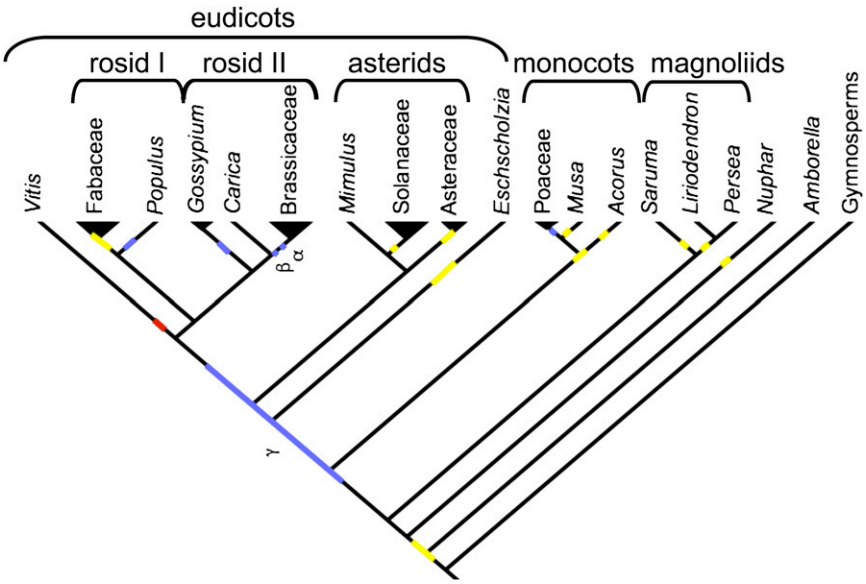

Fig. 1. Simplified summary tree for angiosperms (following the general topology of D. Soltis et al. [2000], with modifications reflecting more recent analyses, including Jansen et al. [2007]; Moore et al. [2007]), depicting putative locations of genome duplication events now inferred for flowering plants relative to major lineages or species with sequenced nuclear genomes or substantial expressed sequence tag data. Tree redrawn to show *Vitis* sister to all other rosids (D. Soltis et al., 2005), and *Carica* is placed as sister to Brassicaceae; *Carica* and Brassicaceae represent Brassicales. The α, β, and γ duplication events of Bowers et al. (2003) have been added to this topology to match what appears to be the best location based on other analyses (see Tang et al., 2008). The timing and nature of the γ event are controversial with estimates of a tetraploid (Velasco et al., 2007) or hexaploid (Jaillon et al., 2007) progenitor ranging from before the divergence of the monocot and eudicot lineages to the last common ancestor of all extant rosids. The α event likely occurred within Brassicaceae. The position of β is less certain; it may also have occurred within Brassicaceae, or perhaps earlier in Brassicales (see Fig. 5). Blue bars represent genome duplications inferred from comparative analyses of physical and/or genetic maps. The red bar on the *Vitis* lineage represents the Velasco et al. (2007) inference of a second distinct tetraploidy evident in the *Vitis* genome (but see Jaillon et al., 2007 for a different interpretation). Yellow bars represent genome-wide duplication events inferred from analyses of paralog pairs found in large EST sets (e.g., Blanc and Wolfe, 2004; Cui et al., 2006).

(APG II, 2003). A second, older duplication is shared by the *Populus* and *Arabidopsis* lineages (Tuskan et al., 2006).

Based on analyses of the complete genome sequence of *Vitis*, Jaillon et al. (2007) suggested that the common ancestor of *Vitis*, *Populus*, and *Arabidopsis* was an ancient hexaploid (this is now considered the γ event) that possibly arose after the divergence of the monocot and eudicot lineages (see also Sampedro et al., 2005). The exact placement of the γ event remains unclear (hence the ambiguity in its placement in Fig. 1); γ may have occurred before the monocot–eudicot divergence. However, a paleohexaploid signature is not apparent in rice, favoring a placement after the divergence of monocots and eudicots. Following the formation of this hypothetical paleohexaploid, there were subsequent, distinct, genome-wide duplication events in the Brassicales and *Populus* lineages (Fig. 1). Velasco et al. (2007) proposed, in contrast to Jaillon et al. (2007), that the second WGD event evident in *Vitis* was more recent (Fig. 1). Velasco et al. proposed that three genome-wide duplications occurred in the ancestors of both *Arabidopsis* and *Populus*: one shared by all eudicots (and perhaps also monocots), a second WGD shared by *Arabidopsis* and *Populus* (cf. Fig. 1), and additional, single WGDs unique to each lineage. Velasco et al. (2007) proposed that *Vitis* has the ancient genome-wide duplication shared by all eudicots, as well as a lineage-specific duplication that may be the result of hybridization. The most ancient of these events is now thought to correspond to the γ event of Bowers et al. (2003).

The nuclear genome of the eudicot *Carica papaya* (Caricaceae, Brassicales) has recently been sequenced (Ming et al., 2008). *Carica* arose from a deep split in the order and is distantly related to Brassicaceae, the family that contains both *Arabidopsis* and *Brassica* (discussed later). There are no recent genome duplications evident in *Carica* and nothing corresponding to the *Arabidopsis* α or β events proposed by Bowers et al. (2003). Because Caricaceae are estimated to have diverged from the Brassicaceae lineage ca. 72 million years ago (Ming et al., 2008), the absence of recent genome duplication is in conflict with prior estimates of a much earlier age for the β genome duplication, which Bowers et al. (2003) suggested could correspond to the origin of the eudicots. This apparent inconsistency is under investigation (J. Bowers, University of Georgia; A. Paterson, unpublished data). However, *Carica* shows evidence of γ, the early event detected in *Vitis*, *Populus*, and *Arabidopsis* that is now interpreted by Jaillon et al. (2007) as a possible ancient triplication event. *Carica* has a reduced gene number, about 10% fewer genes than *Arabidopsis*, which may be largely accounted for by a paucity of genome duplications relative to other sequenced angiosperms. The data now suggest that both the α and β events occurred within Brassicales and are younger than originally proposed by Bowers et al. (2003). The α event likely occurred somewhere within Brassicaceae. However, the position of β is less clear; it may have occurred within Brassicales at some point after the divergence of Caricaceae (Fig. 1; more detail on these duplications is given later).

Although it is now clear that genome-wide duplications have occurred frequently in angiosperms (Fig. 1), the exact number and precise phylogenetic placement of most of these events remain uncertain. Clarifying the timing and placement of these duplication events, particularly the older ones, will require additional genome sequencing, particularly of one or more basal angiosperms, providing the opportunity for ancestral-state reconstruction of numerous genomic characters (e.g., Pryer et al., 2002; Jackson et al., 2006). Nevertheless, even as the number of angiosperm genome sequences increases, the oldest genome duplications will remain difficult to detect because recurrent polyploidy events, together with gene loss, chromosomal inversions, and translocations following genome duplication may obscure evidence of the earliest events in angiosperm history.

Because phylogenetic analyses (e.g., P. Soltis et al., 1999; reviewed in D. Soltis et al., 2005, Leebens-Mack et al., 2005; Jansen et al., 2007; Moore et al., 2007) identify *Amborella trichopoda* (Amborellaceae) as the probable single sister species to all other living flowering plants, complete sequencing of the *Amborella* nuclear genome offers the opportunity to clarify how gene families and genome structure may have evolved within angiosperms (Leebens-Mack et al., 2006; D. Soltis et al., 2008a). In addition, because the branching point for *Amborella* is located "between" gymnosperms and all other angiosperms, a genome sequence for *Amborella* could potentially help characterize genomic features that distinguish angiosperms from other seed plants. An *Amborella* nuclear genome sequence would facilitate efforts to reconstruct characteristics of the "ancestral angiosperm" (Leebens-Mack et al., 2006; D. Soltis et al., 2008a)—the most recent common ancestor of the crown clade.

EST data and lineage-specific duplications—In addition to complete sequencing of the nuclear genome, other major sources of genomic data are the numerous expressed sequence tags (ESTs) that are now available for many flowering plants. A recent survey of GenBank revealed over 50 million ESTs, representing diverse eukaryotic lineages. The thousands of ESTs available for many plants provide a useful "snapshot" of each genome. The rapidly growing EST data sets for diverse angiosperms can be employed to assess ancient polyploidy using a genomics method to estimate whether a genome duplication event may have occurred in a lineage, as well as the approximate age of these gene duplication events (Lynch and Conery, 2000). This method evaluates the frequency distribution of per-site synonymous divergence levels (K_s) for pairs of duplicate genes. A genome-wide duplication event results in thousands of paralogous pairs—all simultaneously duplicated. Evidence of past genome duplications can be seen as peaks in the distribution of K_s values for sampled paralogous pairs (Lynch and Conery, 2000; Blanc and Wolfe, 2004; Schlueter et al., 2004; Cui et al., 2006). Importantly, this method does not require information on the position of genes within the genome, and therefore it can be applied to any species for which there are moderate to large EST sets. However, caution must be used in interpreting K_s values, and it is important to use statistical testing to identify significant deviations from background duplication process (Schlueter et al., 2004; Cui et al., 2006). There are clear examples in which well-known genome duplication events were not detected in K_s distributions (Blanc and Wolfe, 2004; Paterson et al., 2004). As duplicated genes undergo divergence and death processes, it becomes increasingly likely that polyploidy events will go undetected (Cui et al., 2006). Finally, processes such as a sudden proliferation of tandem gene duplications (Blanc and Wolfe, 2004) or a single chromosome duplication could produce a K_s peak through a large-scale process distinct from a true polyploid event.

