

Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms

Niv Sabath¹, Emma E. Goldberg², Lior Glick¹, Moshe Einhorn¹, Tia-Lynn Ashman³, Ray Ming^{4,5}, Sarah P. Otto⁶, Jana C. Vamوسي⁷ and Itay Mayrose¹

¹Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv, Israel; ²Department of Ecology, Evolution & Behavior, University of Minnesota, St Paul, MN 55108-6097, USA; ³Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA; ⁴FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Haixia Institute for Science and Technology, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China; ⁵Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; ⁶Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6J 3S7, Canada; ⁷Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada

Author for correspondence:

Itay Mayrose

Tel: +972 3 640 7212

Email: itaymay@post.tau.ac.il

Received: 17 May 2015

Accepted: 25 August 2015

New Phytologist (2016) **209**: 1290–1300

doi: 10.1111/nph.13696

Key words: bisexuals, dioecy, diversification, extinction, plant sexual systems, speciation.

Summary

- Dioecy, the sexual system in which male and female organs are found in separate individuals, allows greater specialization for sex-specific functions and can be advantageous under various ecological and environmental conditions. However, dioecy is rare among flowering plants. Previous studies identified contradictory trends regarding the relative diversification rates of dioecious lineages vs their nondioecious counterparts, depending on the methods and data used.
- We gathered detailed species-level data for dozens of genera that contain both dioecious and nondioecious species. We then applied a probabilistic approach that accounts for differential speciation, extinction, and transition rates between states to examine whether there is an association between dioecy and lineage diversification.
- We found a bimodal distribution, whereby dioecious lineages exhibited higher diversification in certain genera but lower diversification in others. Additional analyses did not uncover an ecological or life history trait that could explain a context-dependent effect of dioecy on diversification. Furthermore, in-depth simulations of neutral characters demonstrated that such bimodality is also found when simulating neutral characters across the observed trees.
- Our analyses suggest that – at least for these genera with the currently available data – dioecy neither consistently places a strong brake on diversification nor is a strong driver.

Introduction

Sexual reproduction entails the union of a male gamete (sperm) and a female gamete (egg), but in angiosperms the universality of separate sexes at the gametophytic stage is coupled with a diverse array of sexual systems at the sporophytic stage. Some species have separate male and female sexes (dioecy), many species are entirely hermaphroditic, and virtually every variant in between is found (Supporting Information Table S1). Because of the high degree of variability in sexual systems, flowering plants provide an ideal group in which to conduct comparative evolutionary analyses to understand the evolution of sexual systems. In particular, there is a long-standing interest in understanding why dioecy is such a rare sexual system in flowering plants, present in only *c.* 6% of the species (Renner, 2014) compared with *c.* 95% in animals (Jarne & Auld, 2006). The rarity of a trait (here dioecy) can be explained in a number of ways: it may slow speciation, hasten extinction, be slow to emerge from an ancestral state, or disappear faster than it appears (Thomson & Brunet, 1990; Schwander & Crespi, 2009).

Although rare, dioecy is believed to be advantageous as a means to facilitate outcrossing, particularly when other outcrossing mechanisms are absent, thereby preventing inbreeding depression (Thomson & Barrett, 1981; Thomson & Brunet, 1990). In addition, theoretical (Charnov, 1982) and empirical studies (Bañuelos & Obeso, 2004; Montesinos *et al.*, 2006; Tognetti, 2012) have shown that dioecy might be positively selected because it dampens trade-offs between male and female functions. Dioecy may also be selectively favored under certain environmental conditions. For example, Muenchow (1987) argued that dioecy may be favored in understory habitats where nonspecific pollinators, which are often associated with dioecy, tend to be found. Dioecious species have also flourished on the Hawaiian islands, perhaps because dispersal among islands favors long-lived tree species with fleshy fruits, both traits that are associated with dioecy (Schwander & Crespi, 2009).

Despite these potential advantages, species richness of dioecious angiosperm clades is often lower than that of hermaphroditic sister groups (Heilbut, 2000; Kay *et al.*, 2006). Dioecious lineages may suffer a lower diversification rate for a number of reasons. Species

that are dioecious may be less able to colonize new sites because of the need for a mating partner (Baker, 1955; Pannell & Barrett, 1998) or because seed production is restricted to the female half of a population (Heilbuth *et al.*, 2001; Wilson & Harder, 2003), reducing opportunities for allopatric or parapatric speciation. Reduced speciation rates may also result from the higher rates of gene flow and subsequently lower population differentiation among obligate outcrossers (Barrett, 2003). Alternatively, sexual dimorphism driven by stronger sexual selection acting on males to attract pollinators can place dioecious lineages at greater risk of extinction, particularly when pollinator availability is fluctuating (Vamosi & Otto, 2002). Indeed, a survey of the International Union for Conservation of Nature Red List of Threatened Species revealed that dioecious families are enriched for species at risk (Vamosi & Vamosi, 2005), thus hinting that dioecy is associated with elevated extinction risk, at least under present conditions.

Furthermore, dioecious lineages tend to exhibit a suite of correlated traits, including abiotic pollination, biotic dispersal, woody growth form, tropical distribution (Renner & Ricklefs, 1995; Vamosi *et al.*, 2003; Vamosi & Vamosi, 2004), and to some extent polyploidy (Miller & Venable, 2000; Ashman *et al.*, 2013), which in turn may affect their evolution. An analysis of transition rates to dioecy in angiosperms found little support overall for the idea that dioecy is more prone to evolve in lineages that have a woody growth form, tropical distribution, plain flowers, or fleshy fruits than in those lacking them (Vamosi *et al.*, 2003). Rather, there was more support for the hypothesis that correlations between dioecy and ecological traits are caused via differential net diversification (i.e. speciation minus extinction) of dioecious lineages when associated with the aforementioned traits (Vamosi & Vamosi, 2004). Furthermore, these correlations with other traits may themselves impact diversification. For example, the lower diversification rates of dioecious lineages could be explained by an association with polyploidy, a trait that also exhibits lower diversification rates at the genus level (Mayrose *et al.*, 2011).

Previous comparisons of diversification rate of dioecious lineages vs their nondioecious counterparts were based on sister-clade contrasts at the genus and family levels (Heilbuth, 2000; Kay *et al.*, 2006). While sister-clade comparisons have been widely used in comparative evolutionary studies to reveal a character that is associated with altered diversification rates (Mitter *et al.*, 1988; Farrell *et al.*, 1991; de Queiroz, 1998), this method cannot differentiate between effects of a trait on speciation vs extinction, making it difficult to understand the underlying mechanism. In addition, sister-clade contrasts discard phylogenetic information both internal and external to the sister clades and assume only a single trait transition differentiating the two sisters (but see Heilbuth, 2000). Furthermore, a recent analysis demonstrated that sister-clade comparisons can be strongly biased when the derived state has fewer species because the waiting time until the derived state first appears reduces opportunities for diversification in that state (Käfer & Mousset, 2014). Notably, using a resampling test that should correct this problem, Käfer & Mousset (2014) reanalyzed the data of Heilbuth (2000) and concluded that dioecy does not lead to lower net diversification and

may, if anything, increase it. This reanalysis assumed, however, that the most recent common ancestor of each sister-clade pair was nondioecious and ignored extinctions, assumptions that would overcorrect the bias. In a subsequent study, Käfer *et al.* (2014) used an updated data set, including only those sister clades where dioecy was inferred to be the derived state. Using the corrected sister-clade method of Käfer & Mousset (2014), the authors reached a similar conclusion, namely, dioecy tends to increase diversification. Käfer *et al.* (2014) suggested that the rarity of dioecy may instead be attributable to frequent losses of dioecy, which contradicts the classic view that the evolution of dioecy is largely irreversible (Bull & Charnov, 1985; but see Barrett, 2013). Thus, as fundamental as sexual systems are to angiosperm biodiversity, we currently do not understand why dioecious lineages are rare.

Recent developments in comparative phylogenetic methods allow the simultaneous estimation of rates of speciation and extinction, as well as transition rates between states, within a likelihood framework (Maddison *et al.*, 2007; FitzJohn *et al.*, 2009; FitzJohn, 2012). These methods require phylogenetic trees with known trait states for extant taxa, although complete phylogenetic sampling and trait information are not required (FitzJohn, 2012). One important caveat is that these likelihood methods only reveal a correlation between a particular trait and speciation or extinction rate (Maddison *et al.*, 2007) and do not show causation. In particular, even if other factors underlie variation in diversification rate, a statistically significant association between a trait and diversification can be obtained if transitions in the trait happen, by chance, to cause a trait state to be associated with a more diverse portion of the tree (FitzJohn, 2012; Maddison & FitzJohn, 2015). Consequently, these methods could suffer from an elevated false-positive rate when analyzing a single clade as a result of processes that were unaccounted for in the model (Rabosky & Goldberg, 2015). In order to avoid such coincidental associations, Rabosky & Goldberg (2015) suggested performing comparative analyses on multiple data sets, drawing conclusions based on a preponderance of evidence (Mayrose *et al.*, 2011; Roland *et al.*, 2014).

To this end, we gathered data on sexual systems in many taxonomic groups (largely genera) that contain both dioecious and nondioecious species (Tree of Sex Consortium, 2014). In this study, we assembled molecular phylogenies for dozens of genera for which we had detailed sexual system information. This allowed us to investigate the consequences of dioecy for rates of speciation and extinction using the binary state speciation and extinction (BiSSE) likelihood method (Maddison *et al.*, 2007; FitzJohn, 2012). The inferred speciation and extinction rate parameters for dioecious and nondioecious taxa were then compared across genera to determine whether dioecy has a consistent effect on diversification. We further examined whether differences in ecological and life history traits accounted for the observed variation in diversification patterns among these groups. Moreover, we examined whether the diversification rate of dioecious lineages was significantly correlated with the form of the nondioecious taxa (primarily hermaphroditic or primarily monoecious), given that some of the advantages of dioecy may be

shared by monoecious species (reduced inbreeding and an elimination of trade-offs at the flower level between male and female function; Barrett, 2002, 2003). Finally, we conducted extensive simulations to reveal the distribution of the relative diversification rates of dioecious lineages expected for a 'neutral' trait, given the observed trees and transition rates. We close by discussing additional factors that may affect the observed patterns, including limitations of the current data sets.

Materials and Methods

Database

A broad data set of angiosperm sexual systems was assembled as part of a working group sponsored by the National Evolutionary Synthesis Center, with a focus on 77 genera exhibiting both the presence and absence of dioecious species (Tree of Sex Consortium, 2014). The genera used in the current study were chosen largely from the list of angiosperm genera that contain dioecious species (Renner & Ricklefs, 1995) and supplemented with genera analyzed by Miller & Venable (2000), focusing on groups with sufficient genetic data in GenBank (Benson *et al.*, 2013).

While ideally all sexual systems would be considered separately, doing so in a diversification analysis leads to a large number of parameters (speciation and extinction rates for each sexual system, as well as all possible transition rates). We thus grouped together different sexual systems to obtain a manageable number of model parameters. Specifically, we classified the wide array of possible sexual systems (Table S1) into binary categories, using two different classifications. For 'strict dioecy,' we compared dioecious lineages against all other sexual systems which exhibit at least partial cosexuality (i.e. classifying gynodioecy, androdioecy, and polygamodioecy apart from dioecy). For 'broad dioecy,' we grouped all systems with some unisexual plants (including dioecy, gynodioecy, androdioecy, and polygamodioecy), in contrast to the 'bisexual' state where both sexes are found on every plant (hermaphrodites and all forms of monoecy). In addition, the resulting data set was filtered to genera for which the reconstructed phylogeny included at least 10 species and at least two species in each state (dioecy and nondioecy). Taxonomic name resolution was largely performed using The Plant List (v.1.1; <http://www.theplantlist.org/>). We also employed synonymy provided by experts in particular groups (notably Solanaceae Source; <http://solanaceaesource.org>). The resulting data set included 38 genera in the 'strict' classification and 41 in the 'broad' classification (Table S2).

Phylogenetic reconstruction

The phylogeny for each genus was reconstructed using the following general procedures. (A detailed description of the steps is provided in Methods S1). Sequence data for all species in the genus were automatically retrieved from GenBank (Benson *et al.*, 2013) using National Center for Biotechnology Information taxonomy. This was followed by a name resolution procedure that matched species names as they appear in GenBank to accepted taxonomic

names (according to The Plant List and Solanaceae Source). Sequences were then clustered to orthology groups using ORTHOMCL (Li *et al.*, 2003). An appropriate outgroup was chosen and then added to each cluster. The sequences in each cluster were then aligned, filtering out unreliably aligned sequences and positions and then concatenated into one multiple-sequence alignment. Next, ultrametric phylogenetic trees were reconstructed with MRBAYES v.3.2.1 (Ronquist *et al.*, 2012), using the best-fitting nucleotide substitution model inferred for each locus based on MRBAYES (Nylander, 2004; <https://www.abc.se/~nylander/mraic/mraic.html>). Trees were pruned to remove taxa whose position in the tree was highly uncertain (listed in Table S3). To account for phylogenetic uncertainty, our diversification analyses used 100 trees sampled from the Bayesian posterior distribution. All alignments and phylogenies are available in DRYAD (<http://dx.doi.org/10.5061/dryad.7f4v1/1>).

