Interaction among ploidy, breeding system and lineage diversification

Rosana Zenil-Ferguson[1] (iD), J. Gordon Burleigh[2], William A. Freyman[3], Boris Igić[4] (iD), Itay Mayrose[5] and Emma E. Goldberg[3]

[1]Department of Biology, University of Hawai'i Mānoa, Honolulu, HI 96822, USA; [2]Department of Biology, University of Florida, Gainesville, FL 32611, USA; [3]Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, MN 55108, USA; [4]Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA; [5]School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv 6997801, Israel

Author for correspondence:
Rosana Zenil-Ferguson
Tel: +1 808 956 8241
Email: roszenil@hawaii.edu

Summary

• If particular traits consistently affect rates of speciation and extinction, broad macroevolutionary patterns can be interpreted as consequences of selection at high levels of the biological hierarchy. Identifying traits associated with diversification rates is difficult because of the wide variety of characters under consideration and the statistical challenges of testing for associations from comparative phylogenetic data. Ploidy (diploid vs polyploid states) and breeding system (self-incompatible vs self-compatible states) are both thought to be drivers of differential diversification in angiosperms.

• We fit 29 diversification models to extensive trait and phylogenetic data in Solanaceae and investigate how speciation and extinction rate differences are associated with ploidy, breeding system, and the interaction between these traits.

• We show that diversification patterns in Solanaceae are better explained by breeding system and an additional unobserved factor, rather than by ploidy. We also find that the most common evolutionary pathway to polyploidy in Solanaceae occurs via direct breakdown of self-incompatibility by whole genome duplication, rather than indirectly via breakdown followed by polyploidization.

• Comparing multiple stochastic diversification models that include complex trait interactions alongside hidden states enhances our understanding of the macroevolutionary patterns in plant phylogenies.

> Among life history traits, reproductive characters that determine mating patterns are perhaps the most influential in governing macroevolution.
>
> Barrett *et al.* (1996)

Introduction

Species accumulate across the tree of life at different rates. One possible explanation for this phenomenon is that various traits differentially affect rates of diversification. Dramatic increases in phylogenetic and phenotypic data, along with methodological advances, have greatly accelerated the search for traits that influence diversification. Nevertheless, identifying focal traits associated with rates of speciation and extinction remains a challenge (Maddison & FitzJohn, 2015; Rabosky & Goldberg, 2015, 2017; Beaulieu & O'Meara, 2016). One difficulty is that speciation and extinction likely do not depend on a single character, so the biological and environmental contexts in which traits occur can lead to complex interactions that affect lineage diversification (Beaulieu & O'Meara, 2016; Caetano *et al.*, 2018; Herrera-Alsina *et al.*, 2018). Consequently, examining the association of only one character with diversification patterns can be

misleading. Here, we embrace this challenge by jointly investigating two characters thought to influence speciation and extinction rates – ploidy level and breeding system – while allowing for their interactions, and other confounding factors, to change diversification. We also test whether adding one more trait and increasing model complexity is worthwhile.

Polyploidization is a remarkably common mutation in plants (Husband *et al.*, 2013; Zenil-Ferguson *et al.*, 2017). The widespread variation in ploidy has long been considered a salient feature of flowering plant lineages (Stebbins, 1938). An increase in ploidy can alter many traits and affect a variety of evolutionary and ecological processes (Ramsey & Schemske, 2002; Sessa, 2019). At shallow evolutionary time scales, polyploids were found to have an overall lower net diversification rate than diploids across many vascular plant clades (Mayrose *et al.*, 2011, 2015). However, recent genomic studies have inferred numerous paleo-polyploidizations, including some preceding the emergence of highly diverse plant clades (Soltis *et al.*, 2014; Landis *et al.*, 2018), suggesting that whole genome duplications may have played an important role driving innovation and diversification in plants. Evidence of paleo-polyploidy within the genomes of diploid extant plants also implies pervasive diploidization, the

return of polyploids to the diploid state, throughout the angiosperm phylogeny (Dodsworth *et al.*, 2015; Soltis *et al.*, 2015). Our analyses re-examine the association between ploidy and lineage diversification by extending the approach of Mayrose *et al.* (2011, 2015) to include transitions from polyploid to diploid states and potential unobserved factors affecting diversification patterns.

Breeding system shifts – changes in the collection of physiological and morphological traits that determine the likelihood that any two gametes unite – are also remarkably common and crucially affect the distribution and amount of genetic variation in populations (Barrett, 2013). In particular, a variety of genetic self-incompatibility (SI) systems causes plants to reject their own pollen, and the loss of such mechanisms, yielding self-compatibility (SC), is a commonly observed transition in flowering plant evolution (Stebbins, 1974; Igić *et al.*, 2008). Previous analyses reported higher rates of diversification for SI than for SC lineages in Solanaceae (Goldberg *et al.*, 2010). Similarly, heterostylous SI lineages in Primulaceae seem to diversify faster than SC lineages (de Vos *et al.*, 2014), as do outcrossing lineages in Onagraceae (Freyman & Höhna, 2019). Although these findings suggest a consistent macroevolutionary role of breeding system, it is unlikely to be the sole character determining lineage diversification. We investigate the relationship of breeding system to speciation and extinction rates in the context of ploidy and other unobserved factors.

Polyploidy and self-fertilization are widely thought to be associated (Stebbins, 1950). Whole genome duplication may facilitate the transition to selfing by masking inbreeding depression, or self-fertilization may facilitate establishment of polyploids by avoiding the lower fitness of triploids (Levin, 1975; Ramsey & Schemske, 1998; Barringer, 2007; Barrett, 2008; Husband *et al.*, 2008). Additionally, in RNase-based gametophytic SI systems, polyploidization directly causes the loss of SI (Stout & Chandler, 1942; Lewis, 1947). In these systems, SI occurs because haploid self pollen grains, with one S-allele at the locus controlling the SI response, are unable to detoxify the S-RNase produced by the same S-allele in the style (Kubo *et al.*, 2010). The unreduced pollen of diploids, however, can contain two S-alleles expressed in pollen, which jointly provide the ability to detoxify the S-RNases produced by any maternal genotype (Entani *et al.*, 1999; Tsukamoto *et al.*, 2005; Kubo *et al.*, 2010). Initial mutant individuals with pollen containing doubled haploid genomes are consequently capable of self-fertilisation, with exceedingly few exceptions (Hauck *et al.*, 2002; Nunes *et al.*, 2006). RNase-based SI is regarded as ancestral in eudicots (Igić & Kohn, 2001; Steinbachs & Holsinger, 2002), and it is expressed in all SI species of Solanaceae examined to date. The absence of SI polyploids in this family yields a strong correlation between ploidy and breeding system (Robertson *et al.*, 2011).

We address two macroevolutionary questions about the correlated evolution of ploidy and breeding system. First, we investigate their joint influence on rates of speciation and extinction. Each character alone is thought to be associated with differential lineage diversification but, if their effects are not additive, studying them separately may not reveal their combined effect.