We summarize the many genomes thought to be ancient polyploids using analyses of complete genome sequences, as well as ESTs, on a modified summary tree for angiosperms (Fig. 1). Genetic investigation of other taxa using other methods suggests additional ancient polyploidy events. For example, "diploid" members of *Brassica* are, at the least, ancient tetraploids (Kowalski et al., 1994; Lan et al., 2000; Quiros et al., 2001; Fig. 1) and perhaps ancient hexaploids based on analyses of linkage maps with a number of genes clearly represented multiple times (e.g., Lagercrantz and Lydiate, 1996; Lukens et al., 2004). There are also other lineage-specific duplications (see later).

Blanc and Wolfe (2004) investigated 14 model plant species (mostly crop species with known recent polyploid history) for which large EST data sets were available and found spikes in the distributions of older paralogous pairs (with higher K_s values) in nine species, including *Zea* (maize), *Glycine* (soybean), *Gossypium* (cotton), and *Solanum* (tomato and potato). Schlueter et al. (2004) similarly employed the analysis of K_s distributions to sets of paralogous pairs identified in large EST data sets for eight major crop species, including *Glycine*, *Medicago* (alfalfa), *Solanum*, *Zea*, *Sorghum*, *Oryza*, and *Hordeum* (barley), and inferred multiple independent genome duplications in Fabaceae, Solanaceae, and Poaceae over the last 14–60 million years.

When the K_s approach was applied to ESTs from a suite of basal angiosperms via the Floral Genome Project (see Albert et al., 2005), evidence was found for episodes of ancient genome-wide duplication in *Nuphar advena* (Nymphaeaceae), *Persea americana* (avocado; Lauraceae), *Liriodendron tulipifera* (yellow poplar or tulip tree; Magnoliaceae), and *Saruma henryi* (Aristolochiaceae) (Cui et al., 2006). In addition, Cui et al. (2006) detected independent genome duplications in the basal eudicot *Eschscholzia californica* (California poppy; Papaveraceae) and the basal monocot *Acorus americanus* (sweet flag: Acoraceae), both of which were distinct from duplications documented for core eudicots and Poaceae (Fig. 1).

More than one genome-wide duplication event is evident in *Nuphar* (Nymphaeaceae). One of these is likely restricted to Nymphaeaceae (Nymphaeales), but another may correspond to the oldest duplication so far discovered in angiosperms—the latter may date to the common ancestor of all angiosperms except *Amborella*, which so far lacks evidence of ancient polyploidy (Cui et al., 2006; see later) (Fig. 1). Analysis of K_s values also provided evidence for genome-wide duplication in both *Persea* (Lauraceae) and *Liriodendron* (Magnoliaceae). One of these is shared by both families, corroborating evidence based on isozyme data (D. Soltis and Soltis, 1990) for an ancient polyploidy event (about 100 Mya) in their common ancestor. There is weak although inconclusive evidence of a still older WGD in *Persea*, which may correspond to the event suggested for the common ancestor of all angiosperms except *Amborella*. Alternatively, this WGD could perhaps even predate the angiosperms (see later). Testing these and similar hypotheses will minimally require comprehensive transcriptome sequencing for additional basal angiosperms and a complete *Amborella* genome sequence (D. Soltis et al., 2008a).

Are all angiosperms of ancient polyploid origin?—All angiosperm nuclear genomes studied except that of *Amborella* have yielded evidence for multiple genome-wide duplications (Fig. 1). In contrast, other seed plant lineages (with the exception

of Gnetales) have little evidence of polyploidy (Grant, 1981; D. Soltis et al., 2005). As such, the major question is no longer "How many angiosperms are polyploid?", but rather "How many episodes of genome duplication have various lineages experienced?" and "Did specific genes resulting from genome duplication serve to fuel adaptive radiations?"

Despite its relatively high chromosome number of $2n = 26$, very few duplicate gene pairs were detected in *Amborella* based on the initial 8629 ESTs analyzed by Cui et al. (2006), and no K_s signal for ancient genome duplication was observed. The Ancestral Angiosperm Genome Project (AAGP) has now produced more than 20 800 conventional Sanger ESTs and 800 000 ESTs generated via 454 sequencing technology (http://www.454.com/enabling-technology/index.asp) for *Amborella*. This K_s analysis identified 279 paralogous gene pairs with sequence length greater than 300 bp: more than four times the 69 pairs identified by Cui et al. (2006). Even with this greatly expanded EST data set, there is no evidence for an early genome duplication in *Amborella* (Fig. 2; AAGP, unpublished data). However, the lack of evidence for ancient polyploidy in *Amborella* K_s distributions does not preclude the possibility of ancient polyploidy in either this lineage or in the genome of an early angiosperm or angiosperm precursor. The absence of a polyploidy K_s signal in *Amborella* could also occur if evidence of ancient WGD has been eroded by gene death and/or saturation of synonymous substitutions (see Cui et al., 2006) to the point where it is now undetectable in analyses of EST samples. Alternatively, the earliest duplication peak detected in the *Nuphar* analysis (Fig. 2; discussed before) may trace back to an early genome duplication that occurred in the common ancestor of *Nuphar* and all extant angiosperm lineages *other than Amborella.* Complete sequencing of the *Amborella* nuclear genome is needed to test these alternative hypotheses (D. Soltis et al., 2008a).

Duplication of several MADS-box genes, which control floral organ identification and development, also suggests genome doubling early in angiosperm history or perhaps prior to crown angiosperm origins (Kim et al., 2004; Kramer et al., 2004; Buzgo et al., 2005; Zahn et al., 2005; reviewed in Irish, 2006; Kramer and Zimmer, 2006; P. Soltis et al., 2006; D. Soltis et al., 2007). For example, gymnosperms have only one B-function lineage, whereas all angiosperms have at least two such lineages (homologs of *AP3* and *PI*). The two B-function gene lineages, which include homologs of *AP3* and *PI,* respectively, appear to have originated via duplication of a single B-function gene at some point prior to the origin of the angiosperms (Kim et al., 2004). An ancient duplication event of comparable timing occurred in the C-function lineage, forming two lineages in angiosperms, one with *AG* homologs (with roles in stamen and carpel identity), and the other with D function (with a role in ovule formation) (Kramer et al., 2004). Similarly, duplication of *SEP* genes resulted in the *AGL2/3/4* (*SEP1/2/4*) and *AGL9* (*SEP3*) lineages in the common ancestor of the angiosperms (Zahn et al., 2005). WGD is also favored by analysis of protein–protein interaction networks among MADS domain proteins (Veron et al., 2007). Interestingly, the latter study suggests that heterodimerizing factors appear to derive from duplication of homodimerizing ancestors. Duplications of these key floral organ identity genes may have been important in the origin of the flower, but the exact timing of these gene duplications remains unclear. Given that some of these MADS-box gene duplications are thought to be quite old (over 260 million years, well before the origin of the crown angiosperms; see Kim et al., 2004) and others more recent, were there several genome-wide duplication events along the stem lineage leading to the crown angiosperms? Similar questions also apply to early events in angiosperm diversification. Phylogenetic studies have revealed duplication of MADS-box genes *within* the basal eudicot clades; these genes include the *AG, AP3, AP1*, and *SEP* gene lineages (Kramer et al., 1998, 2004; Litt and Irish, 2003; Zahn et al., 2005; reviewed in Kramer and Zimmer, 2006; Irish, 2006). Could these correspond to a genome-wide duplication event early in eudicot evolution (e.g., the γ event; Fig. 1)? MADS protein–protein

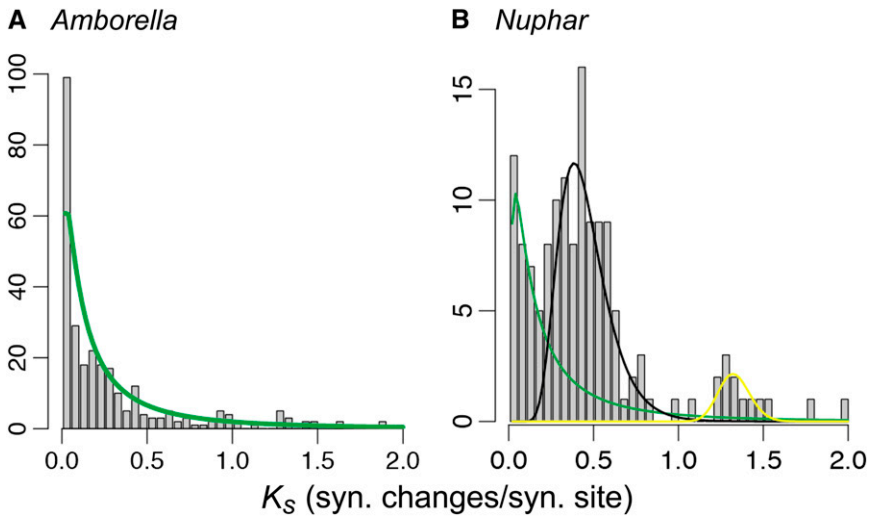

Fig. 2. A greatly expanded *Amborella* expressed sequence tag (EST) data set provides much greater power for detecting ancient genome duplication events. (A) K_s analysis of *Amborella trichopoda* based on 28 000 conventional ESTs (Sanger), and approximately 800 000 454 ESTs; this facilitated the assembly of 701 gene pairs. Although this expanded *Amborella* EST data set provides much greater power to detect ancient polyploidy than was available for the estimation in Cui et al. (2006), there is still no evidence for genome duplication based on this K_s analysis. (B) K_s analysis of 134 gene pairs identified from 8442 *Nuphar* ESTs (adapted from Cui et al. [2006]) shows three components to the K_s plot, including background duplication (green), and two significant older (one in black and one in yellow, Fig. 2) components representing likely ancient polyploidy events.

interaction data imply successive duplications from angiosperm origins through the rosid–asterid split (Veron et al., 2007). Forthcoming asterid genome sequences for *Mimulus* and one or more species of Solanaceae may help resolve this question.