Diversification analysis

With both the strict and broad classification schemes, we estimated diversification rates for dioecious (D) and nondioecious (N) lineages using the BiSSE model (Maddison *et al.*, 2007) as implemented in the *diversitree* R package (FitzJohn, 2012). Six parameters were simultaneously estimated for each genus: the transition rates from N to D (q_{ND}) and from D to N (q_{DN}), the speciation rates of lineages in states N and D (λ_N and λ_D , respectively), and extinction rates for lineages in states N and D (μ_N and μ_D , respectively). Using these estimates, the difference between speciation and extinction rate provides the net diversification rate for each state ($r_N = \lambda_N - \mu_N$ and $r_D = \lambda_D - \mu_D$). We note that extinction rates are difficult to estimate and are particularly sensitive to sampling biases (Rabosky, 2010; but see Beaulieu & O'Meara, 2015). However, the net diversification rate, which is the primary focus of our analysis, does not rely on a precise estimate of the extinction rate and accounts for correlations between speciation and extinction. Nevertheless, we complemented each diversification analysis with one in which the BiSSE model is constrained to enforce equal extinction in both states. Our analyses were performed using the 'skeletal' tree approach (FitzJohn *et al.*, 2009), which accounts for the sampling fraction of species in the given phylogeny out of the total number of species in the clade. Sampling fraction estimates were calculated based on the number of species in a genus as obtained through a literature search (see references in Table S2). To assess the sensitivity of our results to the assumed fraction of sampled species, we also used the total number of accepted and unresolved species from The Plant List (v.1.1; <http://www.theplantlist.org/>), which on average yields twice the number of species (Table S2); this second analysis generated extremely similar results, so only the results using the number of species in the genus reported in the literature are given here.

A Markov chain Monte Carlo (MCMC) sampling approach described in FitzJohn *et al.* (2009) was used to estimate the posterior probability distributions for each of the six parameters, accounting for uncertainty in parameter estimation, incomplete sampling, and phylogeny. An exponential prior distribution was

used for each parameter (see next paragraph for more details). For each of the sampled trees, MCMC chains were started at the maximum likelihood estimates of the model parameters and were run for 2000 steps with the first 10% discarded as burn-in. The MCMC results of all 100 trees for each genus were combined together. To test whether extinction and speciation rates differ between sexual systems, while accounting for uncertainty in the tree and in the transition rate parameters (q_{ND} and q_{DN}), we calculated for each genus the proportion of BiSSE MCMC steps (i.e. the posterior probability (PP)) in which a given rate (i.e. λ , μ or r) was higher for nondioecious lineages than for dioecious lineages. For example, $PP(r_N > r_D)$ is the proportion of post-burn-in steps in which nondioecious lineages were inferred to diversify at a higher rate than dioecious lineages. (Simulations reported in Fig. 6 of FitzJohn *et al.* (2009) indicate that $PP(r_N > r_D)$ does not differ significantly from 50% when diversification rates are trait independent and that this measure is able to detect trait-dependent diversification.)

To examine the robustness of the MCMC results, each genus was subjected to three different MCMC analyses, each with a different prior distribution. Results that are sensitive to the prior choice suggest that the data for that group lack sufficient information to overcome the prior. Specifically, three different exponential prior distributions for each parameter were examined. First, we used the `starting.point.bisse` function of *diversitree* to fit a trait-independent model (assuming sexual system does not influence diversification) to obtain initial parameter estimates for the speciation (λ_0), extinction (μ_0), and transition (q_0) rates. Given these estimates, three different prior distributions were obtained for each parameter by setting the mean of the prior distribution to $p \times f$, where $f = 1, 2, 4$ and p is the estimated rate parameter (i.e. for speciation rates we obtained three exponential priors by setting the mean of the exponential distribution to λ_0 , $2\lambda_0$, and $4\lambda_0$). Genera were not considered robust and were excluded (see Table S2) if $PP(r_N > r_D)$ values differed by $> 20\%$ across the different priors. We report the results obtained using $f = 2$, the prior rate value that was suggested by FitzJohn *et al.* (2009) (all results are available at DRYAD <http://dx.doi.org/10.5061/dryad.7f4v1/1>). To assess significance over the whole data set, we used a one-sample Wilcoxon rank-sum test with a mean equal to 50%, testing the null hypothesis that diversification rates in the D state should be higher than those in the N state half of the time, treating the PP value of each genus as a single data point. In addition, we used Hartigan's dip test (Hartigan & Hartigan, 1985) to test whether there was significant evidence for multimodality in the PP distribution.

We further examined whether the inferred diversification rate differences between dioecious and nondioecious clades were correlated with various ecological or life history traits: growth form (woody vs herbaceous), pollination mode (biotic vs abiotic), dispersal mode (biotic vs abiotic), and geographical distribution (tropical vs temperate). These other traits were coded based on the genus level categorization of Renner (2014). For growth form, Renner (2014) only specified presence/absence of trees, shrubs, herbs, or climbers, and many of the genera contained both woody and herbaceous species. To determine whether the

genus was predominantly woody or herbaceous, we gathered species-level life form data from the literature (Tree of Sex Consortium, 2014), supplemented with data from Zanne *et al.* (2014) and data collected from eFLORA (Brach & Song, 2006). Genera that are mostly ($\geq 70\%$) woody or herbaceous were defined as such. Genera exhibiting a mixture of the growth forms (with no one state above 70%) were excluded from the particular comparison. For geographical distribution, the database of Renner (2014) specified whether the genus is distributed in tropical and/or temperate regions. Again, genera distributed in both tropical and temperate regions were excluded from the particular comparison. In addition to the life history traits described earlier, we also examined whether the types of nondioecious species present in a genus are associated with the inferred diversification rates. We categorized genera in which the nondioecious state is mostly ($\geq 70\%$) hermaphrodite as 'Mostly H'; genera with mostly ($\geq 70\%$) monoecy were labeled as 'Mostly M'; all other genera were dropped from this analysis. For each categorization, we assessed whether support for a diversification rate difference ($PP(r_N > r_D)$) differed significantly from 50% using a one-sample Wilcoxon rank-sum test.

Simulations

Estimation of diversification rates could be influenced by additional factors other than the trait of interest (here sexual system), leading to an elevated false-positive rate (FitzJohn, 2012; Rabosky & Goldberg, 2015). Thus, we used a parametric bootstrapping approach to obtain the null distribution of the relative diversification rates expected for each genus. Specifically, we compared our observed data against simulated data obtained by tracking the evolution of neutral characters (assuming no effect on diversification) on the same set of empirically derived phylogenies. For each genus, we drew a single tree from the posterior set, to represent the true phylogenetic relationships. Using this tree and the known sexual systems of the tips, we estimated the two transition rates (from and to dioecy) and the probability of the root being in each of the two states according to the full six-parameter BiSSE model using a short (200 steps) MCMC chain. We then simulated binary traits along the tree using a Markov model whose only parameters were the estimated transition rates (mk2 model implemented in the package *diversitree*; FitzJohn, 2012) and starting from an ancestral state drawn randomly from the root state probability distribution. Thus, these simulations assumed no direct effect of the character on speciation or extinction. Similar to the analysis of the true data, we kept only simulated data sets that included at least two tips of each state, repeating the procedure until 100 data sets were obtained. The simulated state data were then subjected to the same BiSSE MCMC procedures as the original data, conducted on 20 other trees from the posterior distribution (not all 100, to reduce computation time). We used the distribution of simulated $PP(r_N > r_D)$ values to calculate a per-genus P -value as the proportion of simulated values that are as extreme as the observed $PP(r_N > r_D)$ for that genus. A P -value < 0.025 or > 0.975 can be interpreted as significant support for the hypothesis that the

sexual system affects net diversification in the examined genus. Because the percentage of the rare state could influence diversification estimates, we analyzed an additional and independent set of simulated data sets in which simulated data sets were kept only if the percentage of tips within each state differed by < 10% from that observed in the original data. We denote these two approaches as ‘free tip frequency’ and ‘matched tip frequency’.

Results

Sexual system database

To date, the most comprehensive database detailing the taxonomic distribution of dioecy was first compiled by Renner & Ricklefs (1995) and recently updated (Renner, 2014). The latest database lists 987 genera in which dioecious species are known to exist, *c.* 1/3 of which vary in sexual system within the genus. This database is not suitable for detailed comparative phylogenetic analysis because sexual system assignments are given at the genus rather than species level. To this end, we summarized the Tree of Sex database of species-level sexual system assignments for angiosperm genera with variation in sexual system: 38 genera were included in the ‘strict’ classification and 41 in the ‘broad’ classifications. The ‘strict’ classification includes 475 dioecious and 1535 nondioecious species, whereas the ‘broad’ classification includes 663 species with unisexual plants and 1676 consisting of bisexual plants only (Table S2). In total, the reconstructed phylogenies of these genera include 3380 species and cover on average 42% of the recognized species in these groups.

Diversification analysis

The diversification results using ‘broad’ and ‘strict’ classifications, with and without constraining extinction rates, are summarized in Table 1. Because these analyses yielded similar results (Figs 1, 2, S1, S2; distributions of relative speciation, extinction and transition rates are shown in Fig. S3), we focus here on the results of the ‘broad’ classification (which includes more genera) without constraining extinction rates. Of the 41 genera considered within

the ‘broad’ classification, preliminary BiSSE runs indicated that the results for five data sets were highly dependent on the exact prior choice and were thus deemed unreliable. Dropping these genera, our final data set consisted of 36 genera, whose reconstructed phylogenies encompassed 2702 species (Table S2). Fig. 1 presents the distribution of $PP(r_N > r_D)$ values with each genus represented by the value inferred across the sample of 100 Bayesian trees. (The per-tree $PP(r_N > r_D)$ values, as well as the raw distributions of r_N and r_D , for each genus are shown in Fig. S4). Our results indicate that dioecious lineages do not consistently diversify at higher or lower rates compared with non-dioecious lineages (that is, $PP(r_N > r_D)$ did not deviate significantly from 50%, $P=0.17$; one-sample Wilcoxon rank-sum test). However, the distribution exhibited significant bimodality ($P<0.048$; Hartigans’ dip test). Approximately 67% of the genera exhibited higher diversification in the dioecious state (with four genera showing $PP(r_N > r_D) < 2.5\%$), while 33% exhibited lower diversification (with three genera showing $PP(r_N > r_D) > 97.5\%$). The number of genera in the extremes of these two tails (seven) represents a significant excess over the 5% expected (exact binomial test; $P=0.0018$). The posterior

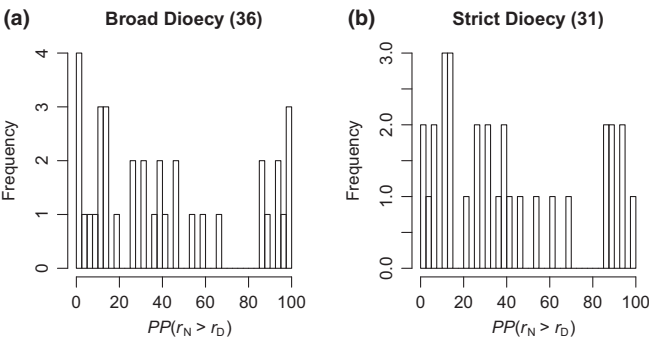


Fig. 1 The proportion of the posterior probability distribution supporting a higher diversification rate for nondioecious species than dioecious species, $PP(r_N > r_D)$, using (a) ‘broad’ and (b) ‘strict’ definitions of dioecy. With either classification, the distribution of $PP(r_N > r_D)$ values did not deviate from 50% ($P=0.13$; one-sample Wilcoxon rank-sum test). In the ‘broad’ classification, the distribution showed significant evidence of multimodality ($P=0.048$; Hartigans’ dip test). In the ‘strict’ classification, the distribution did not exhibit a significant multimodality ($P=0.18$).