Second, we examine the order of transitions in the two characters. Evolution commonly proceeds from diploid to polyploid, and from SI to SC states, but there are two paths by which diploid SI lineages can eventually become SC polyploids. Loss of SI in diploids could be directly caused by polyploidization (as explained above, for RNase-based SI systems), resulting in a one-step pathway to SC polyploids. Alternatively, SI diploids could first transition to SC without an increase in ploidy, and subsequently undergo polyploidization, resulting in a two-step pathway to SC polyploids. Robertson *et al.* (2011) compared the contributions of these two paths, finding that evolution from SI diploids to SC polyploids is more likely to proceed via the one-step pathway over short timescales, but via the two-step pathway over long timescales. They considered only transitions among the states, however, and we investigate whether these results hold true when allowing for differences in lineage diversification.

In the present study, we employ an extended framework of state speciation and extinction models, which simultaneously model transitions between the discrete states of a trait and different rates of speciation and extinction associated with each of those states ('SSE' models; Maddison *et al.*, 2007; FitzJohn, 2012). We start by fitting binary state speciation and extinction models to ploidy and breeding system independently (Maddison *et al.*, 2007). We follow by fitting models that incorporate hidden states, which reduce the chance that the effect of the focal trait (ploidy or breeding system) on diversification is found to be significant when in reality, it may simply represent background heterogeneity in the diversification process that produces the phylogenetic patterns (Beaulieu & O'Meara, 2016). We compare the proposed models against their character-independent counterparts (Beaulieu & O'Meara, 2016) to investigate whether something other than ploidy or breeding system better explains the diversification process. Next, we model ploidy and breeding system jointly to assess their combined influence on diversification, with or without an additional hidden character. Using the ploidy and breeding system model without hidden states, we quantify the relative contributions of the two pathways from SI diploids to SC polyploids. We also aggregate states within these joint models of ploidy and breeding system in order to test whether increasing model complexity from one trait to two traits significantly improves our understanding of the diversification process. Furthermore, we extend all the models involving ploidy to investigate the potential effects of including diploidization. Our results highlight the importance of considering nonadditive effects of traits on net diversification rates under the presence of unobserved factors, in order to detect strong biologically driven processes dictating diversification patterns.

Materials and Methods

Data

Chromosome number data were obtained for all Solanaceae taxa in the Chromosome Counts Database (CCDB; Rice *et al.*, 2015), and the *c.* 14,000 records were curated semiautomatically using the CCDBCURATOR R package (Rivero *et al.*, 2019). CCDB

contains records from original sources that have multiple complex symbol patterns denoting multivalence, or irregularities of chromosome counts. After a first round of automatic cleaning, we examined results by hand and corrected records as necessary. Our hand-curated records were also contrasted against the ploidy dataset from Robertson *et al.* (2011), original references therein, and against ploidy data in the *C*-value DNA dataset from Bennett & Leitch (2012). By comparing three different sources of information, we were able to code taxa as diploid, *D* or polyploid, *P*. For the majority of species, ploidy was assigned according to information from the original publications included in the *C*-value DNA dataset (Bennett & Leitch, 2012). For taxa without ploidy information but with information about chromosome number, we assigned ploidy based on the multiplicity of chromosomes within the genus/family, or based on SI/SC classification. For example, *Solanum betaceum* did not have information about ploidy level, but it has $2n = 24$ chromosomes, and $x = 12$ is the base chromosome number of the genus *Solanum* (Olmstead & Bohs, 2007), so we assigned *S. betaceum* as diploid. Additionally, because of the absence of SI polyploids (explained above and below), species known to be SI could be scored as diploid. Species with more than one ploidy level were assigned the most frequent ploidy level recorded or the smallest ploidy in case of frequency ties.

Breeding system states were scored as self-incompatible, *I* or self-compatible, *C*, based on results curated from the literature (as compiled in Igić *et al.*, 2006; Goldberg *et al.*, 2010; Robertson *et al.*, 2011; Goldberg & Igić, 2012) and original experimental crosses (BI and EEG unpublished). Most species could unambiguously be coded as either *I* or *C* (Raduski *et al.*, 2012). Following previous work, we coded any species with a functional SI system as *I*, even if SC or dioecy was also reported. Dioecious species without functional SI were coded as *C*.

Resolution of taxonomic synonymy followed Solanaceae Source (PBI *Solanum* Project, 2012). Hybrids and cultivars were excluded because ploidy and breeding system can be affected by artificial selection during domestication. Following the reasoning outlined in Robertson *et al.* (2011), we closely examined the few species for which the merged ploidy and breeding system data indicated the presence of self-incompatible polyploids. Although SI populations frequently contain some SC individuals, and diploid populations frequently contain some polyploid individuals, in no case did we find convincing data for a naturally occurring SI polyploid population (discussed in Robertson *et al.*, 2011). Because of the resulting absence of polyploid SI populations, as well as the functional explanation for polyploidy disabling gametophytic SI systems with nonself recognition (see the Introduction), we consider only three observed character states: self-incompatible diploids (*ID*), self-compatible diploids (*CD*) and self-compatible polyploids (*CP*).

Matching our character state data to the largest time-calibrated phylogeny of Solanaceae (Särkinen *et al.*, 2013) yielded 651 species with ploidy and/or breeding system information on the tree. Of these, 368 had information for both states. The number of species in each combination of states is summarised in Fig. 1(a) and in Fig. 1(a) and Fig. S1. We retained all 651 species in

each of the analyses below because pruning away tips lacking breeding system in the ploidy-only analyses (and *vice versa*) would discard data that could inform the diversification models. In total, 372 taxa without any information about breeding system or ploidy were excluded.

The Supporting Information contains citations for the numerous original data sources. The Dryad archive contains the data and tree files used for analyses (doi: 10.5061/dryad.tb7055f).

Models

In order to test our hypotheses about lineage diversification and trait macroevolution, we fit 29 state-dependent speciation and extinction models (BiSSE, MuSSE, HiSSE; Maddison *et al.*, 2007; FitzJohn, 2012; Beaulieu & O'Meara, 2016). SSE models contain parameters that describe per-lineage rates of speciation and extinction, specific to each character state (denoted λ and μ, respectively, with subscripts to indicate the state), along with rates of transitions between states (denoted ρ for polyploidization, δ for diploidization, and q_{IC} for loss of self-incompatibility). The full set of models and all their rate parameters are detailed in Fig. S2. Here, we summarize how each model allows us to assess whether diversification is best explained by variation in ploidy, breeding system, their combination or some unknown factor.
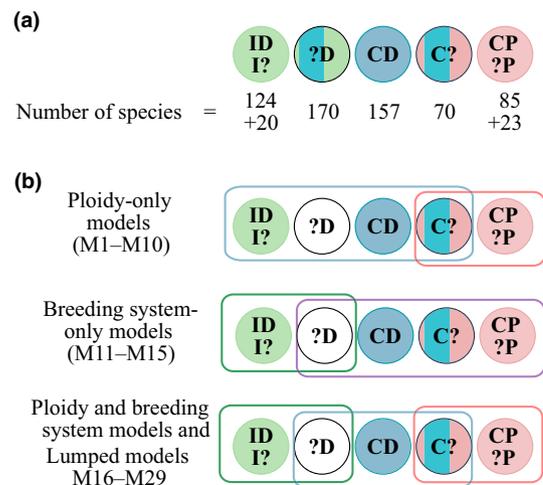