Reconstructing ancestral genomes prior to genome duplication—Now that WGD is known to be both frequent and ubiquitous across angiosperm history, a major challenge is the reconstruction of the ancestral genomes of lineages prior to genome duplication. Reconstruction of the ancestral genomes of lineages prior to genome duplication improves our ability to resolve correlated gene arrangements among taxa (Bowers et al., 2003), which in turn provide a valuable framework for inference of shared ancestry of genes and for the utilization of findings from model organisms to study less well-understood systems. Such alignments have also revealed patterns of differential gene loss following genome duplication, differential gene retention associated with evolution of some morphological complexity, and unexpectedly large variation among taxa in DNA substitution rates (Tang et al., 2008).

Such reconstructions are challenging, however, because whole-genome duplication may be followed, over evolutionary time, by genome downsizing (paralog extinction and the loss of noncoding DNA), rearrangement through intra- and interchromosomal movement of genetic material, and continual gene-duplication events that do not involve WGD. However, reconstruction of ancestral genomes can potentially be achieved through comparative analysis of genome content and structure for extant species, combining information from gene duplication histories (developed through analysis of sequence similarities or gene trees) and chromosomal gene locations to identify duplicated blocks of genes or DNA sequence dispersed among the chromosomes in one or more species. Sequence analysis tools are not directly applicable to this problem because all pairs of duplicates in doubled genomes were generated at the same historical moment. A rapid (linear-time) algorithm to find the ancestral genome that minimizes the genomic distance (minimum number of inversions and translocations) to a present-day genome has been available for some time (El-Mabrouk and Sankoff, 2003). Unfortunately, there are numerous rather different solutions that are equally parsimonious. A new procedure—guided genome halving—seeks to counteract this problem by guiding the reconstruction using one or more reference, or outgroup, genomes (Zheng et al., 2006, 2008a–c; Sankoff et al., 2007). The principle behind the guided genome halving approach is that each time a choice between two equally good construction steps is encountered by the halving algorithm, the outgroup information comes into play to see if one of the choices corresponds more to the structure of the reference genome(s).

Here we use the complete nuclear genome sequences of *Populus* (Salicaceae) and *Vitis* (Vitaceae) to attempt to reconstruct the ancestral genome of *Populus* prior to genome duplication (Fig. 3) as an example of the process of reconstructing ancestral genomes. Paralogous gene pairs in *Populus* and their single orthologs in *Vitis*, where these exist, were used as input. Both *Vitis* and *Populus* are rosids, with Vitaceae sister to all other rosids and Salicaceae a member of Malpighiales (in the eurosid I or Fabidae clade); APG II, 2003; Cantino et al., 2007). We scaled up the guided genome-halving algorithm to handle the very large number of genes analyzable in the *Populus* and *Vitis* data, introducing all pertinent information derivable from the outgroup (*Vitis*) into the halving algorithm as applied to *Populus*. The comparison reveals blocks of genes that have remained unaltered over the perhaps 109–124 million years since the two lineages shared a common ancestor (Fig. 3) (Wikström et al., 2001; D. Soltis et al., 2008b). A recent putative WGD event in *Vitis* (per Velasco et al., 2007) might be problematic in these comparisons, but there was no evidence of this in the analyses we conducted.

Many large, duplicated gene blocks observed in two large segments of the *Populus* genome match to one segment in the *Vitis* genome, attesting to the high degree of conserved synteny (gene order within blocks of sequence) and colinearity (ordering of the syntenic blocks) during divergence of the *Vitis* and *Populus* lineages, as well as retention of duplicate blocks during the diploidization of the ancestral *Populus* tetraploid (Fig. 3). Nevertheless, details of some of the reconstructed chromosomes cannot be assigned a high degree of confidence. For example, much of chromosome 7 consists of short syntenic blocks concatenated by the program based on a "minimum inversions plus translocations" criterion. This juxtaposes genes that are in fact nowhere adjacent in the *Vitis* and *Populus* data. Improvements to existing algorithms, now underway, will exploit new evidence from additional related genomes.

HAVE ANCIENT POLYPLOIDY EVENTS RESULTED IN INCREASED SPECIES RICHNESS?

Given the many purported benefits of polyploidy (e.g., Levin, 1983, 2000, 2002) and the suggested relationship of genome duplication to speciation (Werth and Windham, 1991; Lynch and Force, 2000; Lynch and Conery, 2000), rates of diversification may be higher in polyploid lineages than in diploid groups (due either to increased rates of speciation, decreased rates of extinction, or both). The identification of ancient WGD events at many points in angiosperm phylogeny provides the opportunity to assess the correspondence between inferred genome duplication events and large diversifications, and hence the role of polyploidy in "macro-diversification."

Anecdotal data suggest that polyploid lineages are "successful," but a statistical association of polyploidy and species richness has not been rigorously tested. To address this question, we compared species richness in clades that are ancient polyploids with sister clades that are not. The overall diversification rate (r) for angiosperms was estimated based on the methods described in Magallón and Sanderson (2001). Because the estimation of this parameter is contingent on the rate of extinction (ε), which is an unknown, we estimated r across a range of extinction rates (Alfaro et al., 2007). We also calculated values of r over a range of plausible age estimates for crown group angiosperms (132 Mya, based on first unambiguous occurrence in the fossil record [Hughes, 1994)]; 170 Mya based on recent molecular age estimations [Wikström et al. 2001; Moore et al., 2007]; 250 Mya based on other estimates based on molecular sequence data [Sanderson and Doyle, 2001; Magallón and Sanderson, 2005]). Next we calculated the probability of observing the extant number of species in several putative clades of polyploid origin given these estimated global rates for angiosperms, conditioned on the assumed age of the crown group (e.g., $r_{G\text{-}132}$, see Table 1). All calculations were done using GEIGER 1.0–91 (Harmon et al., 2008).

Current standing diversity of each polyploid clade was tabulated based on values taken from Stevens (2001), as well as Olmstead et al. (in press), and Wojciechowski et al. (2004).

