Recently Formed Polyploid Plants Diversify at Lower Rates

Itay Mayrose,[1]* Shing H. Zhan,[1] Carl J. Rothfels,[2] Karen Magnuson-Ford,[1] Michael S. Barker,[3] Loren H. Rieseberg,[1,4] Sarah P. Otto[1]

Polyploidy (or whole-genome duplication) is a widespread feature of plant genomes, but its importance to evolution has long been debated. Polyploids have been postulated to be evolutionary dead ends because of the inefficiency of selection when genes are masked by multiple copies (*1*). However, most plant species have experienced at least one genome doubling early in their history (*2*), suggesting that rather than being an evolutionary dead end, polyploidy is a route to evolutionary success. A recent study (*3*) confirmed the ubiquity of polyploidy, with about 35% of vascular plant species being recent polyploids ("neopolyploids," having formed since their genus arose), representing 15% of speciation events in flowering plants and 31% in ferns. It remains unknown, however, whether the abundance of polyploids is a consequence of higher diversification rates following polyploidy or of frequent polyploid formation.

We estimated diversification rates of neopolyploids relative to their diploid congeners. We compiled a data set of angiosperm ($n = 49$) and seed-free vascular plant (SFVP, including ferns and lycophytes; $n = 14$) generic-level groups in which ploidy levels could be estimated from cytological and phylogenetic data (*4*). Over 500 ploidy shifts were inferred with a probabilistic model of chromosome number evolution that accounts for aneuploid and polyploid transitions but not diversification rate differences (*5*). This allowed us to label all descendants of a polyploidization event as neopolyploids, even when lacking chromosome data.

Likelihood analyses indicated that 33% of the examined species are neopolyploids (609/2043 for angiosperms and 209/458 for SFVPs), matching earlier estimates (*1*, *3*). Polyploidization events were not distributed uniformly across phylogenies but were disproportionately represented on the tips of the tree of life [$\chi^2_1 = 90.5$ (all data); 48.2 (angiosperms); 45.1 (SFVPs); $P << 0.01$ (*4*)], suggesting that newly formed polyploid lineages generally fail to persist.

To estimate diversification rates, we used the binary state speciation and extinction (BiSSE) model (*6*) to coestimate diversification rates associated with diploids versus neopolyploids. Defining polyploids as those lineages that underwent a polyploidization event since divergence from their generic ancestor, the transition rate from polyploidy to diploidy was set to zero [but see (*4*)]. Across our data set, the speciation rates of neopolyploids were significantly lower than that of diploids ($P < 10^{-3}$; t test), and their extinction rates were significantly higher ($P < 10^{-12}$). Together, neopolyploid lineages exhibit significantly reduced rates of diversification (speciation minus extinction) ($P < 10^{-12}$) (Fig. 1).

The inferred difference in speciation rates between diploids and polyploids may be driven by a greater propensity of diploids to speciate via polyploidization relative to neopolyploids. We extended BiSSE to allow ploidy transitions only at speciation events (*4*) and inferred the frequency of diploid speciation events that involve polyploidization and those that do not (heteroploid and homoploid speciation, respectively). Discounting diploids that underwent heteroploid speciation, the difference in speciation rates between diploids and polyploids was no longer significant ($P > 0.1$). Nevertheless, the diversification rates of polyploids remained significantly lower than that of diploids ($P < 10^{-6}$; fig. S2) because of the higher extinction rate of neopolyploids. The average frequency of heteroploid speciation was 31.7% for all plants, 29.7% for angiosperms, and 38.7% for SFVPs, exceeding previous estimates that ignored extinction rate differences. Our estimates for the rate at which diploids speciate via polyploidization likely represent upper bounds, however, because only phylogenies with variation in ploidy were examined and because ploidy transitions were allowed only at speciation events.

The lower diversification rates of polyploids may seemingly contradict evidence of ancient polyploidization events in the genomes of most angiosperms (*2*). Yet we find that the expected number of paleopolyploidization events is higher than would be observed if diversification rates were equal (*4*). Our results indicate that polyploidy is most often an evolutionary dead end, but the possibility remains that the expanded genomic potential of those polyploids that do persist drives longer-term evolutionary success.

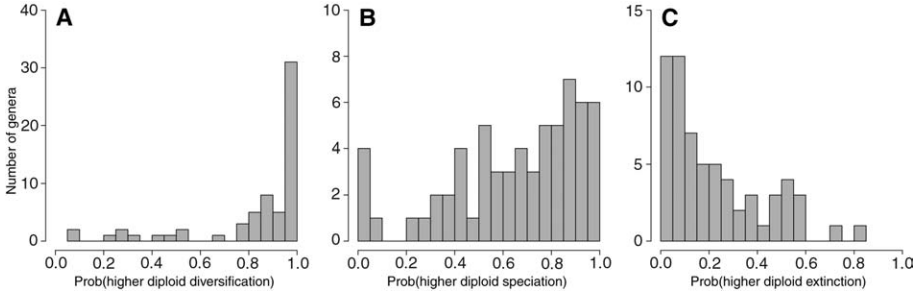

Fig. 1. The posterior probabilities that diploids exhibit higher rates of (**A**) diversification, (**B**) speciation, and (**C**) extinction than polyploids.

References and Notes

1. G. Stebbins, *Chromosomal Evolution in Higher Plants* (Edward Arnold, London, 1971).
2. Y. Jiao *et al.*, *Nature* **473**, 97 (2011).
3. T. E. Wood *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13875 (2009).
4. Materials and methods are available as supporting material on *Science* Online.
5. I. Mayrose, M. S. Barker, S. P. Otto, *Syst. Biol.* **59**, 132 (2010).
6. W. P. Maddison, P. E. Midford, S. P. Otto, *Syst. Biol.* **56**, 701 (2007).

Acknowledgments: We thank R. FitzJohn for instrumental help in using diversitree. Funded by Natural Sciences and Engineering Research Council Discovery grants to L.H.R. and S.P.O. Data have been deposited at Dryad (10.5061/dryad.6hf21).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1207205/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Tables S1 to S3
References (*7* to *103*)

19 April 2011; accepted 22 July 2011
Published online 18 August 2011;
10.1126/science.1207205

[1]Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [2]Department of Biology, Duke University, Durham, NC 27708, USA. [3]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. [4]Department of Biology, Indiana University, Bloomington, IN 47405, USA.

*To whom correspondence should be addressed. Present address: Department of Molecular Biology and Ecology of Plants, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. E-mail: itaymay@post.tau.ac.il

Originally posted 18 August 2011; corrected 1 September 2011

www.sciencemag.org/cgi/content/full/science.1207205/DC1

Supporting Online Material for

Recently Formed Polyploid Plants Diversify at Lower Rates

Itay Mayrose,* Shing H. Zhan, Carl J. Rothfels, Karen Magnuson-Ford, Michael S. Barker, Loren H. Rieseberg, Sarah P. Otto

*To whom correspondence should be addressed. E-mail: itaymay@post.tau.ac.il

Published 18 August 2011 on *Science* Express
DOI: 10.1126/science.1207205

This PDF file includes:

Corrections: Several of the references were updated or completed.

Supporting online material

Materials and methods

Database construction

Our initial set of phylogenies was based on the list of 123 angiosperm and 20 seed-free vascular plant (SFVP) groups presented in Wood et al. (*3*). Twenty-five additional phylogenies were added following a subsequent literature search. These phylogenies are primarily at the genus level with a few phylogenies focused on sections, subgenera, or a cluster of closely related genera. For ease of reference, we refer to the groups examined as genera. A large number of these phylogenies were originally reconstructed from morphological data, and so were unavailable to our likelihood-based analyses, which require branch-lengths to be proportional to time. For these phylogenies, we searched the literature for a molecular-based tree of the same focal genus. The phylogeny was removed if no such study was found.

Molecular sequence data were obtained through TreeBase (http://www.treebase.org/). If not available, accession numbers were collected from the original studies and retrieved from GenBank (*7*). In cases where the original studies were based on multiple data partitions we created a combined data matrix (if such combined data were supported by the original study), as long as the combined analysis reduced the number of species by less than 20%. Otherwise, we selected the data matrix containing the largest number of species and verified that the reconstructed phylogeny (see below) was congruent with the tree presented in the original study. Multiple sequence alignments were created using Muscle (*8*), unless the alignment from the original study could be retrieved through TreeBase, in which case we used the latter. The findModel server (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) was then used to find the best-fitting substitution model for each alignment. Maximum likelihood (ML) trees, which were used in the polyploidy inference analyses (see below), were reconstructed using PhyML (*9*) under the optimal substitution model with rate variation across sites modeled using a gamma distribution with four discrete categories. The tree was rooted using outgroup taxa as specified in the original study. The outgroup was subsequently pruned from the tree, and the tree was ultrametricized using penalized likelihood (*10*) in r8s (*11*). The most appropriate smoothing parameter was chosen based on an initial cross-validation run. Bayesian trees were reconstructed using MrBayes version 3.2 (*12*) under a relaxed molecular clock according to a Brownian motion model (*13*). Each analysis consisted of four Markov chains (with heating according to default settings), run for 225,000 steps and sampled every 2000 steps. The first 125,000 steps were considered as burn-in and discarded from the analyses. The resulting MrBayes trees are already clock-like, and so ultrametrization is unnecessary.

A few datasets contained multiple accessions per species. Special care was taken to ensure that the choice of the included accession did not affect the results. In cases where the multiple accessions formed a monophyletic group in the resulting ML tree, a random representative accession was chosen. If a paraphyletic group for a single species was observed, we chose a sequence from the largest monophyletic group. In cases where no such group was observed but the accession choice did not change the categorization (polyploid versus diploid) of species in the tree (see below), a random accession was chosen. Finally, the study was excluded if the accession choice resulted in different inferences of species as diploids or polyploids.

Chromosome counts were based on the original phylogenetic study, if reported. Additional counts were obtained from the index to plant chromosome numbers database (http://www.tropicos.org/Project/IPCN; (*14*)), the Plant DNA C-values database (http://data.kew.org/cvalues) and surveys of older compendiums (*15, 16*). Where there were discrepancies between sources, we chose the chromosome number reported in the original phylogenetic study.