Fig. 1 Character states used in the models. (a) Each species retained on the phylogeny belonged to one of five possible categories, depending on whether ploidy and/or breeding system were known. Number of species in each category is indicated; for example, 70 species are selfcompatible with unknown ploidy. Character state abbreviations are: *I* for selfincompatible, *C* for selfcompatible, *D* for diploid, *P* for polyploid, ? for unknown. Because polyploidization breaks this form of self-incompatibility, self-incompatible species with unobserved ploidy (*I* ?) are taken to be diploid (*ID*), and polyploid species with unobserved breeding system (?*P*) are taken to be SC (*CP*). (b) Category groupings into states for each model class. In the ploidy-only models (M1–M10), states are coded as *D* & *P* when uncertain/consistent with either state; in the breeding system-only models (M11–M15) such states are coded as *C*; in the ploidy and breeding system models (M16–M29), they are coded as *CD* & *CP*. In the hidden-trait models, all species could take on either of two 'hidden' character states. Two species, *Lycium californicum* and *Solanum bulbocastanum*, are simultaneously *ID* and *CP*, and by adding them the sample adds to the total of 651 taxa used for analyses.

Ploidy and diversification We first employed a model (labeled M1), previously used by Mayrose *et al.* (2011), with each species classified as diploid (*D*) or polyploid (*P*). Although this model can be powerful in studies of trait evolution, it is prone to incorrectly reporting that a trait is associated with diversification differences (Maddison & FitzJohn, 2015; Rabosky & Goldberg, 2015). We therefore define several models that incorporate additional forms of diversification rate heterogeneity. The second ploidy model (M2) includes a binary hidden trait that subdivides each observed state. In this trait-independent model known as CID (Beaulieu & O'Meara, 2016), hidden traits can affect diversification but the observed traits do not. Comparing M1 and M2 allows us to test whether diversification rate heterogeneity is better explained by ploidy or by some unknown factor.

We fit three models in which both ploidy and a hidden trait could influence diversification (M3–M5). These models differ in whether transitions between the hidden states are symmetric (M3) or asymmetric (M4), and whether the polyploidization rate depends on the hidden state (M5). Comparing M1–M5 (as shown in Table 1) allows us to test whether ploidy is associated with diversification differences on top of the differences potentially explained by an unknown factor. We further fit the analogues of these five models but including a rate parameter δ for transitions from polyploid to diploid (M6–M10). These comparisons allow us to assess whether our conclusions about ploidy and diversification are robust to the possibility of diploidization.

Breeding system and diversification We propose five breeding system models following the same logic as the ploidy models above. Under the simplest breeding system and diversification model (M11), species are classified as self-incompatible (*I*) or self-compatible (*C*). This is the same model as in the analysis presented in Goldberg *et al.* (2010) but with an updated phylogeny (Särkinen *et al.*, 2013) and a larger aggregated dataset. We then add models to allow diversification to be influenced by only a hidden trait (M12), or by both breeding system and a hidden trait (M13–M15, with varying degrees of complexity in the hidden trait transitions Table 2). Similar models were used by Freyman & Höhna (2019) to study diversification in Onagraceae.

Self-incompatibility is homologous in all Solanaceae species in which S-alleles have been cloned and controlled crosses performed. All species sampled to date possess a nonself recognition,

RNase-based gametophytic self-incompatibility (shared even with other euasterid families; Ramanauskas & Igić, 2017). Furthermore, species that are distantly related within this family carry closely-related alleles, with deep trans-specific polymorphism at the locus that controls the SI response (Ioerger *et al.*, 1990; Igić *et al.*, 2006). Thus, there is strong evidence in Solanaceae that the *I* state is ancestral in the family, and that the SI mechanism was not regained. For all breeding system models, we estimated a transition rate from *I* to *C* but not the reverse ($q_{CI} = 0$).

Ploidy, breeding system and diversification Ploidy and breeding system might influence lineage diversification individually, but these two traits also have an intricate association (discussed in the Introduction). Therefore, we considered several multistate models that investigate the contribution of both traits and the allowable transitions between them.

The simplest model (M16) classifies each species as either SI diploid (*ID*), SC diploid (*CD*) or SC polyploid (*CP*); recall that SI polyploids do not occur. Each of these states may again be associated with different rates of speciation and extinction, and the allowable transitions are loss of SI within the diploid state (from *ID* to *CD*), loss of SI via polyploidization (from *ID* to *CP*), and polyploidization while SC (from *CD* to *CP*). As for the previous models of only one trait, we then allow diversification to be influenced by only a hidden trait (M17), or by ploidy, breeding system, and a hidden trait (M18–M20) with varying degrees of complexity in the hidden-trait transitions (similar to Caetano *et al.*, 2018; Huang *et al.*, 2018). We also fit the analogous models but allowing for diploidization (M21–M25).

Lumped models The models described so far allow us to assess the contributions of our two focal characters – ploidy and breeding system – to lineage diversification, but they do not reveal whether it is valuable to include both characters in the analysis. To answer this question with statistical model comparisons requires comparing the likelihood of the data given each model. This is impossible for the ploidy and breeding system models presented so far, however, because the data are different for the different models: they use either the *D/P* or the *I/C* or the *ID/CD/CP* state spaces (see Fig. 1 for state classifications). Therefore, the use of different data results in incomparable models.

In order to compare fits of ploidy-only vs breeding system-only vs combined trait models, we use the technique of 'lumping'

Table 1 Bayes factors for ploidy-only models without diplodisation in log-scale.

Model	Marginal log-likelihood	M2	M3	M4	M5	Evidence
M1. D/P	−1283.76	59.90	49.23	60.48	58.82	Every model strongly preferred over M1
M2. CID D/P	**−1223.86**		−10.66	0.579	−1.079	Strong preference for M2 over M1 and M3
M3. D/P + A/B	−1234.52			11.24	9.58	Asymmetric rates strongly preferred over symmetric
M4. D/P + A/B asym	**−1223.28**				−1.658	Moderate evidence for only asymmetric hidden rates
M5. D/P + A/B all asym	−1224.93					

Numbers smaller than −1 indicate moderate preference for the model listed in the row, and values larger than 1 indicate moderate preference for the model listed in the respective column. Conventional threshold for 'strong' preference is an absolute value larger than 10. The character independent model (M2, bold) is strongly preferred over the ploidy-only model (M1) or a hidden state model with symmetric rates (M3). Moderate support exists for the model with ploidy and asymmetric hidden states (M4, bold) over the rest of the ploidy-only models.