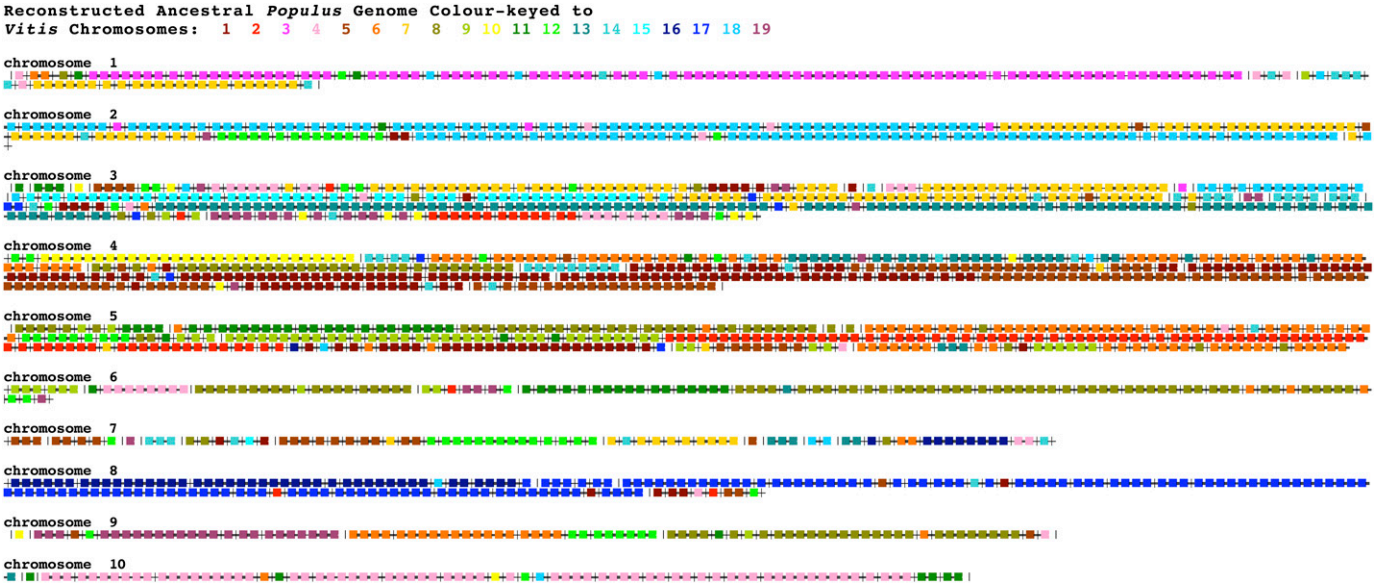

Fig. 3. Reconstruction of the preduplication ancestor of *Populus*. The data contain 2940 genes with two copies in *Populus* and at most one copy in *Vitis*. Of these, 211 "singleton" genes were excluded from the analysis: no two of copies of a singleton have an immediately neighboring gene, an "adjacency," in common. The reconstruction algorithm finds the genome that minimizes the genomic distance, i.e., the number of rearrangements (inversions and translocation), from the ancestral tetraploid to the modern *Populus* genome, while simultaneously minimizing the distance from the diploid version of the ancestor to *Vitis*. The 10 reconstructed chromosomes, containing only the 2020 genes on *Vitis* chromosomes, are shown in wrapped form. Added after the reconstruction: adjacencies present in the original data three times, i.e., twice in *Populus* and once in *Vitis*, are indicated by triple dashes (≡), those present twice by double dashes (=), those present once by single dashes (–), and those for which there is only indirect evidence from the rearrangement analysis by a blank space. Intrachromosomal breakpoints in comparison with one of the *Populus* copies indicated by vertical.

Absolute ages of groups were taken from Wikström et al. (2001). We selected these ages because they represent estimates from the most comprehensive taxon sampling to date. However, taxon sampling by Wikström et al. was not always dense enough across clades of interest (Figs. 4–7) to pinpoint potential duplication events. In these instances, we used the age of the next most inclusive node that would include the inferred duplication event (e.g., age of Fabaceae, when the duplication

TABLE 1. Significant departure in diversification rates of the major plant clades from the global diversification rate estimates. Boldfaced *P* values indicate significantly higher species diversity than expected under the global rate ($r_{G\text{-}132}$) of angiosperms based on a minimum age of 132 Myr, a global rate ($r_{G\text{-}170}$) based on a mean age of 170 Myr (Wikström et al., 2001; Moore et al., 2007), a global rate ($r_{G\text{-}250}$) based on a maximum age of 250 Myr (Sanderson and Doyle, 2001; Magallón and Sanderson, 2005), and a current standing diversity of 261 750 species[a]. ε is the extinction rate.

Clade	No. extant species	Age[b]	ε = 0.0, $r_{G\text{-}132}$ = 0.03690	ε = 0.3, $r_{G\text{-}132}$ = 0.03619	ε = 0.5, $r_{G\text{-}132}$ = 0.03475	ε = 0.7, $r_{G\text{-}132}$ = 0.03186	ε = 0.9, $r_{G\text{-}132}$ = 0.02457
Brassicaceae	3710	22	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Cleomaceae	300	22	**<<0.001**	**<<0.001**	**0.002732658**	**0.01243294**	0.0866002
Within Fabaceae–	7000	63	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Core Poaceae	10 000	15	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Solanoideae	1925	39	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
			ε = 0.0, $r_{G\text{-}170}$ = 0.02866	ε = 0.3, $r_{G\text{-}170}$ = 0.02811	ε = 0.5, $r_{G\text{-}170}$ = 0.02698	ε = 0.7, $r_{G\text{-}170}$ = 0.02475	ε = 0.9, $r_{G\text{-}170}$ = 0.01908
Brassicaceae	3710	22	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Cleomaceae	300	22	**<<0.001**	**<<0.001**	**<<0.001**	**0.002709716**	**0.03925442**
Within Fabaceae–	7000	63	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Core Poaceae	10 000	15	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Solanoideae	1925	39	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
			ε = 0.0, $r_{G\text{-}250}$ = 0.01948	ε = 0.3, $r_{G\text{-}250}$ = 0.01911	ε = 0.5, $r_{G\text{-}250}$ = 0.01835	ε = 0.7, $r_{G\text{-}250}$ = 0.01683	ε = 0.9, $r_{G\text{-}250}$ = 0.01297
Brassicaceae	3710	22	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Cleomaceae	300	22	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**0.008363369**
Within Fabaceae–	7000	63	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Core Poaceae	10 000	15	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**
Solanoideae	1925	39	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**

[a] Based, in part, on Stevens (2001, http://www.mobot.org/MOBOT/Research/APweb/welcome.html)
[b] Based on estimates of Wikström et al. (2001)

event occurred after the origin of the crown group). By doing this, we provided a conservative estimate of diversification rates for each of the clades.

Successfully conducting this exercise is difficult in that it requires both reliable genomic data for ancient polyploidy, plus a sufficient breadth of data to be able to place the WGDs with confidence on a phylogenetic tree. Unfortunately, for most putative examples here, the exact placement of the genome duplication events proposed is not yet clear. Although the data and analyses presented certainly suggest that ancient polyploidy has led to a dramatic increase in species richness in several clades, until additional genomic studies are conducted to pinpoint the placement of those WGD events, the results should be considered preliminary. Nonetheless, in all but a few cases (calculations involving Cleomaceae), *p*-values were found to be highly significant (<<0.001; Table 1), suggesting exceptionally high diversification rates in these polyploid clades.

Poaceae—The estimated age of a genome duplication event characterizing Poaceae (70–50 million years before present: Blanc and Wolfe, 2004; Paterson et al., 2004; Schlueter et al., 2004; Malcomber and Kellogg, 2005) is not only close to the molecular estimates for the age of the family (ca. 89 and 83 million years for stem and crown group Poaceae, respectively; Janssen and Bremer, 2004; see also Bremer, 2002), but also similar to the age of grass fossils (55 million years; Crepet and Feldman, 1991) assignable to the PACCMAD (Panicoideae, Arundinoideae, Centothecoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae) and BEP (Bambusoideae, Ehrhartoideae, Pooideae) clades (Grass Phylogeny Working Group, 2001). The precise phylogenetic placement of the genome duplication is still unclear (as represented by angled line; Fig. 4), but if this event coincided with, or was close to the origin of core Poaceae (indicated by the slash mark on the tree), it would agree with the hypothesis that polyploidy promotes speciation. Core Poaceae are a very large group (658 genera and 9998 species), whereas the three early-branching subclades of Poaceae (Anomochlooideae, Pharoideae, Puellioideae) are all very small, as are the three families most closely related to Poaceae—Flagellariaceae, Joinvilleaceae, and Ecdeicoleaceae (Marchant and Briggs, 2007) (Fig. 4).

Brassicaceae and Cleomaceae—Recent data indicate that a genome-wide triplication occurred in Cleomaceae independently of WGD events in Brassicaceae (Schranz and Mitchell-Olds, 2006). Strengthening support for the occurrence of these doubling events well within Brassicales is the recent genome sequence for *Carica* of Brassicales, which shows no evidence of genome duplications more recent than γ (discussed earlier; Ming et al., 2008). Hence, the data suggest independent WGDs occurred in both Cleomaceae and Brassicaceae (Fig. 5). Virtually all families of Brassicales are very small—most have fewer than 15 species. In contrast, Capparaceae (480 species), Cleomaceae (300 species), and particularly Brassicaceae (3710 species) are all species-rich. Genome-scale data are not yet available for Capparaceae, but evidence for genome duplication in Cleomaceae and Brassicaceae is suggestive of an association between polyploidy and diversification.

Fabaceae—A genome-wide duplication event within the legumes has been documented and is postulated to have occurred immediately prior to the radiation that yielded approximately 7000 species (Lavin et al., 2005; Pfeil et al., 2005; Cannon et al., 2006). However, the placement of this duplication event is much more problematic than those reviewed here. The short branches along the backbone of the tree for the Fabaceae, particularly in the papilionoids, makes placing this duplication and determining the sister group of the clade having the duplication difficult. Because the radiation of the family was so rapid, the uncertainty of the timing of the polyploid event puts just about every node within reach of the event (J. Doyle, Cornell University, personal communication). Considering the placement of the WGD in the common ancestor of *Hologalegina* and phaseoloids (Pfeil et al., 2005; J. Doyle, personal communication), the baphioids, a clade of only five genera and fewer than 20 species, are the immediate sister to the remaining papilionoids, a clade of 7000 species. If this placement is correct, it would again favor the hypothesis that genome duplication may be associated with diversification (Fig. 6).