Diversity counts for each genus, used in the diversification analysis (see below), were taken from the original study if reported. Additional estimates were obtained from Mabberley (*17*) and other sources (listed in Table S1).

Datasets were eliminated if they fell into any of the following categories: (1) less than ten species had sequence data (the exception is *Leavenworthia*, a genus with only eight species but all have both sequence and chromosome number data); (2) chromosome counts were not available from enough species (either fewer than six species with count data or count data available for less than 35% of the sampled species in the group); (3) chromosome counts in the dataset formed an aneuploid series (e.g., 3, 4, 5, 6, 7, 11) that prevented a reliable designation of diploid and polyploid states; or (4) no polyploidy transitions were inferred (see below). The final database included 63 phylogenies with both angiosperm (n = 49) and SFVP (n = 14) genera represented. The average number of sampled species per genus is 39, for an average coverage of 65% of the species in the generic-level groups.

Table S1 lists all phylogenetic studies and associated data used in our analyses as well as the type of loci used in our analyses and the life history characteristics of the various groups. To get a sense of the age of these groups, we also provide in Table S1 the tree height (average distance from the root to the tips in terms of number of nucleotide substitutions per site), which illustrates that the various groups examined exhibit a fairly similar range of tree heights. We do not provide age estimates in years because, in the absence of a calibration point, the rate of a molecular clock would need to be specified for each group. Given a molecular clock, one could then divide the tree height by the number of substitutions per year to obtain an approximate age. For example, as a rough guide, Muse (*18*) estimated the rate of the molecular clock to be 6×10^{-9}, 2×10^{-9}, and 0.3×10^{-9} synonymous substitutions per site per year for nuclear, chloroplast, and mitochondrial genes, respectively. However, the substitution rate is known to vary across lineages [e.g., substitution rates were found to be higher in ferns than in seed plants (*19*)], among loci, and to be influenced by life history attributes [e.g., herbaceous plants exhibit higher rates compared to trees/shrubs and annuals evolve faster than perennials (20)], and should thus be used with great caution.

Polyploidy inference

Given the maximum likelihood tree and the assignments of chromosome numbers to the tips, we next inferred extant taxa as diploid or polyploid relative to the base chromosome number of the group examined, using chromEvol (*5*). This likelihood-based method assesses the fit of models of chromosome number change along the phylogeny and infers transitions in chromosome number along branches of the tree. We ran all eight available models and used the Akaike information criterion (*21*) to select the best model for each dataset. The expected number of ploidy transitions along each branch of the phylogeny was recorded based on the best-fitting model. An extant taxon was categorized as a polyploid if the ML estimate of the number of ploidy transitions from the root to the tip exceeded 0.9 and as diploid otherwise (for the majority of the cases, altering the threshold used to distinguish polyploids from diploids produced similar results and matched those reported in the original phylogenetic studies). By doing so we implicitly treated the root of each phylogeny as diploid. Thus, polyploids are defined here as those lineages that underwent a polyploidization event since the divergence from the common ancestor of the group examined; our definition of polyploidy thus includes only neopolyploids. This procedure does not allow us to differentiate between genome duplications within a species (autopolyploidy) and those resulting from hybridization (allopolyploidy). Determining the origin of large numbers of polyploid species is currently not feasible for the scale of this study, but future databases that distinguish allo- and auto-polyploid species could be analyzed using similar methods to determine their relative effects on diversification.

This methodology allowed us to categorize an extant species as polyploid or diploid regardless of whether chromosome number data were available for that specific taxon. The ploidy inferences were examined manually and compared to those reported in the original phylogenetic study, if available. Table S1 lists the percentage of polyploid species in each dataset as inferred by chromEvol and Figure S3 presents several representative phylogenies with the inferred polyploidization events. In a number of cases, the ploidy inferences seemed questionable for sub-clades with high proportions of missing

chromosome numbers. To be conservative, we changed the ploidy states for these species to "not available" (NA in the diversitree R package) in our diversification analyses.

We note that our methodology for assigning ploidy levels to extant taxa does not account for different diversification rates of diploids and polyploids. This may bias our estimates of ploidy levels among extant taxa, particularly for those taxa with missing chromosome counts. To verify that this potential bias did not substantively affect our conclusions, we repeated the diversification analysis (detailed below) with ploidy levels assigned only to those species with available cytological data, labeling the missing chromosome count data as NA. This procedure yielded very similar results to those reported in the main text. In addition, very similar results were obtained when ploidy-level assignments were not based on the chromEvol reconstructions but instead, following Stebbins (1), regarding all species with more than twice the minimum chromosome number for the genus as polyploids and labeling the missing chromosome count data as NA.

Diversification analysis

As a first indication of ploidy-dependent diversification rates we evaluated whether polyploidization events occur more often than expected along external branches (those leading directly to extant species) or internal branches of the tree. For each phylogeny, the expected number of polyploidization events along each branch, as inferred by chromEvol, was calculated as the relative contribution of the branch to the total tree length (i.e., the branch length divided by the sum of all branch lengths of the tree) times the total number of inferred polyploidization events in the phylogeny. The chi-square test of independence was then applied to test whether the observed distribution of polyploidization events at internal and external branches differ significantly from expectations. In this analysis we excluded external branches that were exceptionally long (more than half of the tree height, defined as the average sum of branch lengths from the root to the tips).

For a more rigorous analysis we applied the binary state speciation and extinction (BiSSE) model (*6*) to estimate diversification rates for diploids and neopolyploids. BiSSE co-estimates six parameters: speciation rates of lineages in state P (polyploid) and D (diploid) (λ_P and λ_D, respectively); extinction rates of lineages in state P and D (μ_P and μ_D); and transition rates from P to D (q_{PD}) and D to P (q_{DP}). Because we defined as polyploid those species that underwent a polyploidization event sometime since divergence from their generic ancestor, we fixed q_{PD} to zero (but see below). This constraint is also compatible with the common assumption that polyploidy is largely an irreversible process, especially over a short time interval (*22*). Our analyses were performed using the "skeletal" tree approach (*23*) implemented in the R package diversitree (http://www.zoology.ubc.ca/prog/diversitree/), which accounts for the sampling fraction of species in the given phylogeny out of the total number of species in that clade, using the null hypothesis that the unsampled species did not differ in proportion of diploid versus polyploid species from the sampled species. Results obtained using the complete sampling assumption were nearly identical (results not shown). In cases where diversity counts were unavailable (for example, when the focal group represented a subgenus without a specified species richness), we ran diversitree under the complete sampling assumption.

The Markov chain Monte Carlo (MCMC) approach described in FitzJohn et al. (*23*) was applied to estimate the probability distributions for each of the five parameters (λ_P, λ_D, μ_P, μ_D, q_{DP}), accounting for uncertainty in parameter estimation and incomplete sampling. In order to account for phylogenetic uncertainty, the analysis was conducted across a set of plausible Bayesian trees and was combined to form one sample. Specifically, exponential priors were placed on the five parameters, with the mean for each set to twice the rate needed to account for the growth of the clade since the most recent common ancestor (see *23*). Posterior distributions were estimated from 50 post burn-in trees sampled by MrBayes. For the first tree in the sample, the initial starting point was determined based on the heuristic starting point estimated by diversitree according to the state-independent birth-death model. The subsequent 49 trees were started from the last point sampled in the previous tree. Each tree was run for 2000 iterations and sampled every tenth step. The first 25% of the chain for each tree was regarded as burn-in and discarded from the analysis. The 50 chains (each corresponding to one tree sampled by MrBayes) were

then concatenated to form a single sample. We note that individual BiSSE MCMC chains converged rather quickly and that results were indistinguishable whether we used 10 or 50 MrBayes trees (we nonetheless used the larger sample).

To test whether estimated extinction and speciation rates differ between polyploids and diploids, we calculated the percentage of BiSSE MCMC steps in which the diploid rate was higher than that of polyploids (the posterior probability of diploids having a higher rate than polyploids). For example, to test whether extinction rates differ, we calculated the percentage of post-burnin steps in which $\mu_D > \mu_P$. To assess significance over the whole dataset of phylogenetic studies, we used a one-sample *t*-test with a mean equal to 50%, testing the null hypothesis that the diploid rate should be higher than the polyploid rate half of the time (the population being all 63 phylogenetic studies). Because our statistic is a proportion we used the probit transformation prior to performing the *t*-test. Table S2 lists the inferred speciation, extinction, and diversification rates for all datasets analyzed.

Extending the BiSSE model to calculate polyploid speciation frequency

The BiSSE analyses described above were based on the assumption that transitions between the two states under study (here, diploid and polyploid) are homogenous with respect to time and occur with equal probability at any point along the branches of the phylogeny. Moreover, in its current implementation, transitions between the two states are decoupled from speciation events: state change cannot occur simultaneously with speciation (*6, 23*). Thus, a speciating diploid lineage will always give rise to two diploid lineages (which may then polyploidize at a later point). These assumptions are potentially problematic when considering a trait that is associated with reproductive isolation, such as polyploidy. To allow state change to occur simultaneously with a speciation event, we derived the BiSSE-ness ("binary state speciation and extinction node-enhanced state shift") model. By doing so, we could also estimate the frequency of speciation events that involve polyploidization, as detailed below.