Table 2 Bayes factors for breeding system-only models in log-scale.

Model	Marginal log-likelihood	M12	M13	M14	M15	Evidence
M11. I/C	−1309.07	41.13	38.59	61.34	61.40	Every model strongly preferred over M11
M12. CID I/C	−1267.93		−2.53	20.21	20.37	Models with asymmetric rates are preferred over M12
M13. I/C + A/B	−1270.47			22.75	22.80	Asymmetric rates strongly preferred over symmetric
M14. I/C + A/B asym	**−1247.72**				0.05	No evidence
M15. I/C + A/B all asym	**−1247.66**					

Numbers smaller than −1 indicate moderate preference for the model listed on the row, and values larger than 1 indicate moderate preference for the model in the respective column. Conventional threshold for 'strong' preference is an absolute value larger than 10. Moderate to strong preference exist for models M14 and M15 (bold) which include a hidden states with asymmetrical rates.

states together (Tarasov, 2019). We use the state space of the *ID/CD/CP* model but constrain the rate parameters to mimic the behavior of the single-trait models. Lumping states requires that the transition rates from the lumped state to the singular state be equal (Tarasov, 2019). First, we lump together *ID* and *CD* to form the diploid state, mimicking the *D/P* model (M26). Proposing a lumped ploidy model by aggregating *ID* and *CD* requires forcing the rate of polyploidization from *ID* and *CD* to *CP* to be equal (i.e. $\rho_0 = \rho_I = \rho_c$), but also requires assuming that the rates of speciation and extinction for *ID* and *CD* to be equal. Therefore, we define the new parameters λ_D and μ_D that are the same for each of the two diploid states *ID* and *CD*. We used the same procedure to lump together *CD* and *CP* to form the self-compatible state, mimicking the *I/C* model (M28). In this particular case, the rate from *CD* to *CP* back to *ID* is zero and equal for both, so the model is lumpable. However, to fully mimic the breeding system model, we assume that the rate of selfing is equal (i.e. $q_0 = q_{IC} = \rho_I$) and the rates of speciation and extinction for both *CD* and *CP* are the same (new parameters λ_C and μ_C). We further add a hidden character to each of these models (M27 and M29), and then compare this group of models (Table 3).

We do not include additional models with diploidization because this reverse ploidy transition renders the models non-lumpable. When including diploidization, transitions from *CP* to *CD* are at rate δ but transitions from *CP* to *ID* do not occur. Because *ID* and *CD* would be lumped to mimic the *D/P* model, this model is nonlumpable when $\delta \neq 0$. Thus, we can compare models to test whether it is advantageous to include both traits, but only when ignoring diploidization.

Pathways to polyploidy Considering ploidy and breeding system together, there are two evolutionary pathways from SI diploid to SC polyploid (Brunet & Liston, 2001; Robertson *et al.*, 2011). In the one-step pathway, the *CP* state is produced directly from the *ID* state when whole genome duplication disables SI. In the two-step pathway, the *CD* state is an intermediate: SI is first lost, and later the SC diploid undergoes polyploidization. We quantify the relative contribution of these pathways to polyploidy in two ways, each using the median estimates of rates from the simplest model that includes both traits (M16). Our results differ from those of Robertson *et al.* (2011) in part because our approach relies on a dated phylogeny and models that allow for state-dependent diversification.

Both of our methods are based on a propagation matrix that describes flow from *ID* to *CP*, as in Robertson *et al.* (2011). We insert an artificial division in the *CP* state, so that one substate contains the *CP* species that arrived via the one-step pathway and the other substate contains the *CP* species that arrived via the two-step pathway. We consider unidirectional change along each step of the pathway in order to separate them into clear alternatives, and because in this family there is no support for regain of SI, and no strong support for diploidization (as discussed in key question (5) in the Results section). First, we consider only the rates of transitions between these states, placing them in the propagation matrix **Q**. The matrix $\mathbf{P} = (\mathbf{Q}t)$ then provides the

Table 3 Testing the addition of a focal trait to a binary state model via Bayes factors (log-scale).

Model	Marginal log-likelihood	Comparison	$K = \log_e(BF(M_i, M_{ii}))$	Preferred model (Evidence)
M26. Lumped D/P	−1463.22	M26 vs M16	4.10	M16 (Moderate)
M16. ID/CD/CP	**−1459.11**			
M27. Lumped D/P + A/B	−1417.67	M27 vs M23	3.678	M23 (Moderate)
M23. ID/CD/CP + A/B	**−1414.00**			
M28. Lumped I/C	−1458.41	M28 vs M16	−0.69	No evidence
M16. ID/CD/CP−	**1459.11**			
M29. Lumped I/C + A/B	−1416.60	M29 vs M23	2.60	M23 (Moderate)
M23. IC/CD/CP + A/B	**−1414.00**			

Numbers smaller than −1 indicate moderate preference for the model in the row, and values larger than 1 indicate moderate preference for the model in the respective column. Conventional threshold for 'strong' preference is an absolute value larger than 10. Three-state models (M16, M23, bold) are moderately preferred over two-state models. Moderate evidence exists towards the inclusion of breeding systems in ploidy models (M16, M23, bold). Moderate to no evidence towards the inclusion of ploidy in breeding system models is indicated by the comparisons of models M28 vs M16 and M29 vs M23 where the difference is the inclusion of ploidy information.

probabilities of changing from one state to any other state after time t. Closed-form solutions for the two pathway probabilities are provided in Robertson *et al.* (2011). Second, we consider not only transitions between states but also diversification within each state. State-dependent diversification can change the relative contributions of the two pathways. For example, if the net diversification rate is small for *CD*, the two-step pathway will contribute relatively less. We therefore include the difference between speciation and extinction along the diagonal elements of the propagation matrix. As before, matrix exponentiation provides the relative chance of changing from one state to any other state after time t. The calculations of the propagation matrix are not probabilities because diversification changes the number of lineages as time passes. We can still use ratios, however, to consider the relative contribution of each pathway, analogous to the normalised age structure in a growing population (Leslie, 1945).

Statistical inference

Model fitting Parameters for each of the 29 models were coded as graphical models and Bayesian statistical inference was performed with RevBayes (Höhna *et al.*, 2016). Scripts for analyses and key results are available in Dryad (doi: 10.5061/dryad.tb7055f). We accounted for incomplete sampling in all analyses by setting the probability of sampling a species at the present to 651/3000 (using the method of FitzJohn *et al.*, 2009) since the Solanaceae family has *c.* 3000 species (PBI *Solanum* Project, 2012). For all models, we assumed that speciation and extinction parameters had log-normal prior distributions with means equal to the expected net diversification rate (number of taxa/(2 × root age)) and standard deviation 0.5. Priors for parameters defining trait changes were assumed to be gamma distributed with parameters $k = 0.5$ and $\theta = 1$. For each model, a Markov chain Monte Carlo (MCMC; Metropolis *et al.*, 1953; Hastings, 1970) was run in the high-performance computational cluster at the Minnesota Supercomputing Institute, which allowed for 5000 generations of burn-in and a minimum of 200 000 generations of MCMC for each of the models. Convergence and mixing of each MCMC chain was determined by ensuring the effective sample size of each parameter was over 200.