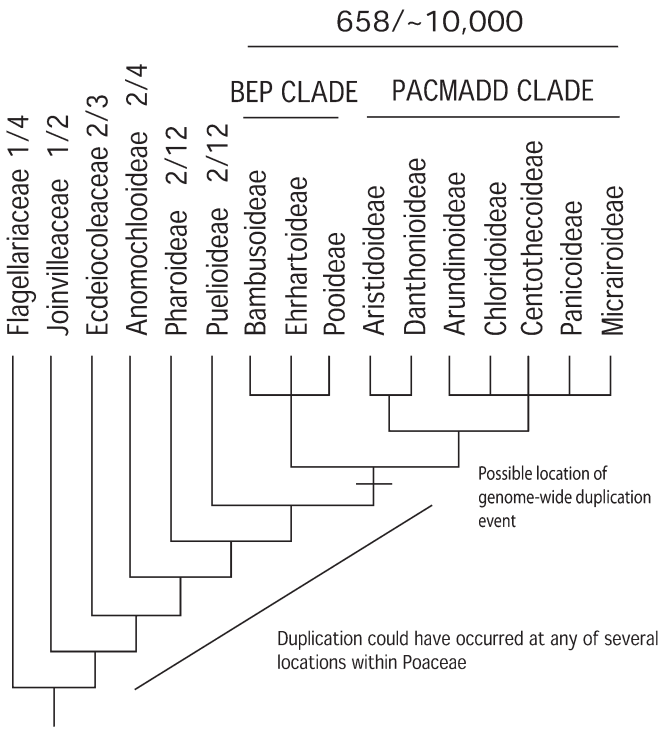

Fig. 4. Placement of inferred genome duplication event detected in Poaceae by Paterson et al. (2004); summary tree modified from that of the Grass Phylogeny Working Group (2001). The precise phylogenetic placement of the genome duplication is still unclear (as represented by angled line), but if this event coincided with, or was close to the origin of core Poaceae (indicated by the slash mark on the tree), it would agree with the hypothesis that polyploidy promotes speciation. The numbers beside each taxon are the number of genera/species.

Solanaceae—A WGD event in Solanaceae has been dated from ca. 50–52 million years before present (Schlueter et al., 2004). Bell et al. (2005) estimated that the age of stem Solanaceae is approximately 49–68 Myr and the crown group 32–50 Myr, in agreement with the estimated age of the duplication as having occurred within Solanaceae. This WGD event characterizes either the clade corresponding to Solanoideae, or perhaps more likely, what has been referred to as the $x = 12$ clade (Olmstead et al., in press) (Fig. 7). The $x = 12$ clade comprises the Solanoideae and Nicotianoideae. Members of Solanaceae

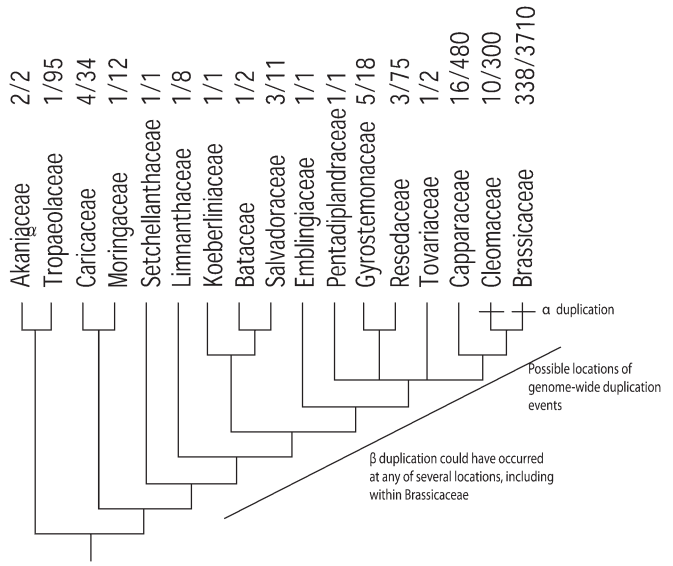

Fig. 5. Placement of inferred genome duplication events in Brassicales—the α and β events occurred in Brassicales, with the α event possibly occurring early in the diversification of Brassicaceae (indicated by slash mark). The precise phylogenetic placement of the β genome duplication is unclear (as represented by angled line), but β likely occurred after the divergence of Caricaceae (see text). An additional triplication event (indicated by slash mark) took place in Cleomaceae (Schranz and Mitchell-Olds, 2006); summary tree modified from D. Soltis et al. (2005) and the Angiosperm Phylogeny Website (Stevens, 2001). The numbers beside each taxon are the number of genera/species.

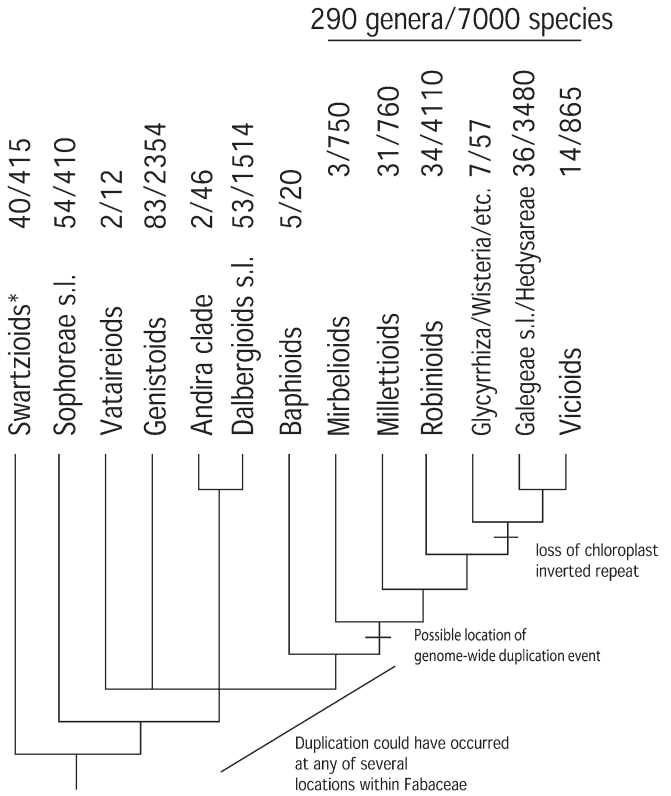

Fig. 6. Putative placement of inferred genome duplication event in Fabaceae detected by Lavin et al. (2005) and Pfeil et al. (2005). The precise phylogenetic placement of the genome duplication is still unclear (as represented by angled line), but one strong possibility for the timing of this event is indicated by the slash mark on the tree; the location of the loss of the chloroplast inverted repeat is also indicted by a slash mark; summary tree modified from Wojciechowski et al. (2004). The numbers beside each taxon are the number of genera/species.

outside of the $x = 12$ clade have a range of chromosome numbers, most from $x = 7$–11, but there are no relic $x = 6$ lineages that could be potential parents of the $x = 12$ clade (Olmstead et al., in press). Hence, our results are consistent with a WGD event that occurred in the ancestor of the $x = 12$ clade or Solanoideae clade (Fig. 7). However, as with the other examples noted, more precision is needed in ascertaining the placement of this ancient duplication.

If this duplication event occurred in either the ancestor of the $x = 12$ clade or the Solanoideae, this ancient WGD event likely had a major impact on diversification. Solanoideae is by far the largest clade within Solanaceae, comprising 61 genera and 1925 species (Stevens, 2001); other clades in Solanaceae are considerably smaller.

ADDITIONAL EVIDENCE AND CONCLUSIONS

Clues regarding the importance of polyploidy in angiosperm diversification may come from studies of recent radiations. Island floras, particularly those inhabiting volcanic chains, may form microcosms for understanding processes behind larger-scale diversifications. Dispersal to volcanic archipelagoes involves the colonization of newly formed habitats that change dramatically over relatively short periods of time with island subsidence. This cycle of environmental instability can repeat itself when further dispersal occurs to younger islands as they form. Early angiosperms may have experienced similar environmental instabilities after the Cretaceous-Tertiary (K-T) Boundary (McElwain and Punyasena, 2007).

The Hawaiian flora has the highest incidence of polyploidy known, and most Hawaiian species are paleopolyploid, having evolved polyploidy prior to the dispersal of their ancestors to Hawaii (Carr, 1998). Both polyploid hybrid and allopolyploid Hawaiian radiations have been studied in detail. The species-rich and morphologically diverse Hawaiian endemic mints, which are high polyploids ($2n = 64$, 66), descend from a relatively recent hybridization event between two different polyploid North American lineages, one with bird-pollinated flowers, the other with insect-pollinated ones (Lindqvist et al., 2003). Their morphological and ecological variation is extensive; plants range from subalpine vines to rainforest shrubs, flowers may have either bird- or insect- pollinated anatomies, and seed dispersal patterns may depend on either dry or fleshy fruits. In another example, it has been shown that the Hawaiian silverswords, with their incredible vegetative diversity, are recent allopolyploids, and the presence of two divergent genomes in their colonizing ancestor may have helped promote adaptive radiation in the alliance (Barrier et al., 1999). Molecular studies of polyploid genome formation have indicated that phenomena such as substantial intragenomic rearrangement and altered gene regulatory relationships, and in allopolyploids, fixed heterozygosity, can contribute to evolutionary flexibility (e.g., Levin, 1983, 2000, 2002; P. Soltis and Soltis, 2000; Wendel, 2000; Wendel and Doyle, 2005; Tate et al., 2005). Polyploid radiations in modern unstable island environments may therefore provide clues as to why some