As originally formulated, BiSSE makes the assumption that no change in state occurs precisely at speciation. Therefore, to calculate the probability that a lineage just prior to node *A* is in state 0 and evolved as observed, BiSSE multiplies the probability that both daughter lineages (*N* and *M*) are in state 0 and evolved as observed by the rate of speciation in state 0, λ_0:

$$D_{A0}(t_A) = D_{N0}(t_A) D_{M0}(t_A) \lambda_0 \tag{S1}$$

(Equation 4a in *6*). An equivalent equation is obtained for character state 1. In this formulation, speciation and character state changes are treated as independent events. Thus, these equations include only the case where the character state is the same for the ancestor *A* and the two descendent lineages. To relax this assumption, which in the case of polyploidy may be particularly unrealistic, we incorporated the possibility that at speciation events the two daughter lineages can change state or not:

$$\begin{aligned} D_{A0}(t_A) = {} & D_{N0}(t_A) D_{M0}(t_A) \lambda_0 (1 - p_{0c}) && \text{no change at speciation} \\ & + \left(\frac{D_{N1}(t_A) D_{M0}(t_A)}{2} + \frac{D_{N0}(t_A) D_{M1}(t_A)}{2} \right) \lambda_0 p_{0c} p_{0a} && \text{one lineage changes} \\ & + D_{N1}(t_A) D_{M1}(t_A) \lambda_0 p_{0c} (1 - p_{0a}) && \text{both lineages change} \end{aligned} \tag{S2}$$

where p_{0c} is the probability that there is a change in character state associated with the speciation process and thus one or both lineages are in state 1 (opposite to that of the ancestral lineage), p_{0a} is the probability that given a change in character state has occurred during speciation this change is asymmetrical such that one lineage changes state and the other one remains in the same state (half of the time this is the *M* lineage and half of the time the *N* lineage), and $1–p_{0c}$ is the probability that both lineages remain in the same state.

Because not all speciation events result in two daughter lineages that survive to the present, similar modifications were made to allow simultaneous speciation with state change along the branches, given that one of the lineages resulting from such speciation events must go extinct before the present and the other lineage must give rise to the descendant species, as observed. The above modifications resulted in the addition of four parameters to form the BiSSE-ness model: p_{0c}, p_{0a}, p_{1c}, p_{1a} (here state 0 being diploid and state 1 polyploid). Here, however, we excluded polyploid to diploid transitions ($p_{1c} = 0$ and $p_{1a} = 0$) since in our generic phylogenies we defined polyploids as those lineages that underwent a polyploidization event since the divergence from their generic ancestor. Additionally p_{0a} was set to one since instant speciation via polyploidy entails one lineage becoming polyploid while the second one remaining in the diploid state, and q_{DP} was set to zero to restrict polyploid transitions to speciation events. The only new parameter is thus p_{0c}, denoting the frequency of speciation via polyploidy while being in the diploid state (termed heteroploid speciation in the main text), with $1–p_{0c}$ being the frequency of homoploid speciation in the diploid state. We note that because of the binary division of taxa into diploids and polyploids, so that tetraploids and hexaploids, for example, are considered in the same polyploid state, we could not differentiate between homoploid and heteroploid speciation among polyploids. Table S3 lists the inferred heteroploid speciation frequency as well as other diversification statistics inferred using BiSSE-ness for all datasets analyzed.

Supporting text

Diversification results allowing for polyploidy reversals

In the diversification analyses detailed in the main text we fixed the rate of polyploid-to-diploid reversals to zero. This irreversibility assumption was introduced for several important reasons. First, allowing reversibility in the BiSSE or BiSSE-ness models would imply that it is possible for a taxon to experience a near-instantaneous halving of its genome (i.e., polyhaploidy), which is not widely accepted. Instead, diploidization is thought to be a gradual process involving the loss and differentiation of genetic material. In addition, polyploid species were defined in our studies as those lineages that underwent a polyploidization event since divergence from the common ancestor of the group examined. Thus, a lineage may be labeled polyploid even if it has a chromosome number and meiotic behavior similar to that inferred for the base of the group (although, in practice, this was not observed). Defined in this way, polyploidy is truly irreversible. Allowing reversibility in BiSSE thus contradicts the ploidy definition that we employ and the ploidy assignments. For these reasons, we believe that constraining the transition rate q_{PD} to zero (no polyploid to diploid transitions) is justified.

Nevertheless, we also explored the possibility of allowing polyploid reversals to diploidy (i.e., without the constraint $q_{PD} = 0$). To ensure that neopolyploidy is still defined with respect to the group examined, the root state was fixed to the diploid state. Two analyses were then performed. The first allowed the transition rate from polyploidy to diploidy to be unconstrained. This often led to very high transition rates; both from diploidy to polyploidy and especially from polyploidy to diploidy, with q_{PD} being on average three times as large as q_{DP}. This typically reduced the inferred extinction rate of polyploids, reducing the signal of diversification differences between polyploids and diploids (no significant differences were observed between diploids and polyploids in their diversification rates, speciation rates, or extinction rates). Upon further investigation, the likelihood surface exhibited a ridge whereby the lack of proliferation of polyploid lineages could be explained by either high extinction or by high rates of reversion back to diploidy. Given that polyploid to diploid reversions are virtually unknown in the plant world, however, we have a strong prior expectation that reversion rates should not be high relative to the rates of polyploidization. We thus repeated our analyses constraining the transition rate from polyploidy to diploidy to be lower than the transition rate from diploidy to polyploidy. Under this model, net-diversification rates of neopolyploids were lower than that of diploids ($p < 10^{-5}$; *t*-test across the 63 trees). This was driven primarily by the higher extinction rates of polyploid lineages ($p < 10^{-9}$), while speciation rates were similar between the two ploidal states ($p > 0.1$). We thus conclude that only under unrealistically high reversion rates from polyploidy to diploidy would polyploids be likely to diversify at rates that are as high as that inferred for diploids.

Simulating the number of ancient polyploidization events within angiosperm species

We used a two-state (diploid and polyploid) birth and death process to simulate the distribution of the number of genome duplications expected in the evolutionary history of an angiosperm species since its divergence from the most recent common ancestor of angiosperms. In these simulations we assumed that variation in ploidy does not affect speciation and extinction rates. The ratio of extinction to speciation was set to 0.6, as estimated by Bokma (*24*) for a diverse set of angiosperm taxa. We then recorded the total number of polyploidization events that occurred in the history of each extant taxon, assuming that a certain fraction of speciation events, f_{het}, involve polyploidization. We ran this process 100 times, for a given f_{het} value, starting from a common diploid ancestor until the total number of species reached 300,000, a somewhat conservative estimate of the total number of angiosperm species (*25*); other estimates ranging from 250,000 to 400,000 gave similar results). Under the extent of heteroploid speciation estimated in our study ($f_{het} = 29.7\%$ for angiosperms), our simulations indicated that if polyploids and diploids were diversifying at equal rates we would find the traces of even more paleopolyploidization events than the 1–4 duplications observed in extant angiosperm species (>94% of the species having 5 or more; Fig. S2a). The distributions of the number of paleopolyploidization events

obtained assuming lower rates of heteroploid speciation were also shifted to the right (Fig S2b-c). In these simulations we did not aim to account for the complex dynamics of diversification rates through time, but rather to illustrate that under a simple birth-death model our finding that neopolyploids do not diversify as much as diploids is not inconsistent with the observation of multiple paleopolyploidy events among extant taxa.

Supporting figures

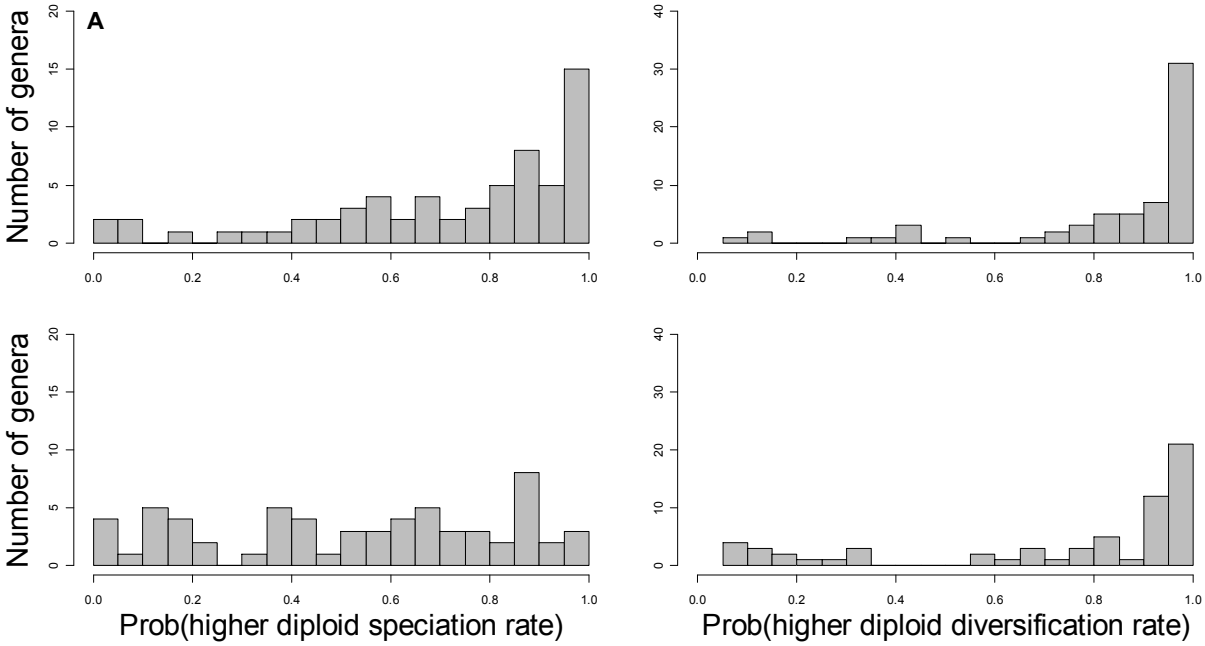

Figure S1|Homoploid diversification rates of diploid and polyploid taxa.
The histogram shows the posterior probability for each phylogeny that the speciation (A+C), and net diversification rate (B+D) of diploid lineages were higher than that of the polyploid lineages across the 63 plant groups studied (as in Figure 1, but here restricting polyploidization to speciation events). A value of 0.8 represents a phylogeny in which diploids exhibited a higher rate than polyploids in 80% of the MCMC steps analyzed. In A and B speciation and diversification rates were calculated based on both homoploid and heteroploid speciation, while in C and D rates include only homoploid speciation. For each dataset, MCMC BiSSE-ness results were pooled across 50 trees sampled by MrBayes.