We report posterior distributions for the net diversification parameters ($\lambda - \mu$) in Fig. 2 and full results of all parameters of the diversification models in Figs S3–S8. Additionally, ancestral states at each node in the phylogeny were sampled jointly during the MCMC analyses every 100 generations. Ancestral state estimations for all models show the maximum *a posteriori* estimates of the marginal probability distributions for each of the 650 internal nodes for each of the models in Fig. 2 (Figs S9–S14).

Model selection key questions We calculated the marginal log-likelihood for each of the models using 50 stepping stone steps under the methodology of Xie *et al.* (2010), implemented in RevBayes (Höhna *et al.*, 2016). Each stepping stone step was found by calculating at least 500 generations of burn-in followed by a total of 1000 MCMC steps (Table 1).

Using the marginal likelihood values, we calculated Bayes factors to answer five key biological and methodological questions:

(**1**) Are diversification patterns only determined by hidden states and not the traits of interest? – Comparison of character-independent models against hidden state (Tables 1, 2, S1–S3).
(**2**) Are hidden states necessary to explain diversification rate heterogeneity? – Comparison of simple models against hidden state models (Table S4).
(**3**) Does a second focal trait add information about the diversification process? – Comparison of lumped models against *IC*/*CD*/*CP* models (Figs 3, S15, S16; Table 3).
(**4**) Are conclusions robust to assumptions about hidden state transitions? – Comparison amongst hidden states models with equal hidden rates and asymmetrical rates (Table S5; Fig. S17).
(**5**) Is there evidence for diploidization? – Comparison amongst log-scale models with and without diploidization (Table S6; Fig. S18).

Each model comparison is reported with a Bayes factor on the natural log-scale: the comparison between models M_0 and M_1 is $K = \log_e(\mathrm{BF}(M_1, \quad M_0)) = \log_e[P(X|M_1) - P(X|M_0)]$. There is 'strong evidence' for M_1 when this value is > 10, moderate support if the value is > 1, and no evidence if the value is between -1 and 1. If the value of K is negative the evidence goes towards M_0 (Kass & Raftery, 1995).

Results

Trait-dependent diversification

Ploidy only When considering ploidy alone, we found a larger net diversification rate for diploids than for polyploids, in agreement with Mayrose *et al.* (2011, 2015). This result holds with (model M1; Fig. 2a) or without the diploidization parameter (M6; Fig. S18a). Incorporating a hidden state in this model, however, reduces the clear separation in diversification rate estimates between diploids and polyploids (M4, Fig. 2b; M9, Fig. S18b). Statistical model comparisons show a clear preference for models in which only a hidden state affects diversification or a hidden state as well as ploidy (M2 and M4; Table 1). Results are similar when diploidization is included (Table S1). Thus, when other traits are ignored, the role of ploidy in net diversification is unclear, with marginal support for diploids having higher diversification rates, but rate differences perhaps better explained by another factor.

Breeding system only When considering breeding system alone (M11; Fig. 2c), we found a larger net diversification rate for SI than for SC species, in agreement with Goldberg *et al.* (2010). When a hidden state is included, the large net diversification rate difference persists for one hidden state but is diminished for the other (M14; Fig. 2d). In the statistical model comparisons, the best two supported models include diversification differences due to both breeding system and to a hidden trait (M14 and M15; Table 2). Breeding system seems to play a role in diversification differences, though a hidden factor does as well.

Ploidy and breeding system together When considering ploidy and breeding system together, the net diversification rate for SI
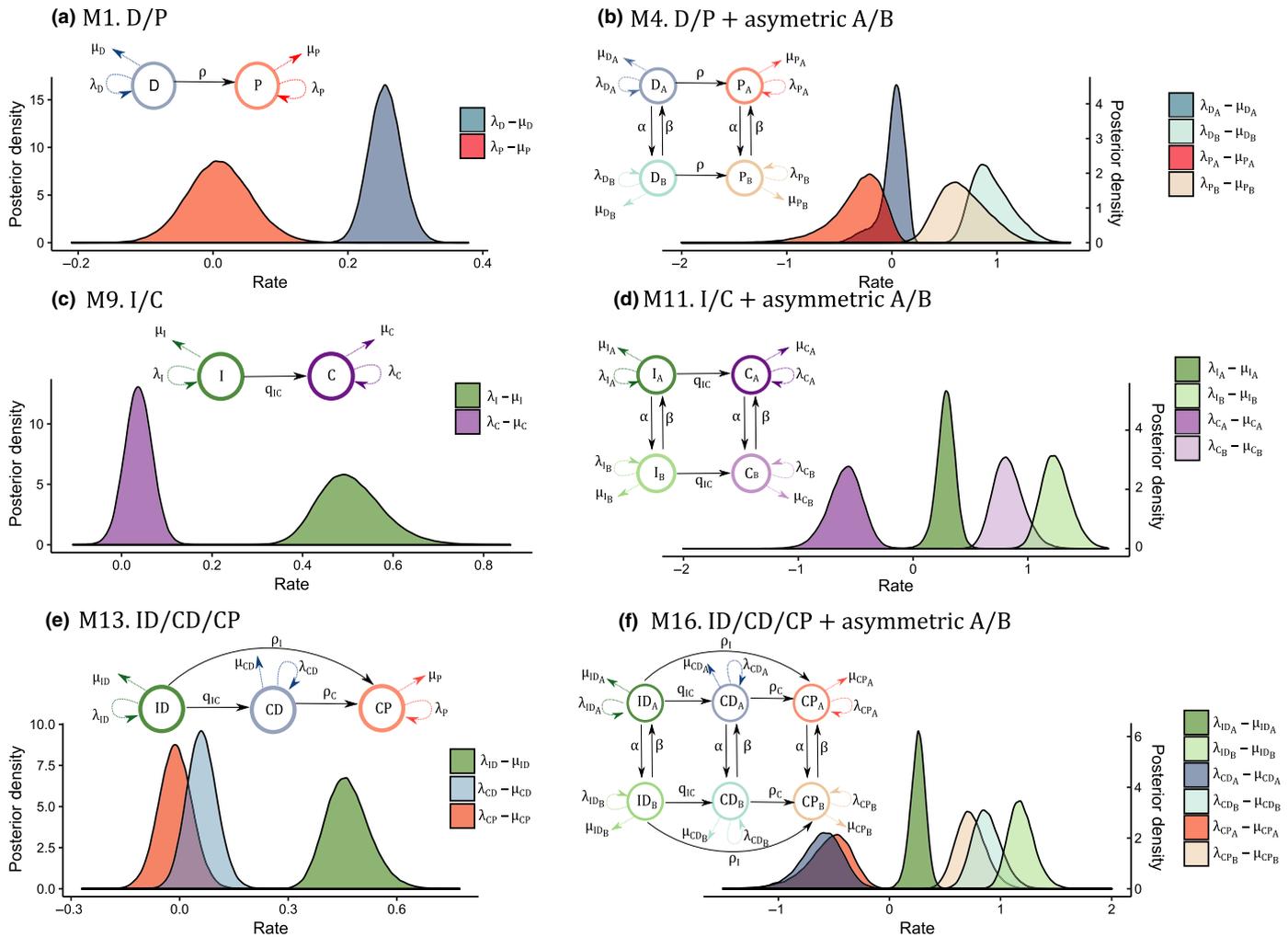
Fig. 2 Net diversification rates for SSE models of focal traits with or without hidden state. (a) Ploidy-only model (M1) showing higher net diversification linked to diploid state D compared with polyploid state P. (b) Ploidy with hidden states model (M4) showing that the net diversification is higher for hidden state B (lighter colors) compared to hidden state A (darker colors) and both diploid and polyploid states within each hidden traits have overlapping net diversification rates. (c) Breeding system-only model (M9) showing higher net diversification linked to self-incompatible I state compared with self-compatible state C. (d) Breeding system with hidden states model (M11) showing diversification differences in both hidden states (light vs dark colors) and little to no overlapping in between self-compatible vs self-incompatible states. (e) Ploidy and breeding system model (M13) showing higher net diversification linked to self-incompatible diploid state ID compared to both self-compatible states despite ploidy level (CD, CP). (f) Ploidy, breeding system and hidden states model (M16) showing a similar pattern that panel (e) within each hidden state A and B. For hidden state B there is a larger overlap of net diversification between states CD_B and ID_B.