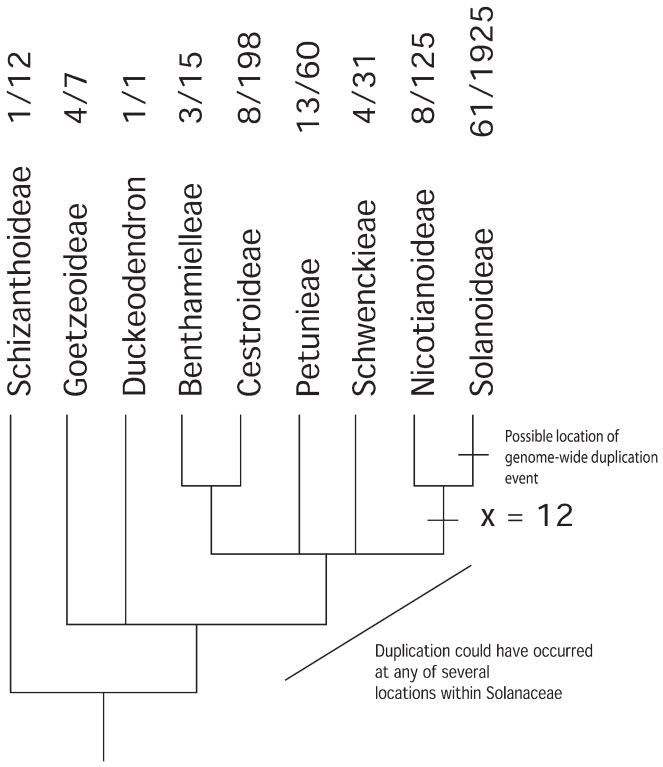

Fig. 7. Putative placement of inferred genome duplication event in Solanaceae detected by Schlueter et al. (2004). The precise phylogenetic placement of the genome duplication is still unclear (as represented by angled line), but one possibility is that this event coincided with the origin of Solanoideae (indicated by the slash mark on the tree); summary tree modified from Olmstead et al. (in press). X = 12 refers to the X = 12 clade of Olmstead et al. (in press). The numbers beside each taxon are the number of genera/species.

angiosperm lineages (e.g., Poaceae and Fabales) underwent high diversification long after their first appearance in the fossil record.

Nowhere else in the history of life is the influence of polyploidy more apparent as a possible diversifying force as it is among the angiosperms. Only the tetraploidization event in the stem lineage of ray-finned fish may come close, having potentially resulted in >20 000 living species (Le Comber and Smith, 2004; Gregory and Mable, 2005). Despite clear evidence for WGD underlying all vertebrates (Muffato and Crollius, 2008), many fewer lineage-specific polyploidy events have been detected than among angiosperms, which have a far shorter combined evolutionary history. Much of this disparity can be attributed to the different nature of plant vs. animal development vis-à-vis compensation for gene dosage (Grant, 1981; Mable, 2004). Subfunctionalization of gene duplicates has been shown to be a prominent molecular evolutionary force for coping with WGD in both flowering plants and ray-finned fish (e.g., Adams et al., 2004; Adams, 2007; Woolfe and Elgar, 2007); still, the angiosperms have diverged into many more habitats, life forms, and biotic interactions, and as such, the importance of their polyploid heritage cannot be underestimated as a force of truly global impact. Future work addressing the influence of polyploidy from the population to whole-lineage levels will clearly be important for understanding these phenomena.

LITERATURE CITED

ADAMS, K. L. 2007. Evolution of duplicated gene expression in polyploid and hybrid plants. *Journal of Heredity* 98: 136–141.

ADAMS, K. L., R. PERCIFIELD, AND J. F. WENDEL. 2004. Organ-specific silencing of duplicated genes in a newly synthesized allotetraploid. *Genetics* 168: 2217–2226.

ADAMS, K. L., AND J. F. WENDEL. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8: 135–141.

ALBERT, V. A., D. E. SOLTIS, J. E. CARLSON, W. G. FARMERIE, P. K. WALL, D. C. ILUT, L. A. MUELLER, ET AL. 2005. Floral gene resources from basal angiosperms for comparative genomics research. *BMC Plant Biology* 5: 5–16.

ALFARO, M. E., F. SANTINI, AND C. D. BROCK. 2007. Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). *Evolution* 61: 2104–2126.

APG II [ANGIOSPERM PHYLOGENY GROUP]. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society* 141: 399–436.

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

BARRIER, M., B. G. BALDWIN, R. H. ROBICHAUX, AND M. D. PURUGGANAN. 1999. Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Molecular Biology and Evolution* 16: 1105–1113.

BELL, C. D., D. E. SOLTIS, AND P. SOLTIS. 2005. The age of the angiosperms: A molecular time scale without a clock. *Evolution* 59: 1245–1258.

BENNETT, M. D., I. J. LEITCH, H. J. PRICE, AND J. S. JOHNSTON. 2003. Comparisons with *Caenorhabditis* (~100 Mb) and Drosophila (~175 Mb) using flow cytometry show genome size in Arabidopsis to be ~157 Mb and thus ~25% larger than the Arabidopsis Genome Initiative estimate of ~125 Mb. *Annals of Botany* 91: 547–557.

BLANC, G., A. BARAKAT, R. GUYOT, R. COOKE, AND I. 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 duplications in the Arabidopsis genome. *Genome Research* 13: 137–144.

BLANC, G., AND K. H. WOLFE. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16: 1667–1678.

BOWERS, J. E., B. A. CHAPMAN, J. RONG, AND A. H. PATERSON. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433–438.

BREMER, K. 2002. Gondwanan evolution of the grass alliance of families (Poales). *Evolution* 56: 1374–1387.

BUZGO, M., P. S. SOLTIS, S. KIM, AND D. E. SOLTIS. 2005. The making of the flower. *Biologist (Columbus, Ohio)* 52: 149–154.

CANNON, S. B., L. STERCK, S. ROMBAUTS, S. SATO, F. CHEUNG, J. GOUZY, X. WANG, ET AL. 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of the National Academy of Sciences, USA* 103: 14959–14964.

CANTINO, P., J. DOYLE, S. GRAHAM, W. JUDD, R. OLMSTEAD, D. E. SOLTIS, P. S. SOLTIS, AND M. DONOGHUE. 2007. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* 56: 822–846.

CARR, G. D. 1998. Chromosome evolution and speciation in Hawaiian flowering plants. *In* T. F. Stuessy and M. Ono [eds.], Evolution and speciation of island plants, 5–47. Cambridge University Press, Cambridge, UK.

CREPET, W. L., AND G. D. FELDMAN. 1991. The earliest remains of grasses in the fossil record. *American Journal of Botany* 78: 1010–1014.

CLAUSEN, J., D. D. KECK, AND W. M. HIESEY. 1945. Experimental studies on the nature of species. II. Plant evolution through amphiploidy and autopolyploidy, with examples from the Madiinae. Carnegie Institute of Washington, Washington, D.C., USA.

CUI, L., P. K. WALL, J. LEEBENS-MACK, B. G. LINDSAY, D. E. SOLTIS, J. J. DOYLE, P. S. SOLTIS. 2006. Widespread genome duplications

throughout the history of flowering plants. *Genome Research* 16: 738–749.

DARLINGTON, C. D. 1937. Recent advances in cytology, 2nd ed. P. Blakiston's, Philadelphia, Pennsylvania, USA.

DOYLE, J. A. 2008. Integrating molecular phylogenetic and paleobotanical evidence on origin of the flower. *International Journal of Plant Sciences* 169: 816–843.

EL-MABROUK, N., AND D. SANKOFF. 2003. The reconstruction of doubled genomes. *SIAM Journal on Computing* 32: 754–792.

FRIIS, E. M., K. R. PEDERSEN, AND P. R. CRANE. 2006. Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Paleogeography, Paleoclimatology, Paleoecology* 232: 251–293.

FROHLICH, M. W., AND M. W. CHASE. 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450: 1184–1189.

GATES, R. R. 1909. The stature and chromosomes of *Oenothera gigas* De Vries. Arch. f. *Zellforschung* 3: 525–552.

GOLDBLATT, P. 1980. Polyploidy in angiosperms: Monocotyledons. *In* W. H. Lewis [ed.], Polyploidy: Biological relevance, 219–239. Plenum Press, New York, New York, USA.

GRANT, V. 1963. The origin of adaptations. Columbia University Press, New York, New York, USA.

GRANT, V. 1981. Plant speciation, 2nd ed. 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.

GREGORY, T. R., AND B. K. MABLE. 2005. Polyploidy in animals. *In* T. R. Gregory [ed.], The evolution of the genome, 428–501. Elsevier Academic Press, San Diego, California, USA.

HARMON, L. J., J. T. WEIR, C. D. BROCK, R. E. GLOR, AND W. CHALLENGER. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics (Oxford, England)* 24: 129–131.

HUGHES, N. F. 1994. The enigma of angiosperm origins. Cambridge University Press, Cambridge, UK.

HUSBAND, B. C. 2004. The role of triploids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal Linnean Society* 82: 537–546.

INTERNATIONAL RICE GENOME SEQUENCING PROJECT. 2005. The map-based sequence of the rice genome. *Nature* 436: 793–800.

IRISH, V. 2006. Duplication, diversification, and comparative genetics of angiosperm MADS-box genes. *Advances in Botanical Research* 44: 127–159.

JACKSON, S., S. ROUNSLEY, AND M. PURUGGANAN. 2006. Comparative sequencing of plant genomes: Choices to make. *Plant Cell* 18: 1100–1104.