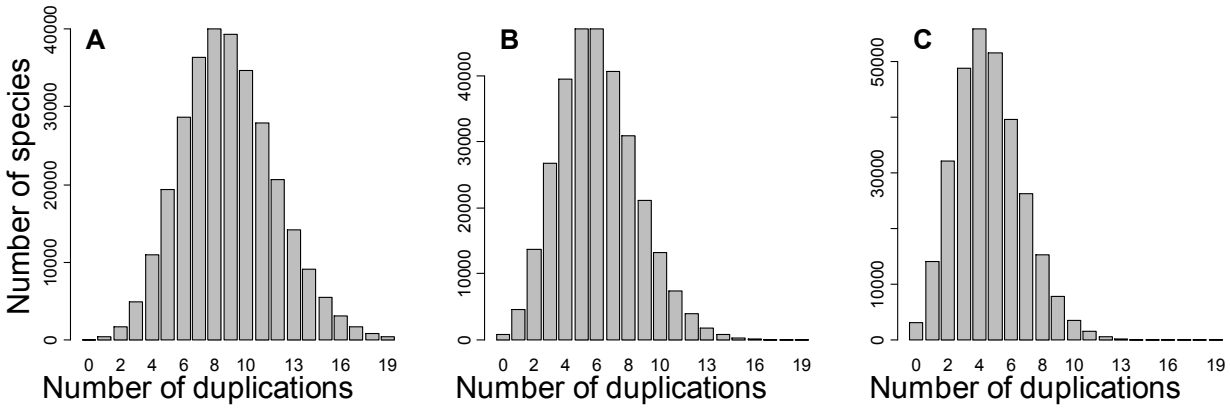

Figure S2|Simulated distribution of the total number of genome duplications that have occurred within extant angiosperm species. A birth and death process was used to simulate the distribution of the number of genome duplications assuming equal net diversification rates of diploids and polyploids and polyploidizations occurring at speciation events with a probability, f_{het}, of (A) 30% (the average estimate of our angiosperm datasets), (B) 20%, and (C) 15%, resulting in an average number of past polyploidization events of 8.9, 6.0, and 4.6, respectively. In all cases the number of angiosperm species was assumed to be 300,000 (*25*), with a ratio of extinction to speciation of 0.6 (*24*).

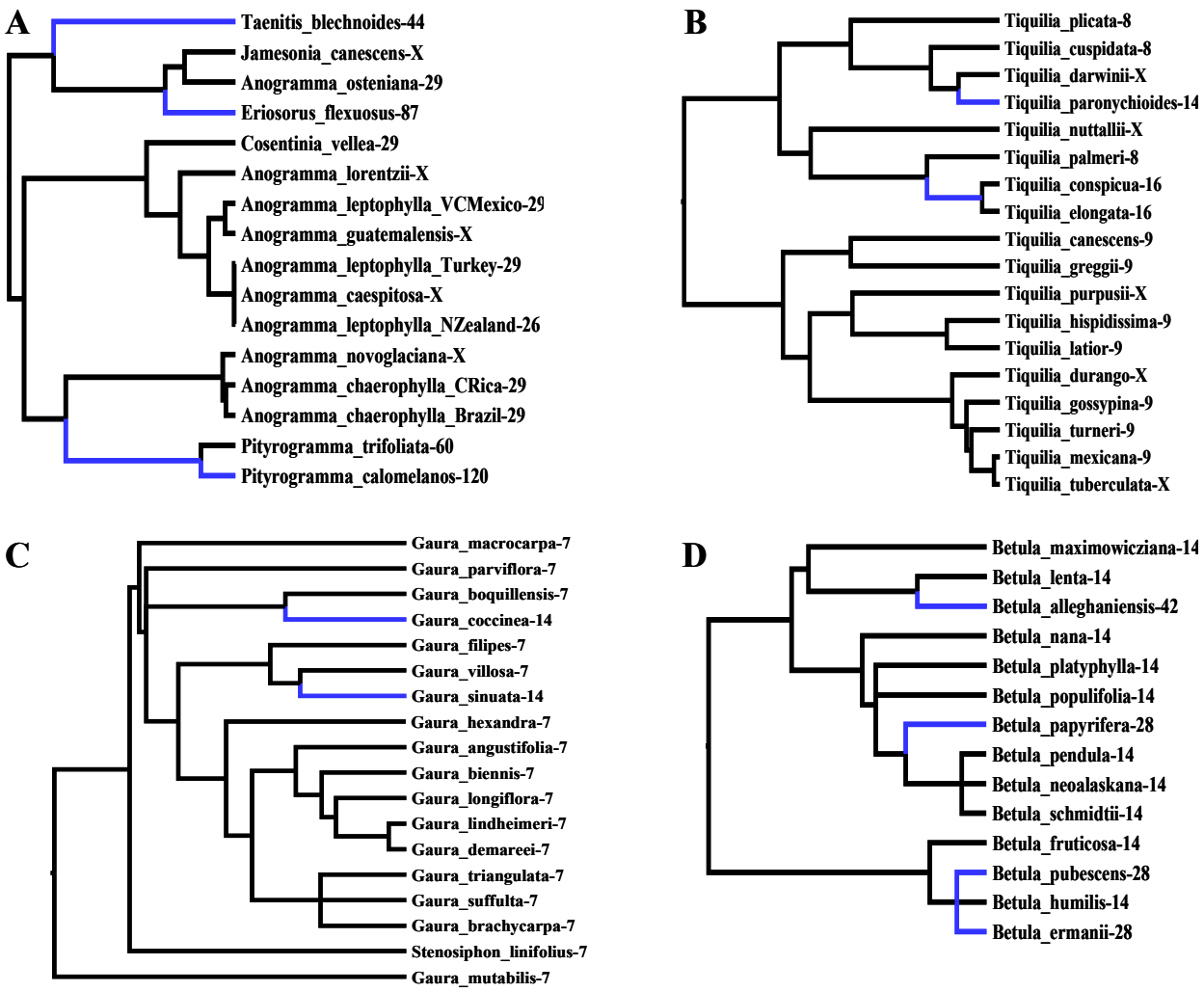

Figure S3| Ploidy transitions for several representative groups. (A) *Anogramma*, (B) *Tiquilia*, (C) *Gaura/Stenosiphon*, (D) *Betula*. Branches of the tree where polyploidization events were inferred to occur using the chromEvol methodology (*5*) are colored in blue. Chromosome counts appear to the right of the species name following a hyphen, where 'X' indicates unavailable cytological data.

Supporting tables

Table S1|Datasets used in this study. SS = number of species sampled with sequence data, SI = number of species recognized in the ingroup, SC = number of sampled species with cytological data, %PP = percentage of neopolyploids as estimated by chromEvol, TH = tree height in terms of average number of substitutions per site[a], LH = life history attributes[b], na = diversity counts not available.

Group	Focal Group	SS	SI	SC	%PP	TH[a]	LH[b]	Marker type[c]	Reference[d]
Eudicots	*Sium s.l.*	14	17	6	0.36	0.051	pr, hb	ITS	(*26*)
Eudicots	*Lathyrus*	52	160	37	0.02	0.033	pr/an, hb	ITS+cp	(*27*)
Eudicots	*Betula*	14	35	14	0.29	0.017	pr, wd	nr	(*28*)
Eudicots	*Tarasa s.l.*	36	na	22	0.33	0.033	pr/an, hb/wd	ITS	(*29*)
Eudicots	*Cuphea*	52	260	39	0.40	0.139	pr/an, hb/wd	ITS	(*30*)
Eudicots	*Fuchsia*	34	110	34	0.35	0.010	pr, wd	ITS+cp	(*31*)
Eudicots	*Gaura/Stenosiphon*	18	22	18	0.11	0.011	pr/an, hb	ITS+nr+cp	(*32*)
Eudicots	*Geum* + allies	23	na	20	0.78	0.025	pr, hb	ITS+cp	(*33*)
Eudicots	*Centaurium*	27	27	26	0.63	0.062	an, hb	ITS	(*34*)
Eudicots	*Primula* sect. *Aleuritia/ Armerina*	19	35	16	0.58	0.012	pr, hb	ITS	(*35*)
Eudicots	*Microseris*	16	16	16	0.38	0.020	pr/an, hb	ITS	(*36*)
Eudicots	*Senecio* sect. *Jacobaea*	26	26	19	0.88	0.026	pr/an, hb/wd	ITS+cp	(*37*)
Eudicots	*Campanula Rapunculus* clade	60	na	43	0.03	0.154	pr, hb	ITS	(*38*)
Eudicots	*Tiquilia* subg. *Tiquilia*	18	20	13	0.17	0.042	pr/an, hb/wd	ITS+nr	(*39*)
Eudicots	*Phacelia* subg. *Phacelia*	50	na	50	0.08	0.076	an/pr, hb	ITS	(*40*)
Eudicots	*Viburnum*	42	175	34	0.10	0.020	pr, wd	ITS+cp	(*41*)
Eudicots	*Actinidia*	35	62	30	0.20	0.008	pr, wd	cp	(*42*)
Eudicots	*Vaccinium* sect. *Macropelma/ Myrtillus/Hemimyrtillus*	50	na	21	0.18	0.017	pr, wd	ITS+cp	(*43*)
Eudicots	*Collomia*	10	15	7	0.30	0.008	an, hb	cp	(*44*)
Eudicots	*Dodecatheon/ Primula* subg. *Auriculastrum*	41	na	17	0.80	0.048	pr, hb	cp	(*45*)
Eudicots	*Achillea*	59	130	54	0.29	0.021	pr, hb/wd	ITS	(*46*)
Eudicots	*Erodium*	67	74	59	0.13	0.059	pr/an, hb	cp	(*47*)
Eudicots	*Pelargonium*	142	280	142	0.20	0.062	pr/an, hb/wd	ITS+cp+mt	(*48*)
Eudicots	*Houstonia*	15	25	15	0.53	0.192	pr/an, hb	ITS+cp	(*49*)
Eudicots	*Achimenes*	20	23	16	0.15	0.027	pr, hb	ITS+cp	(*50*)
Eudicots	*Mentha*	15	18	15	0.73	0.005	pr/an, hb	cp	(*51*)
Eudicots	*Orobanche* + allies	47	na	20	0.74	0.048	pr/an,	cp	(*52*) (*53*)