Table 1 Summary of diversification analyses

Model	Classification	N^a	N^b	P (mean = 50%) ^c	P (dip test) ^d	corr (r, λ) ^e	corr (r, μ) ^f
Six parameters	Strict	38	31	0.13	0.1838	0.9 (0)	0.11 (0.55)
	Broad	41	36	0.17	0.0482	0.93 (0)	0.07 (0.70)
Constrained extinction ($\mu_N = \mu_D$)	Strict	38	30	0.34	0.1710	1 (0)	na ^g
	Broad	41	34	0.31	0.0374	1 (0)	na ^g

^aThe number of genera with at least 10 species total and two species in each state.
^bThe number of genera remaining after removing genera that exhibit sensitivity to parameters.
^c P -value for testing if the distribution of $PP(r_N > r_D)$ deviates from 50% (one-sample Wilcoxon rank-sum test).
^d P -value for testing if the distribution of $PP(r_N > r_D)$ deviates from a uniform or unimodal distribution (Hartigans’ dip test).
^ePearson correlation coefficient (and corresponding P -value) between $PP(r_N > r_D)$ and $PP(\lambda_N > \lambda_D)$.
^fPearson correlation coefficient (and corresponding P -value) between $PP(r_N > r_D)$ and $PP(\mu_N > \mu_D)$.
^gna, not applicable. No correlation is computed because extinction rates are constrained to be equal ($PP(\mu_N > \mu_D)$ is undefined).

distribution of $r_N > r_D$ was unimodal for most genera, as was the distribution of $PP(r_N > r_D)$ across the trees for a genus (Fig. S4). This consistent signal from each genus indicates that the analysis was robust within and among the sampled trees. Rather, the bimodality arose because dioecy was correlated with higher diversification in some genera and lower diversification in others (see the Simulations and sampling issues subsection below for evidence that this bimodality might reflect other unknown processes affecting the shape of the trees).

Overall, the posterior probabilities for diversification rate differences, $PP(r_N > r_D)$, were highly correlated with those inferred for speciation rate ($PP(\lambda_N > \lambda_D)$: $r = 0.93$; $P < 10^{-10}$; $n = 36$) but not for extinction rate ($PP(\mu_N > \mu_D)$: $r = 0.07$; $P = 0.7$), suggesting that diversification rates inferred for these genera generally reflected inferred rates of speciation. Furthermore, $PP(r_N > r_D)$ values were not correlated ($P = 0.7$) with the transition rates asymmetries (i.e. $PP(q_{ND} > q_{DN})$), suggesting that the apparent pattern of diversification is not driven by biased transition rates between the sexual systems.

Results were fairly similar using the 'strict' definitions of dioecy. Of the 38 genera, seven were filtered out based upon sensitivity to the prior, leaving 31 genera. The mean of the $PP(r_N > r_D)$ distribution did not deviate significantly from 50% ($P = 0.13$; one-sample Wilcoxon rank-sum test). Again, the distribution exhibited a bimodal pattern (Fig. 1b), but the bimodality was not significant according to Hartigan's dip test ($P = 0.18$; $D_n = 0.07$), possibly because of the smaller sample size. In addition, one genus displayed $PP(r_N > r_D) < 2.5\%$ and two displayed $PP(r_N > r_D) > 97.5\%$, which represents a slightly higher density in the tails, but not significantly so ($P = 0.2$; exact binomial test).

Results were similar when constraining the extinction rates to be equal (μ_N and μ_D), suggesting that the known difficulty in accurately estimating extinction rates (Rabosky, 2010) does not influence the conclusions using either 'strict' or 'broad' classifications (Table 1; Fig. S1). In all four analyses ('strict' or 'broad' classifications, with and without constraining the extinction rates), genera with more nondioecious species tended to display higher $PP(r_N > r_D)$ values (Fig. S5), as expected whether or not sexual system truly influences diversification rates (e.g. when by chance dioecy happens to be overrepresented in a clade, higher r_D will be inferred and vice versa). Note that several clades have very skewed tip ratios (very low or high percentage dioecy), which strongly limits the power of BiSSE (Davis *et al.*, 2013).

Trait correlations

We tested whether the high degree of variation observed in relative diversification rates for dioecious and nondioecious lineages (Fig. 1) could be attributed to differences among the genera in certain ecological and life history traits (i.e. testing whether dioecy does affect diversification rates but only in the background of another trait). Specifically, we subdivided the genera according to whether they were predominantly herbaceous or woody, temperate or tropical, animal dispersed or abiotically dispersed, animal pollinated or abiotically pollinated, or primarily hermaphroditic or monoecious when considering the

nondioecious species. As shown in Fig. 2, we did not find any significant associations between the relative diversification rates of dioecious lineages and the analyzed traits (all P -values above 0.27).

Simulations and sampling issues

To examine whether our results could be explained solely by elevated false positives from the BiSSE methodology (FitzJohn, 2012; Rabosky & Goldberg, 2015), we used neutrally simulated traits to construct a null hypothesis of the distribution of $PP(r_N > r_D)$ on our empirically derived phylogenies. The results were amalgamated into a single histogram for visualization (Fig. 3), although one must keep in mind that these points are based on the same 36 genera and hence are not fully independent. Interestingly, the $PP(r_N > r_D)$ distributions were strongly bimodal under the two simulated scenarios ('free tip frequency' or 'matched tip frequency'), resembling the distribution obtained for the empirical data set (Fig. 1). Considering the simulated distributions within each genus separately (Figs S6, S7), it is apparent that the distribution was not bimodal in most genera but rather peaked near either the right or left tail of the distribution, leading to an overall bimodality pattern across the genera, as for the true sexual system states. We next sampled 100 data sets, choosing one inferred $PP(r_N > r_D)$ from one simulated data set per genus, thus creating a background distribution for our meta-analysis, and calculated bimodality using two different metrics. In both cases, we found that the bimodality exhibited by the empirical data was not significantly different from the bimodality expected by chance among the simulated data sets. First, the observed density of data points in the two extreme tails ($PP(r_N > r_D) < 0.025$ or > 0.975) fell within the distribution found in the simulated data sets (Fig. 3b,e). Second, the simulated data sets typically showed bimodality, with a significant dip test statistic in 66 ('free tip frequency') or 88 ('matched tip frequency') out of 100 simulated data sets. Again, the observed data fell well within the distribution of simulated test statistics (Fig. 3c,f).

While the shapes of the observed phylogenetic trees often lead randomly generated characters to exhibit evidence for differential diversification (high or low $PP(r_N > r_D)$ values), the observed data suggested a stronger effect of dioecy on diversification for some genera than expected based on the distribution of simulated characters. Specifically, for each genus, the empirically derived $PP(r_N > r_D)$ values were compared to the 100 simulated data sets (Figs S6, S7 for 'free tip frequency' and 'matched tip frequency'). Given that only 100 simulations were performed, a strict multiple corrections test could not be performed, but we focused only on those observed results that fell outside of the full range of the 100 simulations. With the 'free tip frequency' approach, dioecy is inferred to have a more positive effect on diversification than 100% of the simulated characters for *Dodonaea*, *Fragaria*, *Galium*, and *Sidalcea* and a more negative effect for *Pilea* (Fig. S6). Of these, the effect of dioecy on diversification remains significantly different from that for the simulated characters only for *Fragaria* and *Sidalcea* with the 'matched tip frequency' approach (Fig. S7). Using the strict definition of dioecy, dioecy

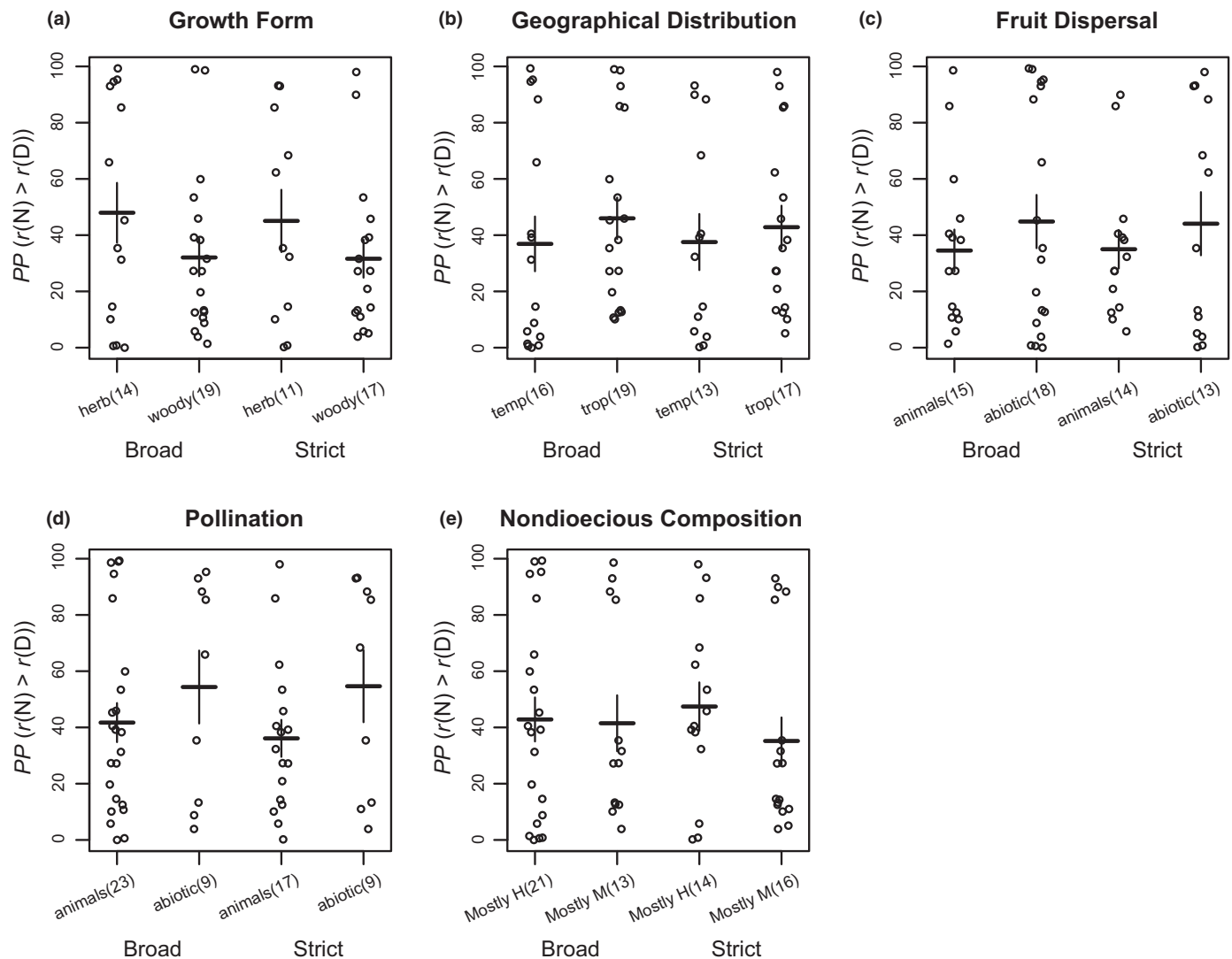


Fig. 2 The distribution of $PP(r_N > r_D)$ values for genera classified according to (a) growth form, (b) geographical distribution, (c) mode of fruit dispersal, (d) mode of pollination, and (e) whether the nondioecious species in the genus were predominantly hermaphroditic ('Mostly H') vs monoecious ('Mostly M'). Each panel presents both 'broad' (left side) and 'strict' (right side) classifications. Horizontal and vertical lines mark the sample means and \pm SE, respectively. For each particular trait, we tested whether the mean of the $PP(r_N > r_D)$ distribution was different from 50% using a one-sample Wilcoxon rank sum test. All P -values were nonsignificant, with $P > 0.27$. In addition, we did not find any significant difference between the $PP(r_N > r_D)$ of the two states of each trait (e.g. between woody and herbaceous) using a two-sample Wilcoxon rank sum test.

was only found to have a negative effect on diversification (in *Gunnera* and *Pilea* with the 'free tip frequency'; *Pilea* with the 'matched tip frequency'). Thus, focusing on genera that are atypical relative to the simulation results, dioecy did not have a consistent effect on diversification, and the direction of effect was not related to other traits putatively purported to influence diversification (e.g. woodiness or tropical distribution). Furthermore, even these genera might not have appeared unusual if we had accounted for uncertainty in the transition rates within the simulated data sets.

However, poor sampling may have also reduced our power to detect departures from neutrally simulated characters. In line with this idea, there was a significantly negative correlation between $PP(r_N > r_D)$ values and the percentage of species within a genus included in the tree ($r = -0.53$; $P = 0.001$). Indeed, the

taxa with the most complete sampling (*Sidalcea*, *Allocasuarina*, *Fraxinus*, *Momordica*, *Rhus*, *Dodonaea*, *Fragaria*, and *Lepechinia*, each with $> 60\%$ sampling on the phylogeny) tended to support higher diversification of dioecious clades, and all exhibited $PP(r_N > r_D)$ values below 50% (0.6, 3.9, 8.8, 10.1, 10.7, 12.7, 14.6 and 19.7%, respectively). This suggests that there might indeed be some clades where dioecy significantly impacts diversification and that increased sampling would improve our power to detect these clades.