diploids was greater than for either SC diploids or SC polyploids, with or without diploidization (M16, Fig. 2e; M21, Fig. S18e). Thus, the difference in net diversification associated with breeding system persists even when ploidy is included in the model. The reverse is not true: the association of ploidy with net diversification in the simplest ploidy-only model (M1; Figs 2a, S18a) appears to be driven by the subset of diploids that are SI, while among SC species, net diversification rates for diploids and polyploids are similar.

When a hidden state is included, the separation in net diversification rate of ID vs CD and CP persists within one hidden state but is reduced in the other (M19, Fig. 2f). The same general pattern remains when diploidization is included (Fig. S18f). Model comparisons clearly favour models that include ploidy, breeding system and the hidden trait, against the character-independent

model in which the focal traits do not influence diversification (Tables S2, S3 with diploidization).

Using the lumped models, we find moderate support for obtaining a significantly better fit by adding breeding system to the ploidy-only model (M26 vs M16, Table 3; Fig. 3b,c). This is also true when a hidden trait is included (M27 vs M23, Table 3; Fig. S15e,f). A similar comparison in which ploidy is added to the breeding system-only model shows no preference for the model that also includes ploidy (M28 vs M16, Table 3; Fig. 3c, e). When including a hidden state, however, the model with both focal traits is moderately preferred over the model with only breeding system (M29 vs M18, Table 3; Fig. S16e,f).

From all of these types of statistical evidence, we conclude that breeding system (and a hidden factor) are strongly associated with diversification differences, and that ploidy plays a smaller role.
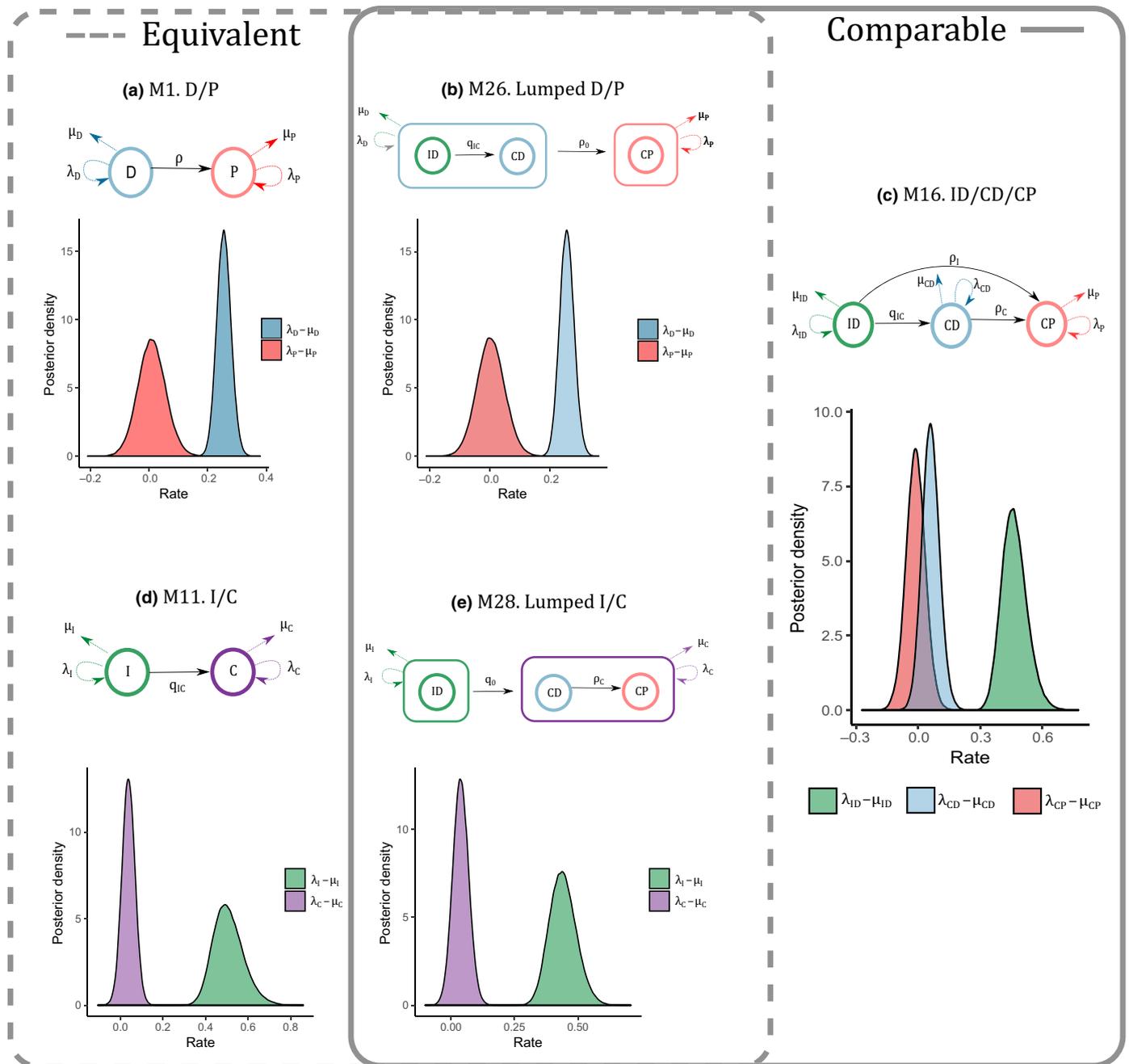
Fig. 3 Comparing two and three-state models using lumped models. (a) Ploidy model only (M1) where data enter as binary *D* and *P*. (b) Lumped model for ploidy (M26) where data are the three-state values (*ID*,*CP*,*CD*) but results are equivalent to model M1. (c) Ploidy and breeding system model (M16) where data enter as the three-state values. Models M26 and M16 are comparable whereas M1 and M16 are not. (d) Breeding system-only model (M11) where data are entered as binary *I* and *C*. (e) Lumped model for breeding system (M28) where data are the three-state values (*ID*,*CP*,*CD*) but results are equivalent to model M11. Model M26 can be compared with model M16 from panel (c). Model comparisons are done via Bayes factors and results are shown in Table 3.