							hb		
Eudicots	*Penstemon*	132	271	90	0.08	0.015	pr/an, hb/wd	cp	(*54*)
Eudicots	*Antirrhinum* + allies	32	na	22	0.63	0.106	pr/an, hb	ITS	(*55*)
Eudicots	*Digitalis/Isoplexis*	23	23	17	0.96	0.019	pr/an, hb	ITS+cp	(*56*)
Eudicots	*Mimulus*	86	120	67	0.42	0.158	pr/an, hb/wd	ITS	(*57*)
Eudicots	*Physalis*	49	na	21	0.04	0.082	pr/an, hb	ITS	(*58*)
Eudicots	*Solanum* subg. *Leptostemonum*	131	350	62	0.02	0.082	pr/an, hb/wd	ITS+cp+nr	(*59*)
Eudicots	*Cerastium*	36	100	28	0.42	0.025	pr/an, hb	cp	(*60*)
Eudicots	*Silene* sect. *Physolychnis*	14	na	12	0.36	0.023	pr/an, hb	cp	(*61*)
Eudicots	*Gunnera*	20	30	9	0.55	0.208	pr, hb	ITS	(*62*)
Eudicots	*Aeonium/Greenovia/Monanthes*	52	63	32	0.08	0.082	pr/an, hb/wd	ITS	(*63*)
Eudicots	*Aichryson*	14	14	6	0.36	0.012	pr/an, hb	ITS+cp	(*64*)
Eudicots	*Graptopetalum* + allies	28	na	12	0.32	0.020	pr, hb/wd	ITS+cp+nr	(*65*)
Eudicots	*Coreopsis*	22	25	21	0.91	0.014	pr/an, hb/wd	ITS+cp	(*66*)
Eudicots	*Leavenworthia*	8	8	8	0.13	0.008	an, hb	cp	(*67*)
Eudicots	*Cucumis*	22	32	20	0.32	0.032	an, hb	ITS	(*68*) (*69*)
Magnoliids	*Aristolochia s.l.*	78	400	62	0.31	0.037	pr, hb/wd	cp	(*70*)
Monocots	*Arisaema*	75	150	27	0.08	0.007	pr, hb	cp	(*71*)
Monocots	*Lemna/Wolffia/ Wolffiella/Spirodella/Landoltia*	38	38	21	0.84	0.038	an, hb	cp	(*72*)
Monocots	*Veratrum s.l.*	26	na	18	0.35	0.029	pr, hb	ITS	(*73*)
Monocots	*Gagea/Lloydia*	58	82	27	0.41	0.063	pr, hb	ITS	(*74*)
Monocots	*Trillium s.l. /Paris s.l.*	25	69	25	0.08	0.018	pr, hb	ITS+cp	(*75*)
Monocots	*Sorghum*	16	25	15	0.50	0.012	pr/an, hb	ITS+cp	(*76*)
Lycophytes	*Isoëtes* ("North American clade")	36	43	39	0.19	0.091	pr, wd	ITS+cp	(*77*) (*78-80*)
Ferns	*Asplenium* subg. *Ceterach* + allies	37	na	23	0.70	0.049	pr, hb	cp	(*81*) (*82*)
Ferns	*Asplenium* (New Zealand *australe* group)	18	na	23	0.56	0.022	pr, hb	ITS+cp	(*83*)
Ferns	*Dryopteris* (Hawaii)	55	na	44	0.51	0.020	pr, hb	cp	(*84*) (*85*)
Ferns	*Anogramma* + allies	16	na	11	0.25	0.056	pr, hb	cp	(*86*) (*87, 88*)
Ferns	*Argyrochosma*	18	20	18	0.33	0.021	pr, hb	cp	(*89*) (*90*)
Ferns	*Cyrtomium* + allies	27	na	20	0.44	0.025	pr, hb	cp	(*91*)
Ferns	*Dryopteris* (China)	62	230	41	0.29	0.079	pr, hb	cp	(*85*)
Ferns	*Dryopteris* (North America)	12	na	12	0.50	0.021	pr, hb	cp	(*92*)
Ferns	*Hymenophyllum*	80	250	51	0.83	0.025	pr, hb	cp	(*93*)
Ferns	*Lygodium*	15	30	10	0.33	0.082	pr, hb	cp	(*94*) (*95*)

Ferns	*Cheilanthes* (*Myriopteris* clade)	37	39	28	0.38	0.031	pr, hb	cp	(*92*) (*96-98*)
Ferns	*Notholaena*	21	38	19	0.14	0.032	pr, hb	cp	(*99*) (*100*)
Ferns	*Pellaea*	18	26	10	0.17	0.035	pr, hb	cp	(*101*) (*98, 102, 103*)

[a]Tree height was calculate based on the ML tree obtained using phyML by traversing the tree from the tips to the root. For each internal node, its height was calculated as the average height of the lineages descending from that node, finally reaching the root node.

[b]Life history characteristics are given based on the majority of species in the group as follows: pr = perennials; an = annuals or biennials; hb = herbaceous; wd = woody (trees/shrubs). Data were obtained through eflora (http://www.efloras.org/), Mabberley (*17*), or the original phylogenetic study.

[c]nr: nuclear; cp: chloroplast; mt: mitochondrial; ITS: the nuclear internal transcribed spacer

[d]Diversity and cytology references are given to the right if different from the main reference or from Mabberley (*17*)

Table S2|BiSSE diversification analysis. $\overline{\lambda}_D$ and $\overline{\lambda}_P$ = average inferred speciation rate of diploids and polyploids over the MCMC BiSSE sample, respectively; $\overline{\mu}_D$ and $\overline{\mu}_P$ = average extinction rate of diploids and polyploids, respectively; %($r_D > r_P$) = the percentage of MCMC BiSSE steps in which the diversification rate of diploid lineages were higher than that of the polyploid lineages; %($\lambda_D > \lambda_P$) and %($\mu_D > \mu_P$) are the percentage of MCMC BiSSE steps in which the speciation and extinction rates of diploid lineages were higher than that of the polyploid lineages, respectively.

Focal Genus	$\overline{\lambda}_D$	$\overline{\lambda}_P$	$\overline{\mu}_D$	$\overline{\mu}_P$	**%($r_D > r_P$)**	**%($\lambda_D > \lambda_P$)**	**%($\mu_D > \mu_P$)**
Sium s.l.	69.29	58.91	47.98	40.9	0.58	0.54	0.5
Lathyrus	253.08	148.5	128.33	586.22	0.81	0.08	0.98
Betula	293.58	171.61	118.93	442.12	0.79	0.1	0.95
Tarasa s.l.	138.64	142.19	34.07	179.75	0.52	0.08	0.9
Cuphea	47.15	45.6	14.8	26.19	0.56	0.32	0.84
Fuchsia	302.65	448.59	73.17	233.75	0.22	0.21	0.55
Gaura/Stenosiphon	236.23	207.85	79.79	548.33	0.66	0.07	0.93
Geum + allies	129.74	59.41	57.68	43.47	0.94	0.57	0.87
Centaurium	111.88	97.37	57.76	189.51	0.61	0.02	1
Primula sect. *Aleuritia/Armerina*	725.99	463.22	302.35	864.45	0.78	0.1	0.99
Microseris	257.78	197.19	117.62	489.44	0.7	0.07	0.96
Senecio sect. *Jacobaea*	311	212.94	151.14	240.03	0.73	0.27	0.88
Campanula: Rapunculus clade	44.49	48.03	11.17	187.25	0.6	0.01	0.96
Tiquilia subg. *Tiquilia*	89.67	132.23	35.15	208.06	0.38	0.08	0.82
Phacelia subg. *Phacelia*	42.63	43.81	7.9	146.56	0.6	0.01	0.96
Viburnum	268.44	93	127.06	386.87	0.97	0.57	1
Actinidia	1574.39	507.8	755.33	2851.16	0.93	0.05	1
Vaccinium sect. *Macropelma/ Myrtillus/Hemimyrtillus*	118.85	59.01	20.38	57.23	0.9	0.26	0.97
Collomia	642.53	776.59	369.14	1242.46	0.48	0.14	0.81
Dodecatheon/ Primula subg. *Auriculastrum*	96.41	208.86	42.18	121.38	0.04	0.19	0.33
Achillea	444.56	1139.26	104.86	1516.58	0.07	0.02	0.85
Erodium	300.8	153.16	205.06	478.27	0.88	0.1	1
Pelargonium	123.62	127.42	27.32	175.51	0.52	0	1
Houstonia	31.74	13.72	23.11	20.27	0.9	0.55	0.84
Achimenes	151.7	124.39	42.8	321.9	0.7	0.08	0.95
Mentha	605.34	708.16	363.28	435.63	0.35	0.46	0.41
Orobanche + allies	85.97	55.74	32.01	59.01	0.83	0.25	0.91
Penstemon	448.61	456.15	87.83	685.47	0.52	0.02	0.97
Antirrhinum + allies	18.97	50.8	8.98	17.39	0.01	0.33	0.07
Digitalis/Isoplexis	281.76	245.59	200.3	179.23	0.53	0.48	0.53
Mimulus	60.12	107.26	27.93	43.88	0.04	0.36	0.08
Physalis	49.43	20.7	10.8	50.65	0.89	0.23	0.98
Solanum subg. *Leptostemonum*	96.95	82.84	24.78	98.08	0.67	0.18	0.94
Cerastium	748.66	410.51	278.33	1275.62	0.88	0.03	1
Silene sect. *Physolychnis*	184.25	326.24	86.48	467.6	0.27	0.07	0.78
Gunnera	67.1	53.91	45.23	87.34	0.67	0.15	0.99
Aeonium/Greenovia/Monanthes	233.87	64.12	171.81	383.09	0.96	0.13	1
Aichryson	240.39	92.95	102.83	345.12	0.89	0.14	0.98
Graptopetalum + allies	5.32	2.74	1.48	2.84	0.9	0.29	0.95
Coreopsis	451.9	160.37	259.7	111.46	0.96	0.73	0.77