Discussion

Here, we asked whether dioecy affects diversification rates by applying probabilistic phylogenetic methods to a large set of genera with both dioecious and nondioecious species. We found no

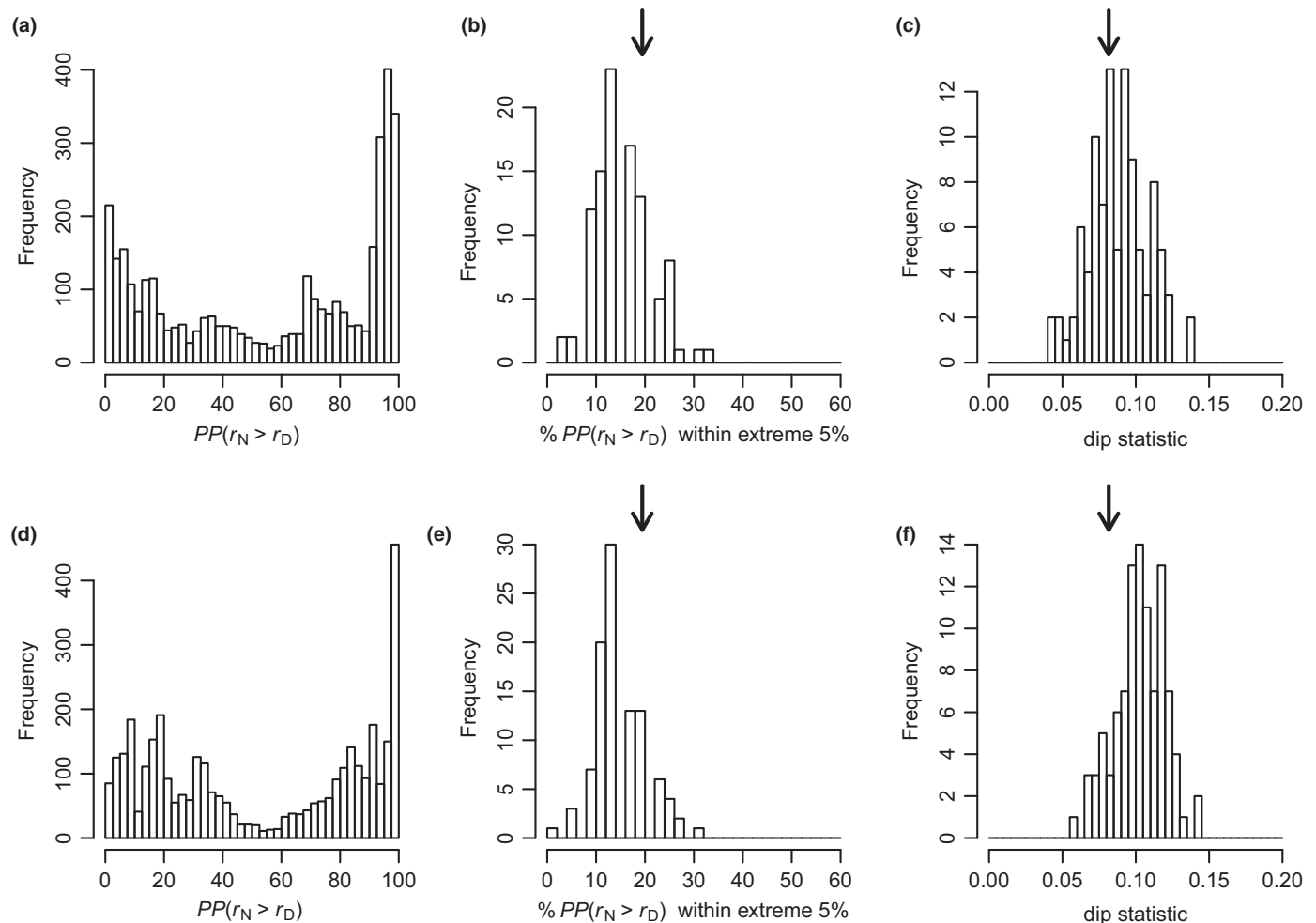


Fig. 3 Comparison of the bimodality exhibited by the observed data (black arrows) to the distribution from simulations, using either (a–c) the ‘free tip frequency’ or (d, e) the ‘matched tip frequency’ approaches. The simulated $PP(r_N > r_D)$ distributions over all 100 simulated data sets over all 36 genera are shown in (a) and (d). The 100 simulated data sets exhibit a similar fraction of genera in the 5% extreme tails (b, e; the percentage of data points $< 2.5\%$ or $> 97.5\%$) and similar dip test statistics (c, f) as the observed data.

consistent effect of dioecy on diversification. Instead, we observed a mixed pattern, with dioecious lineages sometimes exhibiting higher diversification and other times showing the reverse, and we showed that this pattern could be explained by chance. Our results are in line with those of several other studies that used similar methods in other plant groups and did not detect a strong effect of sexual system on diversification rates (Leslie *et al.*, 2013; McDaniel *et al.*, 2013; Villarreal & Renner, 2013). Villarreal & Renner (2013) found no difference between the diversification rates of monoicous and dioicous species applying BiSSE to a 98 species (out of 200) phylogeny of hornworts. Using a large conifer phylogeny and the BiSSE model, Leslie *et al.* (2013) found no effect of sexual system (monoecy or dioecy), cone type (dry or fleshy), or their combination on diversification rates. McDaniel *et al.* (2013) used a genus-level phylogeny of 493 moss genera and found only a nonsignificant trend towards higher diversification for hermaphrodites over dioecious species. Our analysis extends these previous studies by examining multiple clades simultaneously and conducting comprehensive simulations to evaluate the significance of the results.

While low power may have limited the fraction of clades with extreme diversification rate asymmetry (in the tails of the $PP(r_N > r_D)$ distribution), it cannot explain why there was substantial and significant bimodality in the effect of dioecy on diversification. Two different explanations could account for the observed bimodality. The first possibility is that dioecy does have a direct impact on diversification, but only when coupled with certain other traits (e.g. woody dioecious lineages might diversify at higher rates, but not herbaceous dioecious lineages). Our comparisons of diversification rates in the background of different ecological and life history traits revealed no association of the relative diversification rates with the various ecological or life history traits examined (growth form, pollination mode, dispersal mode, and geographical distribution). This conclusion contrasts with that of a previous study (Vamosi & Vamosi, 2004) which found that dioecious sister clades were relatively more speciose when associated with tropical distribution, woody growth form, plain flowers, and fleshy fruits. This difference in outcome might be attributable to the relatively small number of genera in our study, differences in methodology, and/or differences in the taxa

considered. Indeed, across the 994 genera in which dioecy is present (Renner, 2014), we found a significant difference in pollination and dispersal modes between genera that are entirely dioecious and genera that have multiple sexual systems. Fully dioecious genera are more associated with wind pollination ($P=0.025$; $\chi^2=35$; chi-square test) and animal dispersal ($P=0.03$; $\chi^2=20$), suggesting that the set of genera analyzed here does not perfectly represent the entire set of genera in which dioecy is present.

The second – and very likely – possibility is that the observed bimodality reflects other factors not included in our analysis that affect diversification rate, such as a different trait (e.g. geographical range) or temporal changes in speciation and/or extinction rate (FitzJohn, 2012; Rabosky & Goldberg, 2015). Dioecy may then, by chance, be associated with lineages exhibiting higher or lower diversification. The wide taxonomic distribution of dioecy allowed us to use a meta-analysis of multiple independent data sets to test this hypothesis. We used simulations of neutral character states to obtain the null distribution expected for our meta-analysis given the same set of empirically derived phylogenies. We found that such bimodal distributions would be expected by chance. Whatever these unexamined factors may be, we conclude that the effect of dioecy on diversification is not strong enough or consistent enough to have imparted a signal in the data analyzed.

We emphasize that several aspects of the data may have limited the power of our analyses and/or potentially biased our conclusions. First, on average only 42% of species within each genus were included (between 7 and 100% per genus). Although simulation studies showed that such coverage allows detection of differential diversification rates with only modest reductions in power (FitzJohn *et al.*, 2009), greater sampling within genera would have improved our ability to detect the effects of dioecy. Second, we assumed that the sampling fraction was the same for dioecious and nondioecious taxa, which might well be incorrect (e.g. if dioecious species are considered unusual and hence more often investigated and sequenced). However, for each genus, the fraction of dioecious species included on the trees (i.e. with GenBank data) was highly correlated with the fraction of dioecious species in the Tree of Sex database (Tree of Sex Consortium, 2014) ($r=0.94$; Pearson correlation coefficient; Fig. S8). That said, species whose sexual system was not specified in our search of the literature would be missing from both data sets. In particular, many authors might not mention if a species is hermaphroditic, given how common this sexual system is among flowering plants. Underreporting of hermaphroditism might artificially bias our results towards higher diversification of dioecious lineages. Third, we used only genera that contain at least two species of each sexual system. We may thus have preselected genera with higher than average origination rates and/or diversification rates of the derived state (dioecy for most of our genera). Moreover, by including only genera with both dioecious and nondioecious taxa (as required by BiSSE), a large number of clades that are entirely dioecious were ignored (e.g. nearly half of all dioecious angiosperm species are found within 34 clades that are entirely dioecious; Renner, 2014). Ideally, one should use the complete angiosperm phylogeny (or multiple

large-scale phylogenies) in which the sexual system distribution is representative of angiosperms as a whole. This, however, requires much higher coverage of sexual system information as well as methodologies that can appropriately handle the greater computational demands and biological heterogeneity of very large phylogenies. Fourth, our results may have been affected by the somewhat artificial nature of the binary classifications that we made. That said, we observed very similar results for the ‘strict’ and ‘broad’ classifications, even though the ‘broad’ classification contains *c.* 30% more dioecy species than the ‘strict’ classification.

Considering all of the potential biases discussed earlier, our result that dioecy is not consistently associated with either increased or decreased diversification must be considered preliminary and deserving of further research. We hope that the issues we have highlighted in obtaining these results will aid such further investigation.

Conclusions

A quarter of a century ago, Thomson & Brunet (1990) reviewed several hypotheses to explain the rarity of dioecy within flowering plants. These authors challenged, although could not reject, previously proposed correlations and hypotheses for the distribution of dioecy among angiosperms. The authors concluded that only with detailed phylogenetic analysis could these hypotheses be confirmed or rejected. Twenty-five years later, despite detailed investigations in several clades (Renner *et al.*, 2007; Volz & Renner, 2008), no phylogenetic study has examined more than a handful of plant groups, preventing general conclusions from being drawn. Even though we amassed the largest detailed data set to date and carried out careful and detailed phylogenetic analyses, we did not find a consistent effect of dioecy on diversification. Our analysis suggests that the effect of dioecy on speciation and extinction rates, if there is one, is relatively weak and easily overwhelmed by other factors affecting diversification. That is, dioecy does not consistently act as a strong driver of or brake on diversification. However, the possibility remains that future research, with more complete data and potentially improved methods, will clarify whether sexual systems affect rates of speciation and extinction in flowering plants.

Data accessibility

All phylogenies and sexual system assignments are available in DRYAD (<http://dx.doi.org/10.5061/dryad.7f4v1/1>).

Acknowledgements

We thank Susanne Renner for helpful discussions and suggestions. We also thank two anonymous reviewers for useful suggestions. This study was supported by an Edmond J. Safra postdoctoral fellowship awarded to N.S., a grant from the Israel Science Foundation (1265/12) to I.M., a grant from the United States Israel Binational Science Foundation (2013286) to I.M. and E.E.G., a Natural Sciences and Engineering Research Council grant to J.V. and S.P.O., and grants from the National Science

Foundation to T.-L.A. (DEB-1020523 and DEB-1241006) and E.E.G. (DEB-1120279). We have no competing interests.

Author contributions

N.S., I.M., S.P.O., E.E.G., J.V., T.-L.A., and R.M. conceived and designed the study. N.S., M.E. and L.G. conducted the analysis. N.S. and I.M. drafted the article. All authors contributed to and approved the article's final version for publication.