JAILLON, O., J. M. AURY, B. NOEL, A. POLICRITI, C. CLEPET, A. CASAGRANDE, N. CHOISNE, ET AL., FOR THE FRENCH-ITALIAN PUBLIC CONSORTIUM FOR GRAPEVINE GENOME CHARACTERIZATION. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.

JANSSEN, T., AND K. BREMER. 2004. The age of major monocot groups inferred from 800+ *rbcL* sequences. *Botanical Journal of the Linnean Society* 146: 385–398.

JANSEN, R. K., Z. CAI, L. A. RAUBESON, H. DANIELL, C. W. DEPAMPHILIS, J. LEEBENS-MACK, K. F. MÜLLER, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374.

KIM, S., V. A. ALBERT, M. J. YOO, J. S. FARRIS, P. S. SOLTIS, AND D. E. SOLTIS. 2004. Pre-angiosperm duplication of floral genes and regulatory tinkering at the base of angiosperms. *American Journal of Botany* 91: 2102–2118.

KIM, S., J. KOH, M. J. YOO, H. KONG, Y. HU, H. MA, P. S. SOLTIS, AND D. E. SOLTIS. 2005. Expression of floral MADS-box genes in basal angiosperms: Implications on evolution of floral regulators and the perianth. *Plant Journal* 43: 724–744.

KOCH, M. A., AND M. KIEFER. 2005. Genome evolution among cruciferous plants: A lecture from the comparison of the genetic maps of three diploid species—*Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. *American Journal of Botany* 92: 761–767.

KOWALSKI, S., T.-H. LAN, K. FELDMANN, AND A. PATERSON. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved gene order. *Genetics* 138: 499–510.

KRAMER, E. M., R. L. DORIT, AND V. F. IRISH. 1998. Molecular evolution of genes controlling petal and stamen development: Duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* 149: 765–783.

KRAMER, E. M., M. A. JARAMILLO, AND V. S. DI STILIO. 2004. Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. *Genetics* 166: 1011–1023.

KRAMER, E. M., AND E. A. ZIMMER. 2006. Gene duplication and floral developmental genetics of basal eudicots. *Advances in Botanical Research* 44: 354–376.

KUITTINEN, H., A. A. DE HAAN, C. VOGL, S. OIKARINEN, J. LEPPALA, M. KOCH, T. MITCHELL-OLDS, et al. 2004. Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. *Genetics* 168: 1575–1584.

KUWADA, Y. 1911. Meiosis in the pollen mother cells of *Zea mays* L. *Botanical Magazine, Tokyo* 25: 163–181.

LAGERCRANTZ, U., AND D. J. LYDIATE. 1996. Comparative genome mapping in *Brassica*. *Genetics* 144: 1903–1910.

LAN, T. H., T. A. DELMONTE, K. P. REISCHMANN, J. HYMAN, S. P. KOWALSKI, J. MCFERSON, S. KRESOVICH, AND A. H. PATERSON. 2000. An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. *Genome Research* 10: 776–788.

LAVIN, M., P. HERENDENN, AND M. F. WOJCIECHOWSKI. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* 54: 575–594.

LAWRENCE, W. J. C. 1931. The secondary association of chromosomes. *Cytologia* 2: 342–384.

LE COMBER, S. C. L., AND C. SMITH. 2004. Polyploidy in fishes: Patterns and processes. *Biological Journal of the Linnean Society* 82: 431–442.

LEEBENS-MACK, J., L. A. RAUBESON, L. CUI, J. V. KUEHL, M. H. FOURCADE, T. W. CHUMLEY, J. L. BOORE, R. K. JANSEN, AND C. W. DEPAMPHILIS. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 22: 1948–1963.

LEEBENS-MACK, J.-H., K. WALL, J. DUARTE, Z. ZHENG, D. OPPENHEIMER, AND C. DEPAMPHILIS. 2006. A genomics approach to the study of ancient polyploidy and floral developmental genetics. *Advances in Botanical Research* 44: 526–540.

LEITCH, I. J., AND M. D. BENNETT. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* 82: 651–663.

LEITCH, A. R., AND I. J. LEITCH. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* 320: 481–483.

LEITCH, A. R., D. E. SOLTIS, P. S. SOLTIS, I. J. LEITCH, AND J. C. PIRES [eds.]. 2004. Biological relevance of polyploidy: Ecology to genomics. *Biological Journal of the Linnean Society* 82: 409–700.

LEVIN, D. A. 1983. Polyploidy and novelty in flowering plants. *American Naturalist* 122: 1–25.

LEVIN, D. A. 2000. The origin, expansion and demise of plant species. Oxford University Press, New York, New York, USA.

LEVIN, D. A. 2002. The role of chromosomal change in plant evolution. Oxford University Press, New York, New York, USA.

LEWIS, W. H. 1980. Polyploidy in angiosperms: Dicotyledons. *In* W. H. Lewis [ed.], Polyploidy: Biological relevance, 241–268. Plenum Press, New York, New York, USA.

LINDQVIST, C., T. J. MOTLEY, J. J. JEFFREY, AND V. A. ALBERT. 2003. Cladogenesis and reticulation in the Hawaiian endemic mints (Lamiaceae). *Cladistics* 19: 480–495.

LITT, A., AND V. F. IRISH. 2003. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: Implications for the evolution of floral development. *Genetics* 165: 821–833.

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

LÖVE, A., AND D. LÖVE. 1949. The geobotanical significance of polyploidy. I. Polyploidy and latitude. *Portugaliae Acta Biologica (A) R. B. Goldschmidt Special Volume*: 273–352.

LUKENS, L., P. QUIJADA, J. UDALL, J. C. PIRES, M. E. SCHRANZ, AND T. C. OSBORN. 2004. Genome redundancy and plasticity within ancient and recent *Brassica* crop species. *Biological Journal of the Linnean Society* 82: 665–674.

LUTZ, A. M. 1907. A preliminary note on the chromosomes of *Oenothera lamarckiana* and one of its mutants, *O. gigas*. *Science* 26: 151–152.

LYNCH, M., AND A. FORCE. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154: 459–473.

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

MABLE, B. K. 2004. Why polyploidy is rarer in animals that in plants: Myths and mechanisms. *Biological Journal of the Linnean Society* 82: 453–466.

MAGALLÓN, S., AND M. J. SANDERSON. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55: 1762–1780.

MAGALLÓN, S., AND M. J. SANDERSON. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution; International Journal of Organic Evolution* 59: 1653–1670.

MALCOMBER, S. T., AND E. A. KELLOGG. 2005. *SEPALLATA* gene diversification: Brave new whorls. *Trends in Plant Science* 10: 427–435.

MARCHANT, A. D., AND B. G. BRIGGS. 2007. Ecdeiocoleaceae and Joinvilleaceae, sisters of Poaceae (Poales): Evidence from *rbcL* and *matK* data. *Telopea* 11: 437–450.

MASTERSON, J. 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* 264: 421–423.

MCELWAIN, J. C., AND S. W. PUNYASENA. 2007. Mass extinction events and the plant fossil record. *Trends in Ecology & Evolution* 22: 548–557.

MCGRATH, J. M., M. M. JANCSO, AND E. PICHERSKY. 1993. Duplicate sequences with a similarity to expressed genes in the genome of *Arabidopsis thaliana*. *Theoretical and Applied Genetics* 86: 880–888.

MING, R., S. HOU, Y. FENG, Q. YU, A. DIONNE-LAPORTE, J. H. SAW, P. SENIN, et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* L.). *Nature* 452: 991–996.

MOORE, M. J., C. D. BELL, P. S. SOLTIS, AND D. E. SOLTIS. 2007. Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.

MUFFATO, M., AND H. R. CROLLIUS. 2008. Paleogenomics in vertebrates, or the recovery of lost genomes from the mist of time. *BioEssays* 30: 122–134.

MÜNTZING, A. 1936. The evolutionary significance of autopolyploidy. *Hereditas* 21: 263–378.

OLMSTEAD, R. G., L. BOHS, H. A. MIGID, E. SANTIAGO-VALENTIN, V. F. GARCIA, AND S. M. COLLIER. In press. A molecular phylogeny of the Solanaceae. *Taxon*.

OSBORN, T. C., J. C. PIRES, J. A. BIRCHLER, D. L. AUGER, Z. J. CHEN, H.-S. LEE, L. COMAI, A. MADLUNG, R. W. DOERGE, V. COLOT, AND R. A. MARTIENSSEN. 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics* 19: 141–147.

OTTO, S. P., AND J. WHITTON. 2000. Polyploidy: Incidence and evolution. *Annual Review of Genetics* 34: 401–437.

PATERSON, A. H., J. BOWERS, M. BUROW, X. DRAYE, C. ELSIK, C. JIANG, C. KATSAR, et al. 2000. Comparative genomics of plant chromosomes. *Plant Cell* 12: 1523–1539.

PATERSON, A. H., J. E. BOWERS, AND B. A. CHAPMAN. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences, USA* 101: 9903–9908.

PATERSON, A. H., B. A. CHAPMAN, J. C. KISSINGER, J. E. BOWERS, F. A. FELTUS, AND J. C. ESTILL. 2006. Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends in Genetics* 22: 597–602.

PFEIL, B. E., J. A. SCHLUETER, R. C. SHOEMAKER, AND J. J. DOYLE. 2005. Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Systematic Biology* 54: 441–454.