Key questions about diversification and transitions

The above results include several of our statistical model comparison findings. Here we return to the five specific questions we targeted with our model comparisons.

(**1**) Are diversification patterns only determined by hidden states and not the traits of interest? No, our focal traits are supported as having associations with diversification differences. In most cases,

we find moderate to strong preference for models with the focal traits as well as hidden states, over models with only hidden states (Tables 1, 2, S1–S3).

(**2**) Are hidden states necessary to explain diversification rate heterogeneity? Yes, models with hidden states that influence diversification are strongly preferred over models containing only the focal traits (Table S4). This means that there are potentially many factors underlying diversification shifts within the family.

(3) Does a second focal trait add information about the diversification process? Yes, in most cases models with both ploidy and breeding system are preferred over models with only one of the focal traits (Table 3; Figs S15, S16).

(4) Are conclusions robust to assumptions about hidden state transitions? Yes, we found that allowing different types of asymmetry in transitions within and between hidden states did not change our conclusions about net diversification differences. Hidden state models with asymmetric rates are, however, strongly preferred over models with equal rates between hidden states (Table S5), and they show stronger differences between some net diversification rates (Fig. S17). The effect of the asymmetry of the hidden state transitions is better observed in the ancestral state estimations (Figs S10, S12, S14), which show that moving out of state A (dark colors) happens quickly, with rate α, whereas evolving out of hidden state B (light colors) is slow, with rate β.

(5) Is there evidence for diploidization? Perhaps: when comparing models with diploidization against models without it, we found moderate evidence that models containing diploidization are preferred (Table S6). We discuss later some further challenges in identifying diploidization. We found that our main conclusions about net diversification differences are not dependent on whether diploidization is included (Fig. S18).

Pathways to polyploidy

There are two pathways by which SI diploid lineages eventually – given enough time – become SC polyploids. In the one-step pathway, polyploidization directly disables SI. In the two-step pathway, SI is first lost within the diploid state, followed by polyploidy. Determining the relative contribution of these pathways based on our transition rate estimates (median transition rate values from M16), we find that the one-step pathway is more likely on short timescales and the two-step pathway is more likely on long timescales (Fig. 4, left panels). Beginning with a single SI diploid lineage, when not much time has elapsed, the one-step pathway is more likely because it only necessitates a single event to reach the SC polyploid state. When more time has elapsed, the two-step pathway is more likely because the rate of loss of SI within diploids, q_{IC}, is greater than the rate of polyploidization for SI species, ρ_I (Fig. S7). That is, an *ID* lineage is more likely to begin its path to polyploidy with a transition to *CD*, but completing this path to *CP* takes longer. Robertson *et al.* (2011)
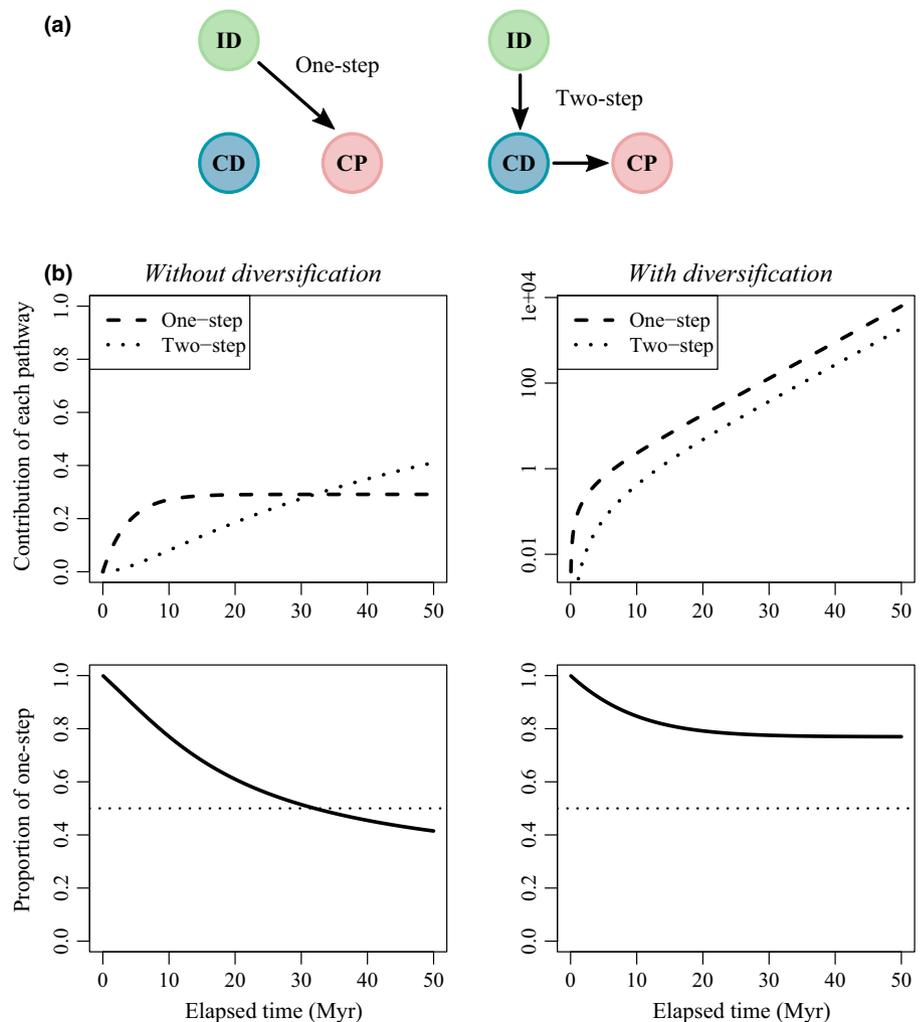

Fig. 4 Contributions of the two pathways to polyploidy. The one-step pathway is direct *ID→CP* transitions. The two-step pathway consists of *ID→CD→CP* transitions. When considering only rates of transitions among the states (ignoring the diversification rate parameters), the one-step pathway dominates on short timescales and the two-step on long timescales (left panels). When also considering diversification within each state, the one-step pathway, in which polyploidy breaks down selfincompatibility (SI), dominates over any timescale (right panels). The top panels show the separate contributions of each pathway. The lower panels show the proportional contribution of the one-step pathway (i.e. one-step/(one-step + two-step)).

reached the same conclusion. Our result is qualitatively unchanged when using transition rate estimates from the model that does not allow diversification differences related to the observed states (M17).