References

- Ashman TL, Kwok A, Husband BC. 2013. Revisiting the dioecy-polyploidy association: alternate pathways and research opportunities. *Cytogenetic and Genome Research* 140: 241–255.
- Baker HG. 1955. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* 9: 347–349.
- Bañuelos M-J, Obeso J-R. 2004. Resource allocation in the dioecious shrub *Rhamnus alpinus*: the hidden costs of reproduction. *Evolutionary Ecology Research* 6: 397–413.
- Barrett SC. 2002. The evolution of plant sexual diversity. *Nature Reviews: Genetics* 3: 274–284.
- Barrett SC. 2003. Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 358: 991–1004.
- Barrett SC. 2013. The evolution of plant reproductive systems: how often are transitions irreversible? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 280: 20130913.
- Beaulieu JM, O'Meara BC. 2015. Extinction can be estimated from moderately sized molecular phylogenies. *Evolution* 69: 1036–1043.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013. GenBank. *Nucleic Acids Research* 41(Database issue): D36–D42.
- Brach AR, Song H. 2006. eFloras: new directions for online floras exemplified by the flora of China project. *Taxon* 55: 188–192.
- Bull JJ, Charnov EL. 1985. On irreversible evolution. *Evolution* 39: 1149–1155.
- Charnov EL. 1982. *The theory of sex allocation*. Princeton, NJ, USA: Princeton University Press.
- Davis MP, Midford PE, Maddison W. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evolutionary Biology* 13: 38.
- Farrell BD, Dussourd DE, Mitter C. 1991. Escalation of plant defense – do latex and resin canals spur plant diversification. *American Naturalist* 138: 881–900.
- FitzJohn RG. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3: 1084–1092.
- FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* 58: 595–611.
- Hartigan JA, Hartigan PM. 1985. The dip test of unimodality. *The Annals of Statistics* 13: 70–84.
- Heilbut JC. 2000. Lower species richness in dioecious clades. *American Naturalist* 156: 221–241.
- Heilbut JC, Ilves KL, Otto SP. 2001. The consequences of dioecy for seed dispersal: modeling the seed-shadow handicap. *Evolution* 55: 880–888.
- Jarne P, Auld JR. 2006. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution* 60: 1816–1824.
- Käfer J, Boer H, Mousset S, Kool A, Dufay M, Marais G. 2014. Dioecy is associated with higher diversification rates in flowering plants. *Journal of Evolutionary Biology* 27: 1478–1490.
- Käfer J, Mousset S. 2014. Standard sister clade comparison fails when testing derived character states. *Systematic Biology* 63: 601–609.
- Kay KM, Voelckel C, Yang JY, Hufford KM, Kaska DD, Hodges SA. 2006. Floral characters and species diversification. In: Harder L, Barrett S, eds. *Ecology and evolution of flowers*. Oxford, UK: Oxford University Press, 311–325.
- Leslie AB, Beaulieu JM, Crane PR, Donoghue MJ. 2013. Explaining the distribution of breeding and dispersal syndromes in conifers. *Proceedings of the Royal Society B-Biological Sciences* 280: 20131812.
- Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13: 2178–2189.
- Maddison WP, FitzJohn RG. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology* 64: 127–136.
- Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* 56: 701–710.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257.
- McDaniel SF, Atwood J, Burleigh JG. 2013. Recurrent evolution of dioecy in bryophytes. *Evolution* 67: 567–572.
- Miller JS, Venable DL. 2000. Polyploidy and the evolution of gender dimorphism in plants. *Science* 289: 2335–2338.
- Mitter C, Farrell B, Wiegmann B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist* 132: 107.
- Montesinos D, de Luis M, Verdu M, Raventos J, Garcia-Fayos P. 2006. When, how and how much: gender-specific resource-use strategies in the dioecious tree *Juniperus thurifera*. *Annals of botany* 98: 885–889.
- Muenchow GE. 1987. Is dioecy associated with fleshy fruit. *American Journal of Botany* 74: 287–293.
- Nylander J. 2004. *MrAIC, pl 1.4.4. (Program distributed by the author.)*. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala University.
- Pannell JR, Barrett SCH. 1998. Baker's law revisited: reproductive assurance in a metapopulation. *Evolution* 52: 657–668.
- de Queiroz A. 1998. Interpreting sister-group tests of key innovation hypotheses. *Systematic Biology* 47: 710–718.
- Rabosky DL. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64: 1816–1824.
- Rabosky DL, Goldberg EE. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology* 64: 340–355.
- Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of botany* 101: 1588–1596.
- Renner SS, Beenken L, Grimm GW, Kocyan A, Ricklefs RE. 2007. The evolution of dioecy, heterodichogamy, and labile sex expression in *Acer*. *Evolution* 61: 2701–2719.
- Renner SS, Ricklefs RE. 1995. Dioecy and its correlates in the flowering plants. *American Journal of Botany* 82: 596–606.
- Rolland J, Condamine FL, Jiguet F, Morlon H. 2014. Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLoS Biology* 12: e1001775.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Schwander T, Crespi BJ. 2009. Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. *Molecular Ecology* 18: 28–42.
- Thomson JD, Barrett SCH. 1981. Selection for outcrossing, sexual selection, and the evolution of dioecy in plants. *American Naturalist* 118: 443–449.
- Thomson JD, Brunet J. 1990. Hypotheses for the evolution of dioecy in seed plants. *Trends in Ecology and Evolution* 5: 11–16.
- Tognetti R. 2012. Adaptation to climate change of dioecious plants: does gender balance matter? *Tree Physiology* 32: 1321–1324.
- Tree of Sex Consortium. 2014. Tree of Sex: a database of sexual systems. *Scientific Data* 1: 140015.
- Vamosi JC, Otto SP. 2002. When looks can kill: the evolution of sexually dimorphic floral display and the extinction of dioecious plants. *Proceedings of the Royal Society B: Biological Sciences* 269: 1187–1194.
- Vamosi JC, Otto SP, Barrett SC. 2003. Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. *Journal of Evolutionary Biology* 16: 1006–1018.

- Vamosi JC, Vamosi SM. 2004. The role of diversification in causing the correlates of dioecy. *Evolution* 58: 723–731.
- Vamosi JC, Vamosi SM. 2005. Present day risk of extinction may exacerbate the lower species richness of dioecious clades. *Diversity and Distributions* 11: 25–32.
- Villarreal JC, Renner SS. 2013. Correlates of monoecy and dioecy in hornworts, the apparent sister group to vascular plants. *BMC Evolutionary Biology* 13: 239.
- Volz SM, Renner SS. 2008. Hybridization, polyploidy, and evolutionary transitions between monoecy and dioecy in *Bryonia* (Cucurbitaceae). *American Journal of Botany* 95: 1297–1306.
- Wilson WG, Harder LD. 2003. Reproductive uncertainty and the relative competitiveness of simultaneous hermaphroditism versus dioecy. *American Naturalist* 162: 220–241.
- Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al.* 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 The distribution of $PP(r_N > r_D)$ values for both 'broad' and 'strict' definitions of dioecy.

Fig. S2 The distribution of $PP(r_N > r_D)$ values with constrained extinction rates.

Fig. S3 The distribution of $PP(\lambda_N > \lambda_D)$, $PP(\mu_N > \mu_D)$, and $PP(q_{ND} > q_{DN})$ values.

Fig. S4 The distribution of r_N and r_D summed over the 100.

Fig. S5 The observed distribution of $PP(r_N > r_D)$ values are associated with the percentage of nondioecious species within each genus.

Fig. S6 The distribution of simulated $PP(r_N > r_D)$ values across 100 simulated data sets using the 'free-tip frequency' approach.

Fig. S7 The distribution of simulated $PP(r_N > r_D)$ values across 100 simulated data sets using the 'matched tip frequency' approach.

Fig. S8 The fraction of dioecious species included on the trees is correlated with the fraction of dioecious species in the database.

Table S1 Plant sexual systems

Table S2 Genera in the study

Table S3 Rogue taxa identified using RogueNaRok

Methods S1 Phylogenetic reconstruction.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About New Phytologist

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <27 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**

New Phytologist Supporting Information Figs S1–S8, Tables S1–S3 and Methods S1

Article title: Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms

Authors: Niv Sabath, Tia-Lynn Ashman, Emma Goldberg, Ray Ming, Sarah P Otto, Jana Vamosi, Moshe Einhorn, Lior Glick, and Itay Mayrose

Article acceptance date: 25 August 2015

The following Supporting Information is available for this article:

Fig. S1 The distribution of $PP(r_N > r_D)$ values for both ‘broad’ and ‘strict’ definitions of dioecy.

Fig. S2 The distribution of $PP(r_N > r_D)$ values with constrained extinction rates.

Fig. S3 The distribution of $PP(\lambda_N > \lambda_D)$, $PP(\mu_N > \mu_D)$, and $PP(q_{ND} > q_{DN})$ values.

Fig. S4 The distribution of r_N and r_D summed over the 100.

Fig. S5 The observed distribution of $PP(r_N > r_D)$ values are associated with the percentage of nondioecious species within each genus.

Fig. S6 The distribution of simulated $PP(r_N > r_D)$ values across 100 simulated data sets using the ‘free-tip frequency’ approach.

Fig. S7 The distribution of simulated $PP(r_N > r_D)$ values across 100 simulated data sets using the ‘matched tip frequency’ approach.

Fig. S8 The fraction of dioecious species included on the trees is correlated with the fraction of dioecious species in the database.

Table S1 Plant sexual systems

Table S2 Genera in the study

Table S3 rogue taxa identified using RogueNaRok

Methods S1 Phylogenetic reconstruction.

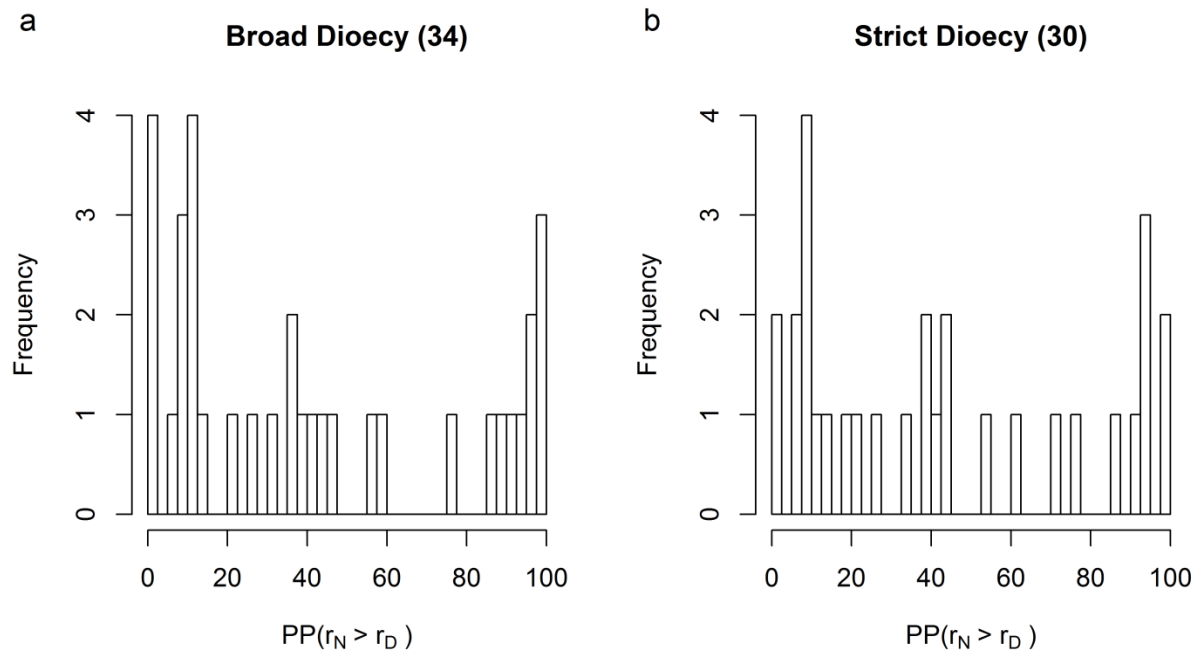


Fig. S1 The distribution of $PP(r_N > r_D)$ values for both (a) ‘broad’ and (b) ‘strict’ definitions of dioecy when the extinction rate is constrained to be equal for nondioecious and dioecious species ($\mu_N = \mu_D$). With either classification, the median of the distribution of $PP(r_N > r_D)$ values did not deviate from 50% ($P > 0.31$; one-sample Wilcoxon rank-sum test). In the ‘broad’ classification, the distribution exhibited a significant evidence of multimodality ($P = 0.037$; Hartigans’ dip test). In the ‘strict’ classification, the distribution did not exhibit a significant multimodality ($P = 0.17$).