Aristolochia s.l.	145.93	274.68	43.64	122.26	0.02	0.2	0.24
Leavenworthia	382.52	326.22	317.2	956.36	0.65	0.15	0.87
Cucumis	128.85	43.72	51.45	215.38	0.93	0.04	1
Arisaema	590.6	439.21	189.77	962.84	0.78	0.09	0.99
Lemna/Wolffia/ Wolffiella/Spirodella/Landoltia	90.04	103.84	54.58	47.32	0.33	0.5	0.29
Veratrum s.l.	200.19	225.62	158.5	133.9	0.41	0.59	0.29
Gagea/Lloydia	42.89	27.86	24.2	25.81	0.83	0.47	0.88
Trillium s.l./Paris s.l.	372.81	173.27	194.23	360.85	0.86	0.35	0.95
Sorghum	204.56	244.22	106.89	232.05	0.42	0.22	0.78
Isoëtes ("North American clade")	202.34	32.8	136.2	182.65	0.99	0.37	1
Asplenium subg. *Ceterach* + allies	97.76	93.32	30.35	79	0.55	0.22	0.81
Asplenium (New Zealand *australe* group)	283.59	63.62	145.62	370.58	0.98	0.08	1
Dryopteris (Hawaii)	311.86	190.6	76.62	479.91	0.88	0	1
Anogramma + allies	94.84	37.62	67.12	69.54	0.9	0.51	0.88
Argyrochosma	177.96	236.47	57.3	401.97	0.41	0.05	0.86
Cyrtomium + allies	237.5	314.01	77.85	620.52	0.35	0	0.99
Dryopteris (China)	395.57	236.16	170.91	311.75	0.9	0.22	1
Dryopteris (North America)	147.2	182.92	77.26	334.64	0.43	0.07	0.9
Hymenophyllum	239.01	151.4	89.26	22.86	0.93	0.82	0.68
Lygodium	69.49	46.9	33.35	80.85	0.75	0.2	0.95
Cheilanthes (*Myriopteris* clade)	168.33	108.98	37.27	341.34	0.79	0.01	1
Notholaena	99.75	60.81	25.72	127.05	0.79	0.13	0.95
Pellaea	16.93	3.89	11.36	12.7	0.97	0.48	0.99

Table S3|BiSSE-ness diversification analysis. %HS = average heteroploid speciation frequency inferred over the MCMC BiSSE sample; $\%(r_D > r_P)$ = the percentage of MCMC BiSSE steps in which the diversification rate of diploid lineages was higher than that of the polyploid lineages; $\%(\lambda_D > \lambda_P)$ and $\%(\mu_D > \mu_P)$ are the percentage of MCMC BiSSE steps in which the speciation and extinction rate of diploid lineages were higher than that of the polyploid lineages, respectively; $\%(r_D > r_P)_h$ and $\%(\lambda_D > \lambda_P)_h$ are the percentage of steps in which the diversification and speciation rate of diploid lineages were higher than that of the polyploid lineages, respectively, accounting for homoploid speciation only.

Focal Group	%HS	$\%(r_D > r_P)$	$\%(\lambda_D > \lambda_P)$	$\%(\mu_D > \mu_P)$	$\%(r_D > r_P)_h$	$\%(\lambda_D > \lambda_P)_h$
Sium s.l.	0.34	0.43	0.58	0.65	0.22	0.34
Lathyrus	0.07	0.96	0.8	0.09	0.96	0.77
Betula	0.35	0.95	0.81	0.1	0.93	0.66
Tarasa s.l.	0.34	0.93	0.77	0.16	0.82	0.44
Cuphea	0.18	0.87	0.62	0.3	0.72	0.42
Fuchsia	0.14	0.68	0.3	0.21	0.55	0.21
Gaura/Stenosiphon	0.22	0.94	0.76	0.1	0.92	0.67
Geum + allies	0.5	0.82	0.95	0.7	0.32	0.55
Centaurium	0.61	1	0.84	0.01	0.98	0.18
Primula sect. *Aleuritia/ Armerina*	0.56	0.99	0.89	0.09	0.95	0.48
Microseris	0.42	0.96	0.85	0.12	0.91	0.61
Senecio sect. *Jacobaea*	0.69	0.91	0.87	0.26	0.34	0.12
Campanula: Rapunculus clade	0.16	0.97	0.66	0.01	0.97	0.59
Tiquilia subg. *Tiquilia*	0.23	0.84	0.49	0.08	0.81	0.36
Phacelia subg. *Phacelia*	0.19	0.96	0.7	0.02	0.96	0.62
Viburnum	0.14	1	0.97	0.63	1	0.95
Actinidia	0.33	1	0.97	0.08	1	0.87
Vaccinium sect. *Macropelma/ Myrtillus/Hemimyrtillus*	0.24	0.98	0.95	0.39	0.91	0.85
Collomia	0.33	0.82	0.54	0.12	0.76	0.35
Dodecatheon/ Primula subg. *Auriculastrum*	0.4	0.44	0.08	0.15	0.2	0.01
Achillea	0.3	0.79	0.06	0.01	0.76	0.01
Erodium	0.13	1	0.94	0.24	1	0.89
Pelargonium	0.16	0.99	0.7	0.04	0.97	0.51
Houstonia	0.33	0.83	0.87	0.53	0.66	0.67
Achimenes	0.24	0.97	0.81	0.13	0.94	0.71
Mentha	0.41	0.43	0.35	0.42	0.19	0.13
Orobanche + allies	0.38	0.91	0.96	0.52	0.6	0.62
Penstemon	0.09	0.97	0.65	0.04	0.96	0.6
Antirrhinum + allies	0.22	0.13	0.01	0.31	0.07	0
Digitalis/Isoplexis	0.65	0.52	0.59	0.55	0.07	0.13
Mimulus	0.06	0.1	0.02	0.32	0.07	0.02
Physalis	0.12	0.98	0.91	0.35	0.97	0.88
Solanum subg. *Leptostemonum*	0.03	0.95	0.7	0.21	0.95	0.68
Cerastium	0.58	1	0.88	0	1	0.55
Silene sect. *Physolychnis*	0.36	0.73	0.34	0.09	0.67	0.16
Gunnera	0.46	0.98	0.74	0.12	0.92	0.37
Aeonium/Greenovia/Monanthes	0.16	1	0.96	0.14	1	0.92
Aichryson	0.41	0.98	0.98	0.28	0.94	0.89
Graptopetalum + allies	0.31	0.96	0.98	0.54	0.85	0.88

Coreopsis	0.6	0.73	0.94	0.81	0.15	0.43
Aristolochia s.l.	0.01	0.14	0.16	0.42	0.13	0.15
Leavenworthia	0.29	0.87	0.64	0.1	0.85	0.52
Cucumis	0.44	1	0.96	0.06	0.99	0.84
Arisaema	0.19	0.99	0.86	0.17	0.98	0.77
Lemna/Wolffia/ Wolffiella/Spirodella/Landoltia	0.5	0.4	0.43	0.5	0.05	0.1
Veratrum s.l.	0.06	0.3	0.43	0.61	0.26	0.39
Gagea/Lloydia	0.2	0.86	0.88	0.59	0.68	0.71
Trillium s.l./Paris s.l.	0.11	0.94	0.89	0.39	0.93	0.85
Sorghum	0.29	0.8	0.59	0.3	0.63	0.36
Isoëtes (“North American clade”)	0.17	1	1	0.73	1	1
Asplenium subg. *Ceterach* + allies	0.45	0.75	0.57	0.26	0.35	0.11
Asplenium (New Zealand *australe* group)	0.65	1	0.99	0.05	1	0.78
Dryopteris (Hawaii)	0.59	1	0.99	0.01	1	0.44
Anogramma + allies	0.16	0.86	0.96	0.72	0.78	0.92
Argyrochosma	0.4	0.88	0.5	0.05	0.82	0.23
Cyrtomium + allies	0.57	0.96	0.45	0.01	0.91	0.06
Dryopteris (China)	0.29	1	0.97	0.34	1	0.87
Dryopteris (North America)	0.53	0.91	0.55	0.06	0.81	0.19
Hymenophyllum	0.32	0.77	0.97	0.84	0.13	0.6
Lygodium	0.3	0.95	0.84	0.28	0.9	0.69
Cheilanthes (*Myriopteris* clade)	0.5	1	0.97	0.02	1	0.71
Notholaena	0.2	0.97	0.87	0.19	0.96	0.8
Pellaea	0.27	0.98	0.98	0.58	0.96	0.95

References and Notes

1. G. Stebbins, *Chromosomal Evolution in Higher Plants* (Edward Arnold, London, 1971).

2. Y. Jiao *et al*., Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97 (2011). doi:10.1038/nature09916 Medline

3. T. E. Wood *et al*., The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13875 (2009). doi:10.1073/pnas.0811575106 Medline

4. Materials and methods are available as supporting material on *Science* Online.

5. I. Mayrose, M. S. Barker, S. P. Otto, Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* **59**, 132 (2010). doi:10.1093/sysbio/syp083 Medline

6. W. P. Maddison, P. E. Midford, S. P. Otto, Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* **56**, 701 (2007). doi:10.1080/10635150701607033 Medline

7. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, E. W. Sayers, GenBank. *Nucleic Acids Res.* **39**, D32 (2011). doi:10.1093/nar/gkq1079 Medline

8. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792 (2004). doi:10.1093/nar/gkh340 Medline

9. S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696 (2003). doi:10.1080/10635150390235520 Medline

10. M. J. Sanderson, Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol. Biol. Evol.* **19**, 101 (2002). Medline

11. M. J. Sanderson, r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301 (2003). doi:10.1093/bioinformatics/19.2.301 Medline

12. F. Ronquist, J. P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572 (2003). doi:10.1093/bioinformatics/btg180 Medline

13. J. L. Thorne, H. Kishino, I. S. Painter, Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647 (1998). Medline

14. P. Goldblatt, D. E. Johnson, Eds., *Index to Plant Chromosome Numbers* (Missouri Botanical Garden, St. Louis, MO, 1978).

15. A. Federov, *Chromosome Numbers of Flowering Plants* (Academic Press, Leningrad, 1969).

16. Á. Löve, D. Löve, R. E. G. Pichi Sermolli, *Cytotaxonomical Atlas of the Pteridophyta* (Strauss and Cramer, Vaduz, Liechtenstein, 1977).