PRYER, K. M., H. SCHNEIDER, E. A. ZIMMER, AND J. A. BANKS. 2002. Deciding among green plants for whole genome studies. *Trends in Plant Science* 7: 550–554.

QUIROS, C. F., F. GRELLET, J. SADOWSKI, T. SUZUKI, G. LI, AND T. WROBLEWSKI. 2001. *Arabidopsis* and *Brassica* comparative genomics: Sequence, structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions. *Genetics* 157: 1321–1330.

RAMSEY, J. R., AND D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467–501.

RAMSEY, J. R., AND D. W. SCHEMSKE. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 33: 589–639.

RAPP, R. A., AND J. F. WENDEL. 2005. Epigenetics and plant evolution. *New Phytologist* 168: 81–91.

RIESEBERG, L. H., AND J. H. WILLIS. 2007. Plant speciation. *Science* 317: 910–914.

SAMPEDRO, J., Y. LEE, R. E. CAREY, C. DEPAMPHILIS, AND D. J. COSGROVE. 2005. Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. *Plant Journal* 44: 409–419.

SANDERSON, M. J., AND J. A. DOYLE. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. *American Journal of Botany* 88: 1499–1516.

SANKOFF, D., C. ZHENG, AND Q. ZHU. 2007. Polyploids, genome halving and phylogeny. *Bioinformatics (Oxford, England)* 23: i433–i439.

SAARELA, J. M., H. S. RAI, J. A. DOYLE, P. K. ENDRESS, S. MATHEWS, A. D. MARCHANT, B. G. BRIGGS, AND S. W. GRAHAM. 2007. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* 446: 312–315.

SCHLUETER, J. A., P. DIXON, C. GRANGER, D. GRANT, L. CLARK, J. J. DOYLE, AND R. C. SCHOENMAKER. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47: 868–876.

SCHRANZ, E. M., AND T. MITCHELL-OLDS. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18: 1152–1165.

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., P. S. SOLTIS, M. W. CHASE, M. E. MORT, D. C. ALBACH, M. ZANIS, V. SAVOLAINEN, et al. 2000. Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

SOLTIS, D. E., V. A. ALBERT, J. LEEBENS-MACK, J. PALMER, R. WING, C. DEPAMPHILIS, H. MA, et al. 2008a. The *Amborella* Genome Initiative: A genome for understanding the evolution of angiosperms. *Genome Biology* 9: 402 .

SOLTIS, D. E., C. D. BELL, S. KIM, AND P. S. SOLTIS. 2008b. The origin and early evolution of the angiosperms. *Annals of the New York Academy of Sciences* 1133: 3–25.

SOLTIS, D. E., AND P. S. SOLTIS. 1990. Isozyme evidence for ancient polyploidy in primitive angiosperms. *Systematic Botany* 15: 328–337.

SOLTIS, D. E., AND P. S. SOLTIS. 1999. Polyploidy: Origins of species and genome evolution. *Trends in Ecology & Evolution* 14: 348–352.

SOLTIS, D. E., P. S. SOLTIS, M. W. CHASE, AND P. K. ENDRESS. 2005. Phylogeny, evolution, and classification of flowering plants. Sinauer, Sunderland, Massachusetts, USA.

SOLTIS, D. E., P. S. SOLTIS, D. W. SCHEMSKE, J. F. HANCOCK, J. N. THOMPSON, B. C. HUSBAND, AND W. S. JUDD. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56: 13–30.

SOLTIS, D. E., P. S. SOLTIS, AND J. A. TATE. 2003. Advances in the study of polyploidy since *Plant Speciation*. *New Phytologist* 161: 173–191.

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 Sciences, USA* 97: 7051–7057.

SOLTIS, P. S., D. E. SOLTIS, AND M. W. CHASE. 1999. Angiosperm phylogeny inferred from multiple genes: A research tool for comparative biology. *Nature* 402: 402–404.

SOLTIS, P. S., D. E. SOLTIS, S. KIM, A. CHANDERBALI, AND M. BUZGO. 2006. Expression of floral regulators in basal angiosperms and the origin and evolution of ABC-function. *Advances in Botanical Research* 44: 483–506.

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

STEBBINS, G. L. JR. 1940. The significance of polyploidy in plant evolution. *American Naturalist* 74: 54–66.

STEBBINS, G. L. 1947. Types of polyploids: Their classification and significance. *Advances in Genetics* 1: 403–429.

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

STEBBINS, G. L. 1971. Chromosomal evolution in higher plants. Addison-Wesley, London, UK.

STEBBINS, G. L. 1985. Polyploidy, hybridization, and the invasion of new habitats. *Annals of the Missouri Botanical Garden* 72: 824–832.

STEVENS, P. F. 2001 [onward]. Angiosperm phylogeny website, version 8, June 2007 (and more or less continuously updated since). Website http://www.mobot.org/MOBOT/research/APweb.

TANG, H., J. E. BOWERS, X. WANG, R. MING, M. ALAM, AND A. H. PATERSON. 2008. Synteny and collinearity in plant genomes. *Science* 320: 486–488.

TATE, J. A., D. E. SOLTIS, AND P. S. SOLTIS. 2005. Polyploidy in plants. *In* T. R. Gregory [ed.], The evolution of the genome, 371–426. Elsevier Academic Press, San Diego, California, USA.

THOMPSON, J. N., B. M. CUNNINGHAM, K. A. SEGRAVES, D. M. ALTHOFF, AND D. WAGNER. 1997. Plant polyploidy and insect/plant interactions. *American Naturalist* 150: 730–743.

THOMPSON, J. N., S. L. NUISMER, AND K. MERG. 2004. Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biological Journal Linnean Society* 82: 503–510.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV, U. HELLSTEN, N. PUTNAM *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604.

VAN DE PEER, Y., AND A. MEYER. 2005. Large-scale gene and ancient genome duplications. *In* T. R. Gregory [ed.], The evolution of the genome, 329–370. Elsevier Academic Press, San Diego, California, USA.

VELASCO, R., A. ZHARKIKH, M. TROGGIO, D. A. CARTWRIGHT, A. CESTARO, D. PRUSS, M. PINDO, ET AL. 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2: e1326, doi:10.1371/journal.pone.0001326.

VERON, A. S., K. KAUFMANN, AND E. BORNBERG-BAUER. 2007. Evidence of interaction network evolution by whole-genome duplications: A case study in MADS-box proteins. *Molecular Biology and Evolution* 24: 670–678.

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

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

WENDEL, J. F., AND J. J. DOYLE. 2005. Polyploidy and evolution in plants. *In* R. Henry, [ed.], Diversity and evolution in plants, 97–117. CABI Publishing, Oxon, UK.

WERTH, C. R., AND M. D. WINDHAM. 1991. A model for divergent, allopatric speciation of polyploid Pteridophytes resulting from silencing of duplicate-gene expression. *American Naturalist* 137: 515.

WIKSTRÖM, N., V. SAVOLAINEN, AND M. W. CHASE. 2001. Evolution of the angiosperms: Calibrating the family tree. *Proceedings of the Royal Society of London, B, Biological Sciences* 268: 2211–2220.

WOJCIECHOWSKI, M. F., M. LAVIN, AND M. J. SANDERSON. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846–1862.

WOOLFE, A., AND G. ELGAR. 2007. Comparative genomics using *Fugu* reveals insights into regulatory subfunctionalizaation. *Genome Biology* 8: R53.

YOGEESWARAN, K., A. FRARY, T. L. YORK, A. AMENTA, A. H. LESSER, J. B. NASRALLAH, S. D. TANKSLEY, AND M. E. NASRALLAH. 2005. Comparative genome analyses of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Research* 15: 505–515.

YU, J., J. WANG, W. LIN, S. LI, H. LI, J. ZHOU, P. NEI et al. 2005. The genomes of *Oryza sativa*: A history of duplications. *PLoS Biology* 3: e38, doi:10.1371/journal.pbio.0030038 .

ZAHN, L. M., H. KONG, J. H. LEEBENS-MACK, S. KIM, P. S. SOLTIS, L. L. LANDHERR, D. E. SOLTIS. et al. 2005. The evolution of the *SEPALLATA* subfamily of MADS-box genes: A pre-angiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169: 2209–2223.

ZHENG, C., Q. ZHU, AND D. SANKOFF. 2006. Genome halving with an outgroup. *Evolutionary Bioinformatics* 2: 319–326.

ZHENG, C., P. K. WALL, J. LEEBENS-MACK, C. DEPAMPHILIS, V. A. ALBERT, AND D. SANKOFF. 2008a. The effect of massive gene loss following whole genome duplication on the algorithmic reconstruction of the ancestral *Populus* diploid. *In* P. Markstein and Y. Xu [eds.], Computational Systems Bioinformatics, Volume 7, Proceedings of the CSB Conference, Stanford University, California, August 2008, 261–271. Imperial College Press, London.

ZHENG C., Q. ZHU, Z. ADAM, AND D. SANKOFF. 2008b. Guided genome halving: hardness, heuristics and the history of the Hemiascomycetes. *Bioinformatics* 24: i96–i104.

ZHENG C., Q. ZHU, AND D. SANKOFF. 2008c. Descendants of whole genome duplication within gene order phylogeny. *Journal of Computational Biology* 15: 1–18.