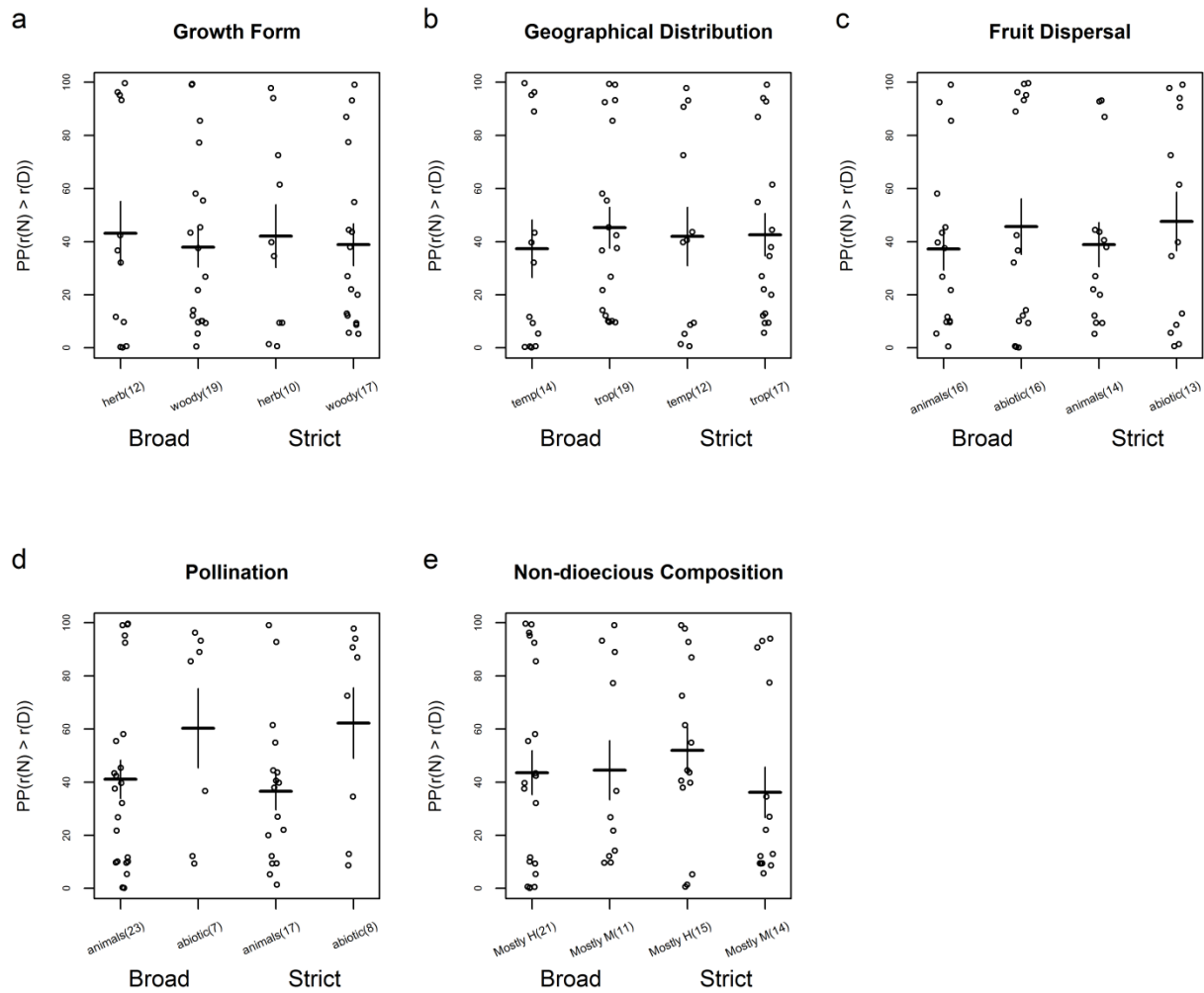


Fig. S2 The distribution of $PP(r_N > r_D)$ values inferred using BiSSE with constrained extinction rates ($\mu_N = \mu_D$) for genera classified according to (a) growth form, (b) geographical distribution, (c) mode of fruit dispersal, (d) mode of pollination, and (e) according to whether the nondioecious species in the genus were predominantly hermaphroditic ('Mostly H') vs monoecious ('Mostly M'). Each panel presents both 'broad' (left) and 'strict' (right) classifications of dioecy. Horizontal and vertical lines mark the sample means and \pm SE, respectively. For each particular trait, we tested whether the mean of the $PP(r_N > r_D)$ distribution was different from 50% using a one-sample Wilcoxon rank-sum test. All P -values were insignificant with $P > 0.72$.

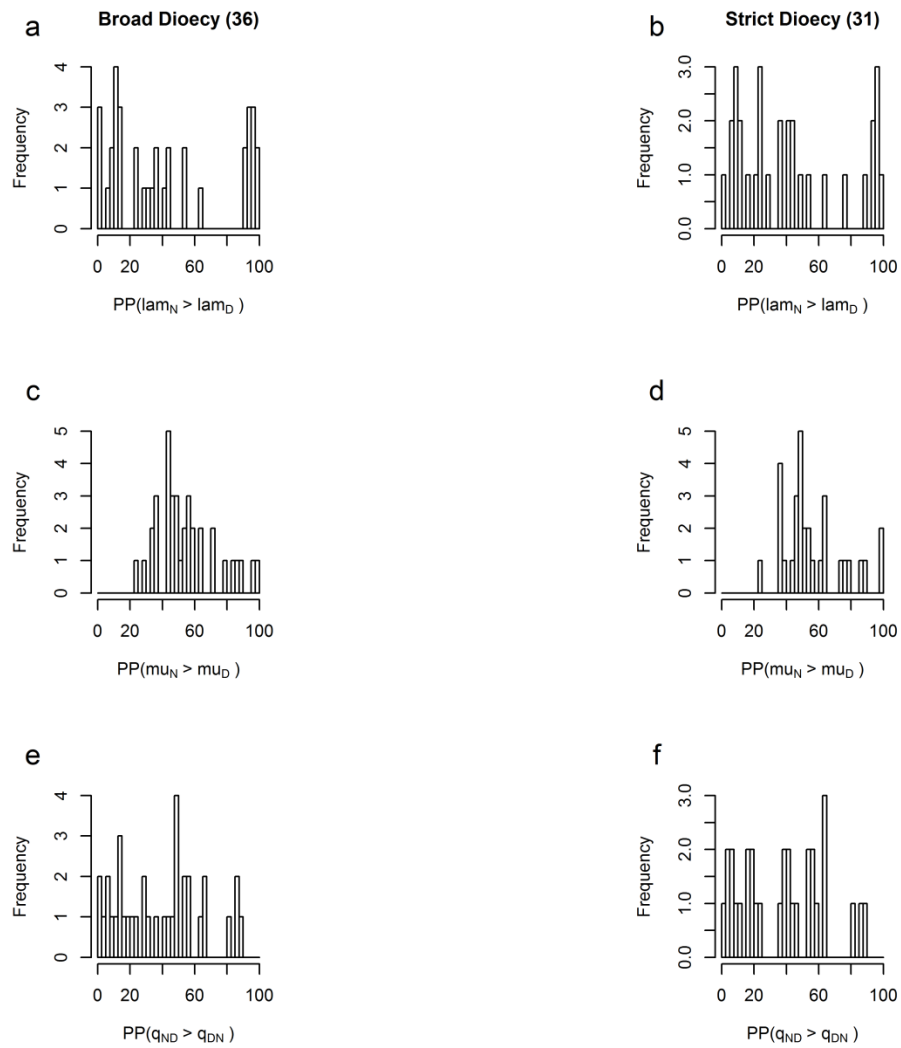


Fig. S3 The distribution of $PP(\lambda_N > \lambda_D)$, $PP(\mu_N > \mu_D)$, and $PP(q_{ND} > q_{DN})$ values for both (a, c, e) 'broad' and (b, d, f) 'strict' classifications of dioecy, based on BiSSE analyses of 36 and 31 genera, respectively.

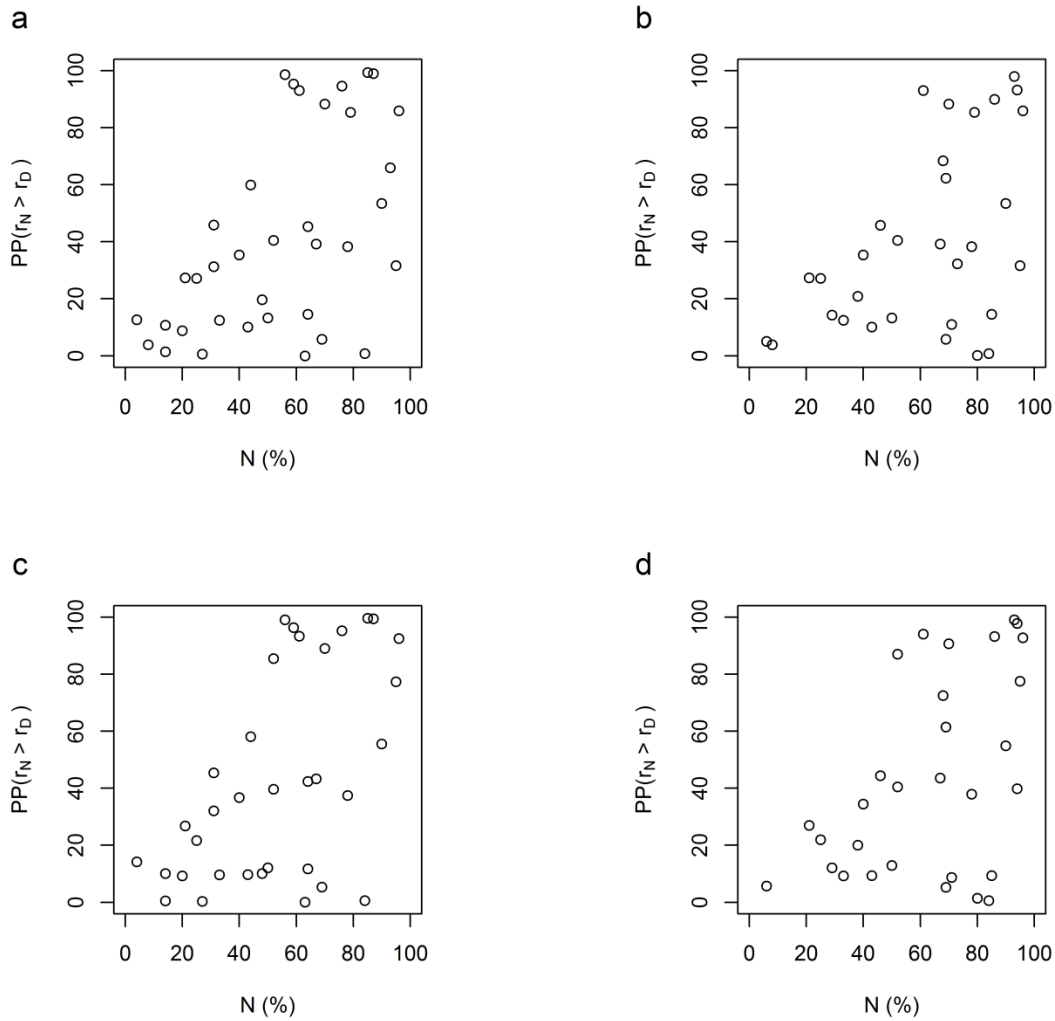


Fig. S5 The observed distribution of $PP(r_N > r_D)$ values inferred by BiSSE for each genus are associated with the percentage of nondioecious species within each genus. Results are shown for the main analysis (a, 'broad' classification $r = 0.55$, $P = 0.0005$, Pearson's correlation coefficient; b, 'strict' classification $r = 0.50$, $P = 0.004$, Pearson's correlation coefficient) and with constraining the extinction rates (c, 'broad' classification $r = 0.55$, $P = 0.0007$, Pearson's correlation coefficient; d, 'strict' classification $r = 0.48$, $P = 0.008$, Pearson's correlation coefficient).

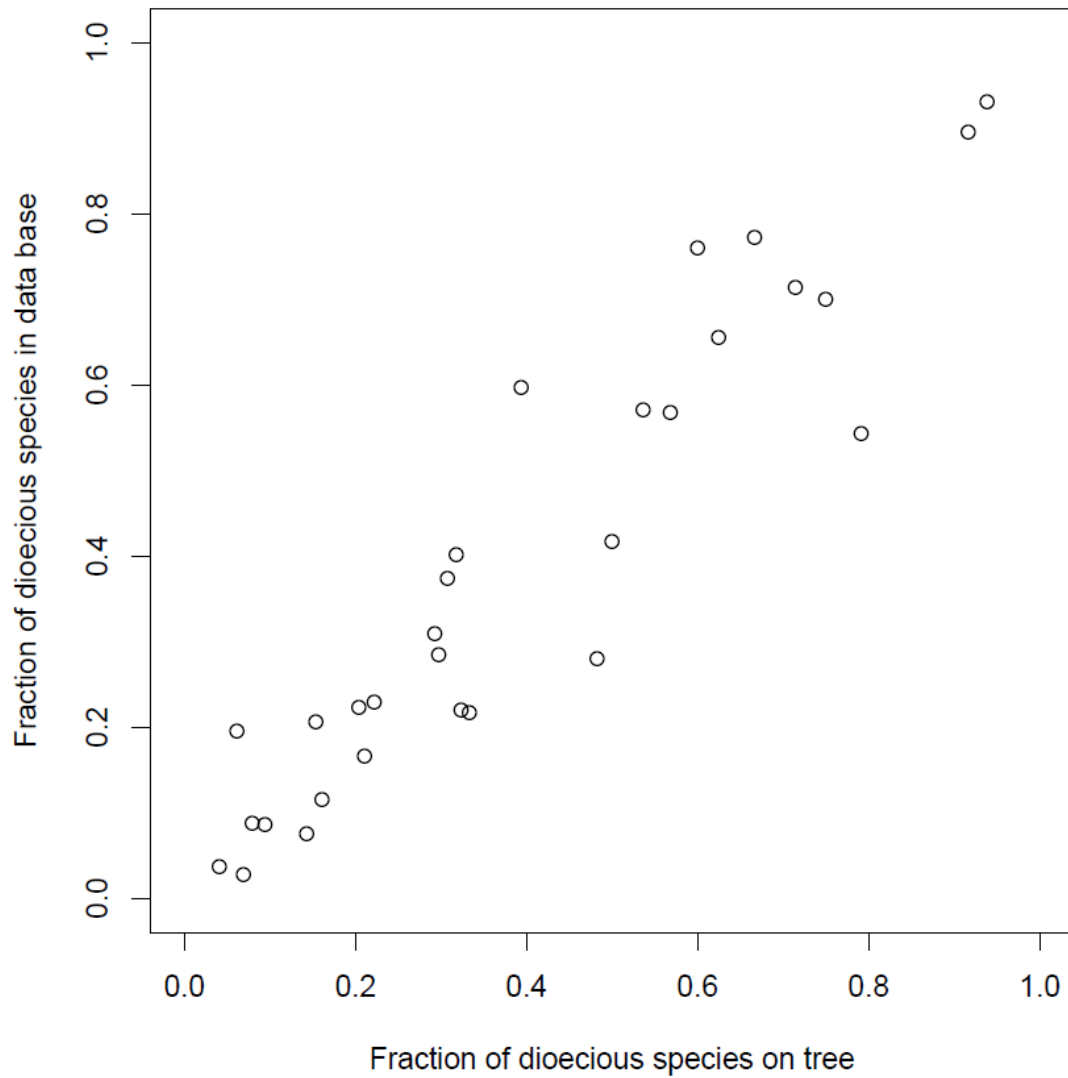


Fig. S8 The fraction of dioecious species included on the trees (i.e. with Genbank data) is correlated with the fraction of dioecious species in the Tree of Sex database¹ ($r = 0.94$, Pearson correlation coefficient).