17. D. J. Mabberley, *Mabberley's Plant-Book: A Portable Dictionary of Plants, Their Classification and Uses* (Cambridge Univ. Press, Cambridge, ed. 3, 2008).

18. S. V. Muse, Examining rates and patterns of nucleotide substitution in plants. *Plant Mol. Biol.* **42**, 25 (2000). doi:10.1023/A:1006319803002 Medline

19. P. G. Wolf *et al*., The evolution of chloroplast genes and genomes in ferns. *Plant Mol. Biol.* **76**, 251 (2011). doi:10.1007/s11103-010-9706-4 Medline

20. S. A. Smith, M. J. Donoghue, Rates of molecular evolution are linked to life history in flowering plants. *Science* **322**, 86 (2008). doi:10.1126/science.1163197 Medline

21. H. Akaike, A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **119**, 716 (1974).

22. L. A. Meyers, D. A. Levin, On the abundance of polyploids in flowering plants. *Evolution* **60**, 1198 (2006). Medline

23. R. G. FitzJohn, W. P. Maddison, S. P. Otto, Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* **58**, 595 (2009). doi:10.1093/sysbio/syp067 Medline

24. F. Bokma, Testing for equal rates of cladogenesis in diverse taxa. *Evolution* **57**, 2469 (2003). Medline

25. G. T. Prance, H. Beentje, J. Dransfield, R. Johns, The tropical flora remains undercollected. *Ann. Mo. Bot. Gard.* **87**, 67 (2000). doi:10.2307/2666209

26. K. Spalik, S. R. Downie, The evolutionary history of *Sium* sensu lato (Apiaceae): Dispersal, vicariance, and domestication as inferred from ITS rDNA phylogeny. *Am. J. Bot.* **93**, 747 (2006). doi:10.3732/ajb.93.5.747 Medline

27. G. J. Kenicer, T. Kajita, R. T. Pennington, J. Murata, Systematics and biogeography of *Lathyrus* (Leguminosae) based on internal transcribed spacer and cpDNA sequence data. *Am. J. Bot.* **92**, 1199 (2005). doi:10.3732/ajb.92.7.1199 Medline

28. P. Järvinen *et al*., Phylogenetic relationships of *Betula* species (Betulaceae) based on nuclear ADH and chloroplast matK sequences. *Am. J. Bot.* **91**, 1834 (2004). doi:10.3732/ajb.91.11.1834 Medline

29. J. A. Tate, B. B. Simpson, *Syst. Bot.* **28**, 723 (2003).

30. S. A. Graham, J. Freudenstein, M. Luker, A phylogenetic study of *Cuphea* (Lythraceae) based on morphology and nuclear rDNA ITS sequences. *Syst. Bot.* **31**, 764 (2006). doi:10.1600/036364406779696004

31. P. E. Berry, W. J. Hahn, K. J. Sytsma, J. C. Hall, A. Mast, Phylogenetic relationships and biogeography of *Fuchsia* (Onagraceae) based on noncoding nuclear and chloroplast DNA data. *Am. J. Bot.* **91**, 601 (2004). doi:10.3732/ajb.91.4.601 Medline

32. G. D. Hoggard, P. J. Kores, M. Molvray, R. K. Hoggard, The phylogeny of *Gaura* (Onagraceae) based on ITS, ETS, and trnL-F sequence data. *Am. J. Bot.* **91**, 139 (2004). doi:10.3732/ajb.91.1.139 Medline

33. J. Smedmark, T. Eriksson, *Syst. Bot.* **27**, 303 (2002).

34. G. Mansion, L. Zeltner, F. Bretagnolle, Phylogenetic patterns and polyploid evolution within the Mediterranean genus *Centaurium* (Gentianaceae - Chironieae). *Taxon* **54**, 931 (2005). doi:10.2307/25065479

35. A. Guggisberg, G. Mansion, E. Conti, Disentangling reticulate evolution in an arctic-alpine polyploid complex. *Syst. Biol.* **58**, 55 (2009). doi:10.1093/sysbio/syp010 Medline

36. U. Lohwasser, A. Granda, F. R. Blattner, Phylogenetic analysis of *Microseris* (Asteraceae), including a newly discovered Andean population from Peru. *Syst. Bot.* **29**, 774 (2004). doi:10.1600/0363644041744446

37. P. B. Pelser, K. Van Den Hof, B. Gravendeel, R. Van Der Meijden, The systematic value of morphological characters in *Senecio* sect. *Jacobaea* (Asteraceae) as compared to DNA sequences. *Syst. Bot.* **29**, 790 (2004). doi:10.1600/0363644041744482

38. J. M. Park, S. Kovacic, Z. Liber, W. M. M. Eddie, G. M. Schneeweiss, Phylogeny and biogeography of isophyllous species of *Campanula* (Campanulaceae) in the Mediterranean area. *Syst. Bot.* **31**, 862 (2006). doi:10.1600/036364406779695924

39. M. J. Moore, A. Tye, R. K. Jansen, Patterns of long-distance dispersal in *Tiquilia* subg. *Tiquilia* (Boraginaceae): Implications for the origins of amphitropical disjuncts and Galapagos Islands endemics. *Am. J. Bot.* **93**, 1163 (2006). doi:10.3732/ajb.93.8.1163 Medline

40. C. Gilbert, J. Dempcy, C. Ganong, R. Patterson, G. S. Spicer, Phylogenetic relationships within *Phacelia* subgenus *Phacelia* (Hydrophyllaceae) inferred from nuclear rDNA ITS sequence data. *Syst. Bot.* **30**, 627 (2005). doi:10.1600/0363644054782251

41. M. J. Donoghue, B. G. Baldwin, J. Li, R. C. Winkworth, *Viburnum* phylogeny based on chloroplast *trnK* intron and nuclear ribosomal ITS DNA sequences. *Syst. Bot.* **29**, 188 (2004). doi:10.1600/036364404772974095

42. J. Chat, B. Jáuregui, R. J. Petit, S. Nadot, Reticulate evolution in kiwifruit (Actinidia, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *Am. J. Bot.* **91**, 736 (2004). doi:10.3732/ajb.91.5.736 Medline

43. E. A. Powell, K. Kron, *Syst. Bot.* **27**, 768 (2002).

44. L. A. Johnson, R. L. Johnson, Morphological delimitation and molecular evidence for allopolyploidy in *Collomia wilkenii* (Polemoniaceae), a new species from northern Nevada. *Syst. Bot.* **31**, 349 (2006). doi:10.1600/036364406777585865

45. A. R. Mast, D. M. Feller, S. Kelso, E. Conti, Buzz-pollinated *Dodecatheo*n originated from within the heterostylous *Primula* subgenus *Auriculastrum* (Primulaceae): A seven-region cpDNA phylogeny and its implications for floral evolution. *Am. J. Bot.* **91**, 926 (2004). doi:10.3732/ajb.91.6.926 Medline

46. Y. P. Guo, F. Ehrendorfer, R. Samuel, Phylogeny and systematics of *Achillea* (Asteraceae-Anthemideae) inferred from nrITS and plastid trnL-F DNA sequences. *Taxon* **53**, 657 (2004). doi:10.2307/4135441

47. O. Fiz, P. Vargas, M. L. Alarcón, J. Aldasoro, Phylogenetic relationships and evolution in *Erodium* (Geraniaceae) based on *trnL-trnF* sequences. *Syst. Bot.* **31**, 739 (2006). doi:10.1600/036364406779695906

48. F. T. Bakker, A. Culham, P. Hettiarachi, T. Touloumenidou, M. Gibby, Phylogeny of *Pelargonium* (Geraniaceae) based on DNA sequences from three genomes. *Taxon* **53**, 17 (2004). doi:10.2307/4135485

49. S. A. Church, D. R. Taylor, Speciation and hybridization among *Houstonia* (Rubiaceae) species: The influence of polyploidy on reticulate evolution. *Am. J. Bot.* **92**, 1372 (2005). doi:10.3732/ajb.92.8.1372 Medline

50. E. H. Roalson, L. E. Skog, E. A. Zimmer, *Syst. Bot.* **28**, 593 (2003).

51. J. Bunsawat, N. E. Elliott, K. L. Hertweck, E. Sproles, L. A. Alice, Phylogenetics of *Mentha* (Lamiaceae): Evidence from chloroplast DNA sequences. *Syst. Bot.* **29**, 959 (2004). doi:10.1600/0363644042450973

52. J.-F. Manen, C. Habashi, D. Jeanmonod, J. M. Park, G. M. Schneeweiss, Phylogeny and intraspecific variability of holoparasitic *Orobanche* (Orobanchaceae) inferred from plastid rbcL sequences. *Mol. Phylogenet. Evol.* **33**, 482 (2004). doi:10.1016/j.ympev.2004.06.010 Medline

53. H. Uhlich, J. Pusch, K.-J. Barthel, *Die SommerwurzartenEuropas, Gattung Orobanche* (Westarp Wissenschaften, Magdeburg, Germany, 1995).