Table S1 Plant sexual systems

Sexual system ^a	Description	Strict dioecy ^c	Broad dioecy ^c
Hermaphrodite	Plants whose flowers have both male and female parts	N	N
Monoecy	Plants with separate male and female flowers on the same plant	N	N
Dioecy	All plants are either female or male	D	D
Gynodioecy	Both female and hermaphrodite plants present	N	D
Androdioecy	Both male and hermaphrodite plants present	N	D
Gynomonoecy	Female and hermaphrodite flowers within a plant	N	N
Andromonoecy	Male and hermaphrodite flowers within a plant	N	N
Polygamodioecy ^d	Male, female, and hermaphrodite plants present	N	D
Polygamomonoecy	Male, female, and hermaphrodite flowers within a plant	N	N

^aTerms describe the sexual system morphologically; genetic or functional sexual system may differ.

^c‘Strict’ and ‘Broad’ dioecy classifications: D, dioecy; N, not dioecy. ^dSometimes called ‘trioecy’.

Table S2 Genera in the study

	Genus	Family	Species richness ^a	Source	TPL species richness ^b	No. sp. tree ^c	Strict N ^d	Strict D ^d	Broad N ^d	Broad D ^d	No. loci	Alignment length
1	<i>Allocasuarina</i>	Casuarinaceae	58	²	61	45	3	33	3	33	1	1800
2	<i>Asparagus</i>	Asparagaceae	200	³	235	29	15	14	15	14	3	9314
3	<i>Atriplex</i>	Amaranthaceae	300	⁴	332	71	26	11	26	11	2	1500
4	<i>Bauhinia</i>	Leguminosae	300	⁵	457	80	27	2	25	4	3	4847
5	<i>Begonia</i> ^{fh}	Begoniaceae	1400	⁶	1806	231	167	5	167	5	2	5064
6	<i>Boehmeria</i>	Urticaceae	65	⁷	168	21	3	3	3	3	2	4866
7	<i>Bursera</i> ^{fg}	Burseraceae	100	⁸	138	85	2	61	0	66	5	3211
8	<i>Cirsium</i> ^e	Compositae	250	⁹	963	59	47	0	34	11	3	5538
9	<i>Cordia</i>	Boraginaceae	350	¹⁰	484	91	26	3	26	3	2	1753
10	<i>Croton</i>	Euphorbiaceae	1250	¹¹	1280	256	211	9	211	9	2	2953
11	<i>Dodonaea</i>	Sapindaceae	64	¹²	77	51	3	46	2	47	2	1177
12	<i>Ficus</i> ^e	Moraceae	750	¹³	984	244	239	0	132	107	2	4890
13	<i>Elatostema</i>	Urticaceae	300	¹⁴	595	22	2	3	2	3	4	3750
14	<i>Fragaria</i>	Rosaceae	24	¹⁵	107	16	11	4	9	5	14	24670
15	<i>Fraxinus</i>	Oleaceae	43	¹⁶	88	43	29	12	8	33	5	5503
16	<i>Gaertnera</i>	Rubiaceae	68	¹⁷	87	28	21	6	21	6	4	3116
17	<i>Galium</i>	Rubiaceae	400	¹⁸	707	111	74	19	58	35	2	5482
18	<i>Garcinia</i>	Clusiaceae	400	¹⁹	429	52	6	10	5	11	3	3135
19	<i>Gunnera</i>	Gunneraceae	35	²⁰	71	20	15	4	15	4	2	3378
20	<i>Lepechinia</i> ^f	Lamiaceae	43	²¹	42	31	19	7	12	13	5	9488
21	<i>Leptinella</i> ^h	Compositae	42	²²	40	28	22	4	19	6	2	1812
22	<i>Mammillaria</i> ^e	Cactaceae	145	²³	654	110	106	0	104	2	6	7886
23	<i>Lycium</i> ^{fh}	Solanaceae	70	²⁴	223	63	58	5	55	7	1	3960
24	<i>Momordica</i>	Cucurbitaceae	59	²⁵	131	45	19	25	19	25	1	3330
25	<i>Ocotea</i>	Lauraceae	300	²⁶	585	43	19	22	18	23	2	2172
26	<i>Phoradendron</i>	Santalaceae	235	²⁷	323	24	5	19	5	19	3	1760
27	<i>Pilea</i>	Urticaceae	657	²⁸	694	97	20	13	20	13	2	1777
28	<i>Piper</i> ^{fh}	Piperaceae	2000	²⁹	2688	171	24	22	24	22	4	5377
29	<i>Poa</i>	Poaceae	575	³⁰	710	109	76	5	75	6	3	3928
30	<i>Rhus</i>	Anacardiaceae	35	³¹	462	29	4	10	2	12	4	6846
31	<i>Ribes</i>	Grossulariaceae	200	³²	457	66	45	21	45	21	4	2419
32	<i>Rubus</i>	Rosaceae	750	³³	6456	98	10	5	10	5	2	3240
33	<i>Rumex</i>	Polygonaceae	200	³⁴	646	36	23	11	19	13	2	5250
34	<i>Schisandra</i>	Schisandraceae	23	³⁵	27	12	4	8	4	8	3	6461
35	<i>Sidalcea</i> ^e	Malvaceae	25	³⁶	42	24	23	0	6	16	5	3319
36	<i>Silene</i> ^f	Caryophyllaceae	700	³⁷	1603	181	104	7	81	14	4	8068
37	<i>Siparuna</i>	Siparunaceae	72	³⁸	65	13	3	9	3	9	1	403
38	<i>Solanum</i>	Solanaceae	1500	³⁹	2676	400	291	11	291	11	3	7754
39	<i>Thalictrum</i>	Ranunculaceae	190	⁴⁰	622	91	73	14	73	14	3	7417
40	<i>Valeriana</i> ^f	Caprifoliaceae	176	⁴¹	551	102	48	5	17	37	3	7745
41	<i>Vitis</i>	Vitaceae	60	⁴²	436	30	18	3	3	18	3	9285

42	<i>Wurmbea</i>	Colchicaceae	48	⁴³	50	22	9	4	9	5	2	6985
----	----------------	--------------	----	---------------	----	----	---	---	---	---	---	------

^aThe total number of species per genus. ^bThe number of species per genus (accepted + unresolved) according to TPL 1.1. ^cThe number of species on the phylogenetic tree. ^dThe number of species on the phylogenetic tree in 'Strict' and 'Broad' dioecy classifications: D, dioecy; N, not dioecy. ^eExcluded from the 'Strict' analysis based on insufficient sampling. ^fExcluded from the 'Strict' analysis based on sensitivity to prior choice. ^gExcluded from the 'Broad' analysis based on insufficient sampling. ^hExcluded from the 'Broad' analysis based on sensitivity to prior choice.

Methods S1 Phylogenetic reconstruction.

Our phylogenetic reconstruction pipeline was composed of the following steps.

(1) For each genus all the sequences found in GenBank were downloaded using NCBI e-utilities (Entrez).

(2) Name resolution was performed using Taxonome⁴⁵, which provides the ability to match synonymous taxon names to accepted names, accounting for likely misspellings (e.g. when the gender of the species name does not match the genus name). As the underlying database for names, we used a local repository of synonymous and accepted names that were created based on The Plant List (V1.1; <http://www.theplantlist.org/>) and Solanaceae Source (<http://solanaceaesource.org/>).

(3) All non-ITS sequences were clustered into orthology groups using orthoMCL v2.0.3⁴⁶. The clustering of ITS sequences is detailed in step (5) below. To ensure that all sequences in a cluster are orthologous, we added an extra filtering step to the orthoMCL pipeline so that no hits obtained by BLAST all-vs-all contain sequences where one sequence is more than twice the size of the other sequence. (This eliminated cases where a long sequence, e.g. the whole chloroplast genome, is mapped to several unrelated loci.)

(4) When multiple accessions were found for a locus from a species, a single representative accession was chosen by picking the accession with the highest average bitscore based on BLAST all-against-all.

(5) Sequences belonging to the internal transcribed spacer (ITS) region were treated differently because the ITS1 and ITS2 loci can be very similar, causing clustering tools such as orthoMCL to cluster these two distinct DNA regions together. Thus, clustering of ITS sequences was based on the sequence description as specified in GenBank rather than relying solely on the sequence

data. Sequences containing the ITS regions were first classified to three nonoverlapping groups: sequences containing ITS1 region only, sequences containing ITS2 region only, and sequences containing ITS1 followed by ITS2. These sequences were then aligned using a 2-step alignment strategy as suggested by Katoh *et al.* (2013)⁴⁷. First, only sequences containing both ITS1 and ITS2 were aligned thereby creating a backbone alignment. Then, ITS1 and ITS2 sequences were iteratively added to the backbone alignment using the add-fragments option of MAFFT⁴⁷.

(6) For each genus, an outgroup was selected as follows. For phylogenies based on a single locus, we automatically located outgroup sequences by looking for similar sequences to those in the cluster, using BLAST, and choosing the sequence that was most similar to the sequences within the cluster but outside the genus (i.e. a non-ingroup sequence was not considered a candidate outgroup if it was more similar to the query sequence than some ingroup sequences). The BLAST search was performed against the nonredundant database of GenBank using as a query a representative sequence in the cluster. (This sequence was chosen by executing BLAST all-against-all for the sequences in the cluster and selecting the sequence with the highest average bitscore.) In addition, we verified that the length of the outgroup sequence was at most twice the length of the representative sequence and at least half its length. In case this criterion was not met, the next sequence in the BLAST search was chosen. In order to get enough results from the BLAST search, we set the number of returned BLAST results (-max_target_seqs option) to the number of the sequences in the cluster multiplied by 50, with a minimum of 1000 results.

The procedure for locating a common outgroup for phylogenies based on many loci was more challenging, because a taxon that is a good outgroup for one locus may not have data for other loci. Thus, for each locus, a list of candidate outgroup species was obtained as detailed above. Each of the candidate outgroup species was then examined, looking for the one in common to as many loci as possible. In rare cases where a single appropriate outgroup was not found, a ‘chimera’ of sequences belonging to different species from a single outgroup genus was considered. For example, when seeking an outgroup for the genus *Asparagus*, the pipeline constructed an outgroup that is a concatenation of sequences from two different species

belonging to the genus *Dracaena* (each individual species having a small number of sequenced loci available). Other genera with chimera outgroup are *Begonia* and *Viscum*.

(7) Multiple sequence alignment of each orthologous gene group (cluster) along with its assigned outgroup was performed using MAFFT⁴⁷. The ‘—adjustdirection’ option was used to verify that the direction of all the sequences was the same. Alignment confidence was tested using GUIDANCE⁴⁸ with default parameters (removing columns or rows with reliability scores below 0.93 and 0.6, respectively). GUIDANCE can filter both unreliable sequences (rows) and unreliable columns. If unreliable rows were found, they were removed from the sequence list and GUIDANCE was executed again to filter unreliable columns given the rows-filtered alignment. Two species were filtered out in this way: *Croton skutchii* and *Momordica welwitschii*.

(8) The best-supported model of sequence evolution was determined using MrAIC⁴⁹ for each locus and each genus separately. The separate alignments were then concatenated to form a single concatenated alignment.

(9) A partitioned Bayesian phylogenetic reconstruction was performed for each genus with MrBayes v3.2.1⁵⁰, using two independent runs, each with one cold and three heated chains. A relaxed clock model was used to ensure that the resulting trees were ultrametric, yet allowing for rate variation across lineages. To this end, the birth–death clock model (chosen because it was found to fit the data better than the uniform and the coalescence models; not shown) was applied together with the independent gamma rates model for across lineages variation. A partial topological constraint was enforced, rooting the tree with the taxon used as the outgroup, as identified above. We first ran MrBayes for 500,000 steps to identify taxa whose position in the tree was highly uncertain (‘rogue taxa’) using RogueNaRok⁴⁴ with default parameters and with raw improvement threshold of 0.1. After removing rogue taxa (listed in Table S3), we ran MrBayes for 2,000,000 steps. Trees were sampled every 2,000 generations, with the initial 25% of the sample treated as burn-in and discarded from the final sample. 100

of these trees were then randomly sampled without replacement from these 1000 trees and outgroup species removed for use in BiSSE.