54. A. D. Wolfe *et al*., Phylogeny, taxonomic affinities, and biogeography of *Penstemon* (Plantaginaceae) based on ITS and cpDNA sequence data. *Am. J. Bot.* **93**, 1699 (2006). doi:10.3732/ajb.93.11.1699 Medline

55. R. K. Oyama, D. A. Baum, Phylogenetic relationships of North American *Antirrhinum* (Veronicaceae). *Am. J. Bot.* **91**, 918 (2004). doi:10.3732/ajb.91.6.918 Medline

56. C. Bräuchler, H. Meimberg, G. Heubl, Molecular phylogeny of the genera *Digitalis L.* and *Isoplexis* (Lindley) *Loudon* (Veronicaceae) based on ITS- and trnL-F sequences. *Plant Syst. Evol.* **248**, 111 (2004). doi:10.1007/s00606-004-0145-z

57. P. M. Beardsley, S. E. Schoenig, J. B. Whittall, R. G. Olmstead, Patterns of evolution in western North American *Mimulus* (Phrymaceae). *Am. J. Bot.* **91**, 474 (2004). doi:10.3732/ajb.91.3.474 Medline

58. M. Whitson, P. Manos, Untangling *Physalis* (Solanaceae) from the Physaloids: A two-gene phylogeny of the Physalinae. *Syst. Bot.* **30**, 216 (2005). doi:10.1600/0363644053661841

59. R. Levin, N. R. Myers, L. Bohs, Phylogenetic relationships among the "spiny solanums" (*Solanum* subgenus *Leptostemonum*, Solanaceae). *Am. J. Bot.* **93**, 157 (2006). doi:10.3732/ajb.93.1.157

60. A. C. Scheen *et al*., Northern hemisphere biogeography of *Cerastium* (Caryophyllaceae): Insights from phylogenetic analysis of noncoding plastidnucleotide sequences. *Am. J. Bot.* **91**, 943 (2004). doi:10.3732/ajb.91.6.943 Medline

61. M. Popp, P. Erixon, F. Eggens, B. Oxelman, Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. *Syst. Bot.* **30**, 302 (2005). doi:10.1600/0363644054223648

62. L. Wanntorp, H. E. Wanntorp, M. Källersjö, *Syst. Bot.* **27**, 512 (2002).

63. M. Mort, D. E. Soltis, P. S. Soltis, J. Francisco-Ortega, A. Santos-Guerra, *Syst. Bot.* **27**, 271 (2002).

64. K. N. Fairfield, M. E. Mort, A. Santos-Guerra, Phylogenetics and evolution of the Macaronesian members of the genus *Aichryson* (Crassulaceae) inferred from nuclear and chloroplast sequence data. *Plant Syst. Evol.* **248**, 71 (2004). doi:10.1007/s00606-004-0190-7

65. R. Acevedo-Rosas, K. Cameron, V. Sosa, S. Pell, A molecular phylogenetic study of *Graptopetalum* (Crassulaceae) based on ETS, ITS, RPL16, and TRNL-F nucleotide sequences. *Am. J. Bot.* **91**, 1099 (2004). doi:10.3732/ajb.91.7.1099 Medline

66. D. J. Crawford, M. E. Mort, Phylogeny of Eastern North American *Coreopsis* (Asteraceae-Coreopsideae): Insights from nuclear and plastid sequences, and comments on character evolution. *Am. J. Bot.* **92**, 330 (2005). doi:10.3732/ajb.92.2.330 Medline

67. J. B. Beck, I. A. Al-Shehbaz, B. A. Schaal, *Leavenworthia* (Brassicaceae) revisited: Testing classic systematic and mating system hypotheses. *Syst. Bot.* **31**, 151 (2006). doi:10.1600/036364406775971732

68. J. Garcia-Mas, A. J. Monforte, P. Arús, Phylogenetic relationships among *Cucumis* species based on the ribosomal internal transcribed spacer sequence and microsatellite markers. *Plant Syst. Evol.* **248**, 191 (2004). doi:10.1007/s00606-004-0170-y

69. S. S. Renner, H. Schaefer, A. Kocyan, Phylogenetics of *Cucumis* (Cucurbitaceae): Cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). *BMC Evol. Biol.* **7**, 58 (2007). doi:10.1186/1471-2148-7-58 Medline

70. T. Ohi-Toma *et al*., Molecular phylogeny of *Aristolochia* sensu lato (Aristolochiaceae) based on sequences of *rbcL*, *matK*, and *phyA* genes, with special reference to differentiation of chromosome numbers. *Syst. Bot.* **31**, 481 (2006). doi:10.1600/036364406778388656

71. S. S. Renner, L.-B. Zhang, J. Murata, A chloroplast phylogeny of *Arisaema* (Araceae) illustrates Tertiary floristic links between Asia, North America, and East Africa. *Am. J. Bot.* **91**, 881 (2004). doi:10.3732/ajb.91.6.881 Medline

72. D. H. Les, D. Crawford, E. Landolt, J. D. Gabel, R. Kimball, *Syst. Bot.* **27**, 221 (2002).

73. W. B. Zomlefer, W. M. Whitten, N. H. Williams, W. S. Judd, *Syst. Bot.* **28**, 250 (2003).

74. M. Zarrei, S. Zarre, P. Wilkin, M. Rix, Systematic revision of the genus *Gagea* Salisb. (Liliaceae) in Iran. *Bot. J. Linn. Soc.* **154**, 559 (2007). doi:10.1111/j.1095-8339.2007.00678.x

75. S. B. Farmer, E. E. Schilling, *Syst. Bot.* **27**, 674 (2002).

76. S. L. Dillon *et al*., *Sorghum laxiflorum* and *S. macrospermum*, the Australian native species most closely related to the cultivated S. bicolor based on ITS1 and ndhF sequence analysis of 25 *Sorghum* species. *Plant Syst. Evol.* **249**, 233 (2004). doi:10.1007/s00606-004-0210-7

77. S. B. Hoot, N. S. Napier, W. C. Taylor, *Syst. Bot.* **31**, 449 (2006).

78. W. C. Taylor, N. T. Luebke, D. M. Britton, R. J. Hickey, D. F. Brunton, in *Isoëataceae Reichenbach -Quillwort Family*, vol. 2 of *Flora of North America* (Missouri Botanical Garden, St. Louis, MO, 1993).

79. D. F. Brunton, D. M. Britton, *Rhodora* **100**, 261 (1998).

80. D. F. Brunton, D. M. Britton, Rush quillwort (*Isoetes junciformis*, sp. nov.), a new pteridophyte from southern Georgia. *Am. Fern J.* **89**, 187 (1999). doi:10.2307/1547421

81. C. J. Van den Heede, R. L. L. Viane, M. W. Chase, Phylogenetic analysis of *Asplenium* subgenus *Ceterach* (Pteridophyta: Aspleniaceae) based on plastid and nuclear ribosomal ITS DNA sequences. *Am. J. Bot.* **90**, 481 (2003). doi:10.3732/ajb.90.3.481 Medline

82. H. Schneider *et al*., Chloroplast phylogeny of asplenioid ferns based on *rbcL* and *rnL-F* spacer sequences (Polypodiidae, Aspleniaceae) and its implications for biogeography. *Syst. Bot.* **29**, 260 (2004). doi:10.1600/036364404774195476

83. L. R. Perrie, P. J. Brownsey, Insights into the biogeography and polyploid evolution of New Zealand *Asplenium* from chloroplast DNA sequence data. *Am. Fern J.* **95**, 1 (2005). doi:10.1640/0002-8444(2005)095[0001:IITBAP]2.0.CO;2

84. J. M. Geiger, T. A. Ranker, Molecular phylogenetics and historical biogeography of Hawaiian *Dryopteris* (Dryopteridaceae). *Mol. Phylogenet. Evol.* **34**, 392 (2005). doi:10.1016/j.ympev.2004.11.001 Medline

85. C. X. Li, S. G. Lu, Phylogenetics of Chinese *Dryopteris* (Dryopteridaceae) based on the chloroplast rps4-trnS sequence data. *J. Plant Res.* **119**, 589 (2006). doi:10.1007/s10265-006-0003-x Medline

86. T. Nakazato, G. J. Gastony, *Syst. Bot.* **28**, 490 (2003).

87. E. Schuettpelz, H. Schneider, L. Huiet, M. D. Windham, K. M. Pryer, A molecular phylogeny of the fern family Pteridaceae: Assessing overall relationships and the

affinities of previously unsampled genera. *Mol. Phylogenet. Evol.* **44**, 1172 (2007). doi:10.1016/j.ympev.2007.04.011 Medline

88. Tropicos, Missouri Botanical Garden, www.tropicos.org (2011).

89. E. M. Sigel, M. D. Windham, L. Huiet, G. Yatskievych, K. M. Pryer, *Syst. Bot.* **36**, 554 (2011).

90. M. D. Windham, in *Pteridophytes and Gymnosperms*, vol. 2 of *Flora of North America* (Missouri Botanical Garden, St. Louis, MO, 1993).

91. J.-M. Lu, D.-Z. Li, L. M. Gao, X. Cheng, D. Wu, Paraphyly of *Cyrtomium* (Dryopteridaceae): Evidence from *rbcL* and *trnL-F* sequence data. *J. Plant Res.* **118**, 129 (2005). doi:10.1007/s10265-005-0201-y Medline

92. Data available through TreeBase (www.treebase.org), study accession no. S11651.

93. S. Hennequin, A. Ebihara, J. Y. Dubuisson, H. Schneider, Chromosome number evolution in *Hymenophyllum* (Hymenophyllaceae), with special reference to the subgenus *Hymenophyllum*. *Mol. Phylogenet. Evol.* **55**, 47 (2010). doi:10.1016/j.ympev.2010.01.001 Medline

94. P. Madeira, R. Pemberton, T. Center, *Biol. Control* **45**, 308 (2008).

95. J. Hanks, thesis, City University, New York, NY (1998).

96. M. D. Windham, G. Yatskievych, *Am. J. Bot.* **90**, 1788 (2003).

97. J. T. Mickel, *Phytologia* **41**, 431 (1979).

98. J. T. Mickel, A. R. Smith, *The Pteridophytes of Mexico* (New York Botanical Garden Press, New York, 2004).

99. C. J. Rothfels, M. D. Windham, A. L. Grusz, G. J. Gastony, K. M. Pryer, Toward a monophyletic Notholaena (Pteridaceae): Resolving patterns of evolutionary convergence in xeric-adapted ferns. *Taxon* **57**, 712 (2008).

100. C. J. Rothfels, in Tree of Life Web Project, http://tolweb.org (2008).

101. R. E. B. Kirkpatrick, Investigating the monophyly of *Pellaea* (Pteridaceae) in the context of a phylogenetic analysis of cheilanthoid ferns. *Syst. Bot.* **32**, 504 (2007). doi:10.1600/036364407782250616

102. W. Bouma, P. Ritchie, L. R. Perrie, Phylogeny and generic taxonomy of the New Zealand Pteridaceae ferns from chloroplast *rbc* L DNA sequences. *Aust. Syst. Bot.* **23**, 143 (2010). doi:10.1071/SB09047

103. M. D. Windham, in *Pteridophytes and Gymnosperms*, vol. 2 of *Flora of North America* (Missouri Botanical Garden, St. Louis, MO, 1993).