References

- 1 Tree of Sex Consortium. Tree of Sex: A database of sexual systems. *Scientific Data* **1**, 140015, doi:10.1038/sdata.2014.15 (2014).
- 2 Steane, D. A., Wilson, K. L. & Hill, R. S. Using matK sequence data to unravel the phylogeny of Casuarinaceae. *Molecular Phylogenetics and Evolution* **28**, 47-59, doi:10.1016/S1055-7903(03)00028-9 (2003).
- 3 Kubota, S., Konno, I. & Kanno, A. Molecular phylogeny of the genus *Asparagus* (Asparagaceae) explains interspecific crossability between the garden asparagus (*A. officinalis*) and other *Asparagus* species. *Theoretical and Applied Genetics* **124**, 345-354, doi:10.1007/s00122-011-1709-2 (2012).
- 4 Kadereit, G., Mavrodiev, E. V., Zacharias, E. H. & Sukhorukov, A. P. Molecular Phylogeny of Atripliceae (Chenopodioideae, Chenopodiaceae): Implications for Systematics, Biogeography, Flower and Fruit Evolution, and the Origin of C-4 Photosynthesis. *American Journal of Botany* **97**, 1664-1687, doi:10.3732/ajb.1000169 (2010).
- 5 Hao, G., Zhang, D. X., Zhang, M. Y., Guo, L. X. & Li, S. J. Phylogenetics of *Bauhinia* subgenus *Phanera* (Leguminosae : Caesalpinioideae) based on ITS sequences of nuclear ribosomal DNA. *Botanical Bulletin of Academia Sinica* **44**, 223-228 (2003).
- 6 Forrest, L. L., Hughes, M. & Hollingsworth, P. M. A phylogeny of *Begonia* using nuclear ribosomal sequence data and morphological characters. *Systematic Botany* **30**, 671-682, doi:Doi 10.1600/0363644054782297 (2005).
- 7 Ma Shu, Z., Jiarui, C., Friis, I. & Wilmot-Dear, C. M. *Flora of China*, <<http://flora.huh.harvard.edu/china/PDF/PDF05/Boehmeria.pdf>> (2003) [accessed 1 January 2015].
- 8 Becerra, J. X. & Venable, D. L. Nuclear ribosomal, DNA phylogeny and its implications for evolutionary trends in Mexican *Bursera* (Burseraceae). *American Journal of Botany* **86**, 1047-1057, doi:Doi 10.2307/2656622 (1999).
- 9 Häffner, E. & Hellwig, F. H. Phylogeny of the tribe Cardueae (Compositae) with emphasis on the subtribe Carduinae: an analysis based on ITS sequence data. *Willdenowia* **29**, 27-39 (1999).

- 10 Gottschling, M., Miller, J. S., Weigend, M. & Hilger, H. H. Congruence of a phylogeny of Cordiaceae (Boraginales) inferred from ITS1 sequence data with morphology, ecology, and biogeography. *Annals of the Missouri Botanical Garden* **92**, 425-437 (2005).
- 11 Berry, P. E., Hipp, A. L., Wurdack, K. J., Van Ee, B. & Riina, R. Molecular phylogenetics of the giant genus *Croton* and tribe Crotonaeae (Euphorbiaceae sensu stricto) using its and trnL-trnF DNA sequence data. *American Journal of Botany* **92**, 1520-1534, doi:DOI 10.3732/ajb.92.9.1520 (2005).
- 12 Harrington, M. *Phylogeny and evolutionary history of Sapindaceae and Dodonaea*, PhD thesis, James Cook University, Townsville, Australia, (2008).
- 13 Ronsted, N., Weiblen, G. D., Savolainen, V. & Cook, J. M. Phylogeny, biogeography, and ecology of *Ficus* section *Malvanthera* (Moraceae). *Molecular Phylogenetics and Evolution* **48**, 12-22, doi:10.1016/j.ympev.2008.04.005 (2008).
- 14 Hadiah, J. T. & Conn, B. J. Usefulness of morphological characters for infrageneric classification of *Elatostema* (Urticaceae). *Blumea* **54**, 181-191, doi:10.3767/000651909X476139 (2009).
- 15 DiMeglio, L. M., Staudt, G., Yu, H. R. & Davis, T. M. A Phylogenetic Analysis of the Genus *Fragaria* (Strawberry) Using Intron-Containing Sequence from the ADH-1 Gene. *Plos One* **9**, e102237.
- 16 Hinsinger DD, Basak J, Gaudeul M, Cruaud C, Bertolino P, Frascaria-Lacoste N, Bousquet J. The Phylogeny and Biogeographic History of Ashes (*Fraxinus*, Oleaceae) Highlight the Roles of Migration and Vicariance in the Diversification of Temperate Trees. *Plos One* **8**, e80431.
- 17 Malcomber, S. T. Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: Evidence of a rapid radiation in a widespread, morphologically diverse genus. *Evolution* **56**, 42-57 (2002).
- 18 Mitova, M. I., Anchev, M. E., Handjieva, N. V. & Popova, S. S. Iridoid patterns in *Galium* L. and some phylogenetic considerations. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences* **57**, 226-234 (2002).

- 19 Santosa, E. Phylogenetic analysis of mangosteen (*Garcinia mangostana* L.) and its relatives based on morphological and inter simple sequence repeat (ISSR) markers. *SABRAO Journal of Breeding and Genetics* **45**, 478-490 (2013).
- 20 Wanntorp, L., Wanntorp, H. E. & Kallersjo, M. Phylogenetic relationships of *Gunnera* based on nuclear ribosomal DNA ITS region, rbcL and rps16 intron sequences. *Systematic Botany* **27**, 512-521 (2002).
- 21 Drew, B. T. & Sytsma, K. J. The South American radiation of *Lepechinia* (Lamiaceae): phylogenetics, divergence times and evolution of dioecy. *Botanical Journal of the Linnean Society* **171**, 171-190, doi:10.1111/j.1095-8339.2012.01325.x (2013).
- 22 Himmelreich, S., Breitwieser, I. & Oberprieler, C. Phylogeny, biogeography, and evolution of sex expression in the southern hemisphere genus *Leptinella* (Compositae, Anthemideae). *Molecular Phylogenetics and Evolution* **65**, 464-481, doi:10.1016/j.ympev.2012.07.001 (2012).
- 23 Butterworth, C. A. & Wallace, R. S. Phylogenetic studies of *Mammillaria* (Cactaceae) - insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *American Journal of Botany* **91**, 1086-1098, doi:DOI 10.3732/ajb.91.7.1086 (2004).
- 24 Fukuda, T., Yokoyama, J. & Ohashi, H. Phylogeny and biogeography of the genus *Lycium* (Solanaceae): Inferences from chloroplast DNA sequences. *Molecular Phylogenetics and Evolution* **19**, 246-258, doi:10.1006/mpev.2001.0921 (2001).
- 25 Schaefer, H. & Renner, S. S. A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia. *Molecular Phylogenetics and Evolution* **54**, 553-560, doi:10.1016/j.ympev.2009.08.006 (2010).
- 26 Chanderbali, A. S., van der Werff, H. & Renner, S. S. Phylogeny and historical biogeography of Lauraceae: Evidence from the chloroplast and nuclear genomes. *Annals of the Missouri Botanical Garden* **88**, 104-134, doi:Doi 10.2307/2666133 (2001).

- 27 Coder, K. American mistletoe (*Phoradendron serotinum* var. *serotinum*) infection in trees. Athens, GA, USA: Warnell School, University of Georgia, University Outreach Publication (2008).
- 28 Monro, A. K. Revision of species-rich genera: A phylogenetic framework for the strategic revision of *Pilea* (Urticaceae) based on cpDNA, nrDNA, and morphology. *American Journal of Botany* **93**, 426-441, doi:Doi 10.3732/Ajb.93.3.426 (2006).
- 29 Jaramillo MA, Callejas R, Davidson C, Smith JF, Stevens AC, Tepe EJ. A Phylogeny of the Tropical Genus *Piper* Using ITS and the Chloroplast Intron psbJ-petA. *Systematic Botany* **33**, 647-660, doi:Doi 10.1600/036364408786500244 (2008).
- 30 Gillespie, L. J. & Soreng, R. J. A phylogenetic analysis of the bluegrass genus *Poa* based on cpDNA restriction site data. *Systematic Botany* **30**, 84-105, doi:Doi 10.1600/0363644053661940 (2005).
- 31 Yi, T. S., Miller, A. J. & Wen, J. Phylogeny of *Rhus* (Anacardiaceae) based on sequences of nuclear Nia-i3 intron and chloroplast trnC-trnD. *Systematic Botany* **32**, 379-391, doi:Doi 10.1600/036364407781179635 (2007).
- 32 Sinters, A. E. & Soltis, D. E. Phylogenetic relationships in *Ribes* (Grossulariaceae) inferred from ITS sequence data. *Taxon* **52**, 51-66, doi:Doi 10.2307/3647301 (2003).
- 33 Alice, L. A. & Campbell, C. S. Phylogeny of *Rubus* (Rosaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *American Journal of Botany* **86**, 81-97, doi:Doi 10.2307/2656957 (1999).
- 34 Talavera M, Balao F, Casimiro-Soriguer R, Ortiz MÁ, Terrab A, Arista M, Ortiz PL, Stuessy TF, Talavera S. Molecular phylogeny and systematics of the highly polymorphic *Rumex bucephalophorus* complex (Polygonaceae). *Molecular Phylogenetics and Evolution* **61**, 659-670, doi:10.1016/j.ympev.2011.08.005 (2011).
- 35 Denk, T. & Oh, I. C. Phylogeny of Schisandraceae based on morphological data: evidence from modern plants and the fossil record. *Plant Systematics and Evolution* **256**, 113-145, doi:10.1007/s00606-005-0327-3 (2005).
- 36 Andreasen, K. & Baldwin, B. G. Reexamination of relationships, habital evolution, and phylogeography of checker mallows (*Sidalcea*; Malvaceae) based on molecular

- phylogenetic data. *American Journal of Botany* **90**, 436-444, doi:Doi 10.3732/Ajb.90.3.436 (2003).
- 37 Ghahremaninejad, F., Angaji, A., Etemad, M., Vahidynia, F. & Attar, F. Molecular taxonomy and phylogeny of *Silene* species (Caryophyllaceae) using DNA-based markers. *Journal of Applied Environmental and Biological Sciences* **4**, 125-132 (2014).
- 38 Renner, S. S., Schwarzbach, A. E. & Lohmann, L. Phylogenetic position and floral function of *Siparuna* (Siparunaceae : Laurales). *International Journal of Plant Sciences* **158**, S89-S98, doi:Doi 10.1086/297509 (1997).
- 39 Weese, T. L. & Bohs, L. A three-gene phylogeny of the genus *Solanum* (Solanaceae). *Systematic Botany* **32**, 445-463, doi:Doi 10.1600/036364407781179671 (2007).
- 40 Soza, V. L., Brunet, J., Liston, A., Smith, P. S. & Di Stilio, V. S. Phylogenetic insights into the correlates of dioecy in meadow-rues (*Thalictrum*, Ranunculaceae). *Molecular Phylogenetics and Evolution* **63**, 180-192, doi:10.1016/j.ympev.2012.01.009 (2012).
- 41 Bell, C. D. & Donoghue, M. J. Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians. *Organisms Diversity & Evolution* **5**, 147-159, doi:10.1016/j.ode.2004.10.014 (2005).
- 42 Wan YZ, Schwaninger HR, Baldo AM, Labate JA, Zhong G-Y, Simon CJ. A phylogenetic analysis of the grape genus (*Vitis* L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. *BMC Evolutionary Biology* **13**, 141 (2013).
- 43 Case, A. L., Graham, S. W., Macfarlane, T. D. & Barrett, S. C. H. A phylogenetic study of evolutionary transitions in sexual systems in Australasian *wurmbea* (Colchicaceae). *International Journal of Plant Sciences* **169**, 141-156, doi:10.1086/523368 (2008).
- 44 Aberer, A. J., Krompass, D. & Stamatakis, A. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology* **62**, 162-166, doi:10.1093/sysbio/sys078 (2013).
- 45 Kluyver, T. A. & Osborne, C. P. Taxonome: a software package for linking biological species data. *Ecology and Evolution* **3**, 1262-1265, doi:10.1002/ece3.529 (2013).

- 46 Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**, 2178-2189, doi:10.1101/gr.1224503 (2003).
- 47 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
- 48 Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Research* **38**, W23-28, doi:10.1093/nar/gkq443 (2010).
- 49 Nylander J. 2004. *MrAIC.pl* 1.4.4. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden (2004).
- 50 Ronquist F, Teslendo M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539-542, doi:10.1093/sysbio/sys029 (2012).