The preceding conclusions, however, ignore the changes in numbers of lineages in each state due to speciation and extinction. By analogy, envisioning the states as stepping stones, the extent to which each stone grows or shrinks over time affects the utility of each possible path. Allowing different net diversification rates for each state (again using median rate estimates from M16), we find a qualitative difference in the relative pathway contributions. The lower rate of net diversification in the *CD* state, relative to *ID*, means that relatively fewer lineages are available to complete the second step of the two-step pathway, ending in *CP*. Consequently, even over long timescales, we find that the two-step pathway contributes less to the formation of polyploids (Fig. 4, right panels) when considering diversification as well as transitions.

Discussion

Species are composed of vast assemblages of variable traits. Many traits are both heritable and possibly affect the propensity of species to perish or multiply (Lewontin, 1970). Examining the effects of complex trait combinations on lineage diversification, however, remains challenging. Focusing first on ploidy and then on breeding system, we found that considering each trait in isolation provides an incomplete story. Considering them together, and in conjunction with another hidden factor, provides a more complete picture of macroevolutionary dynamics within Solanaceae. We hope our work serves as an example of how phylogenetic comparative methods can be used to disentangle the contributions of interacting traits to heterogeneous lineage diversification, and how to statistically argue for increasing complexity in diversification modelling.

Interacting traits and lineage diversification

Previous analyses of the effects of ploidy on diversification found that diploids are associated with greater net diversification rates than polyploids across many angiosperm clades (Mayrose *et al.*, 2011, 2015). We obtain a similar outcome when examining ploidy alone in Solanaceae (Fig. 2a), but a consistent effect of ploidy on diversification is not clear when we incorporate a hidden factor linked to diversification (Fig. 2b). Previous analyses of breeding system in this family found that SI may cause higher diversification rates, compared with SC (Goldberg *et al.*, 2010; Fig. 2c). Our analyses that include a hidden trait recover the same pattern, with one important difference. On the background of one hidden state, we recover a net diversification rate for SC species that is positive and greater than the diversification rate of SI species on the background of the other hidden state (Fig. 2d). Therefore, SC may not be a 'dead end' when coupled with some unknown trait combinations or processes not modelled by these two traits. Our analyses also reveal that models of joint evolution of ploidy and breeding system are statistically preferred, and hint

at how various trait combinations may be linked with diversification. We find that the highest net diversification rate is associated with SI diploids, while SC diploids have a lower diversification rate that overlaps with the net diversification of SC polyploids (Fig. 2e). Thus, breeding system appears to provide a relatively better explanation of diversification rate differences, with ploidy providing a secondary effect within SC species.

Throughout our numerous model comparisons, we find that inclusion of hidden states provides a considerably better fit (Table S4). This is consistent with the expectation that many processes, beyond those associated with the focal traits, can affect inference of speciation and extinction. It is, however, unclear exactly which processes are captured by the hidden states. For example, our results show that, to a varying extent, breeding system functions as a hidden state in the ploidy-centred analysis, and *vice versa* (Figs S15, S16). But the strong statistical support for processes not well explained by ploidy and breeding system (Table S4) tempts one to interpret the remaining variation as the effect of other measurable traits. For example, our data appear to show a rapidly diversifying Australasian clade of mostly SC species within *Solanum*, which suggests that geography may play a role. Nevertheless, it is also possible that the addition of hidden states instead explains variation stemming from any of a number of unrelated processes or methodological artifacts, as previously discussed by Beaulieu & O'Meara (2016). In the absence of additional information, the hidden states can be viewed as a statistical trick, providing an easy way to model extra heterogeneity without directly representing a specific trait.

As more trait information becomes available for macroevolutionary studies, it is not only important to recognize the role of hidden states as a part of a general modelling approach, but also to question whether adding more traits to diversification studies is justified. More generally, we find that lumped models are useful for assessing the value of adding additional traits to already complex diversification models (Fig. 3; Table 3).

Although we fit an extensive set of models in order to relax a variety of assumptions, we did not explore the process of trait change in conjunction with speciation. That is, our models all assume anagenetic trait evolution and ignore cladogenetic shifts. Anagenetic and cladogenetic changes can be separated with phylogenetic models (Mayrose *et al.*, 2011; Goldberg & Igić, 2012; Magnuson-Ford & Otto, 2012). These have been applied to estimate the relative contribution of anagenetic and cladogenetic shifts in breeding system (Goldberg & Igić, 2012) and polyploidy (Zhan *et al.*, 2016; Freyman & Höhna, 2019). We did not explore cladogenetic trait change because of the requisite increase in parameter space and model number, outside the scope of our already complex and extensive modelling framework. Although Goldberg & Igić (2012) found that allowing cladogenetic changes did not substantially affect inference of net diversification rates associated with breeding system, future work could test whether this process affects diversification rate estimates with the more complex state and parameter spaces of our other models.

Pathways to polyploidy

With evolution predominantly in the direction from diploid to polyploid, and from SI to SC, surviving lineages will tend to become SC polyploids. We find that in Solanaceae, the pathway to this state is more likely to consist of a single step ($ID \rightarrow CP$) than two steps ($ID \rightarrow CD \rightarrow CP$; Fig. 4). Although this question focuses on the process of state transitions, we also show that its answer is affected by the process of lineage diversification. We used a simple mathematical approach to investigate the contributions of the two pathways, but future work could instead rely on stochastic character mapping to estimate the numbers of each type of transition more directly.

Macroevolutionary transition rates represent a combination of time spent waiting for individuals with a new character state, and for that new state to become widespread within the species. For our traits, this consists of mutations that break SI or generate polyploid individuals, and selective pressures that cause fixation (or loss) of these mutants. Estimates of mutation rates are highly uncertain, but the chance of breakdown of SI within diploids is perhaps 10^{-5} per pollen grain; this includes breakdown by autopolyploidy (Lewis, 1979), which is by itself estimated to occur approximately within the same order of magnitude (Ramsey & Schemske, 1998). By contrast, we infer a macroevolutionary transition rate from ID to CP that is 2.5-times greater than the rate from ID to CD, indicating that selection restricts the fixation of new polyploids more than of new SC mutants (Fig. 4; Robertson *et al.*, 2011).

Our findings prompt several further questions about the macroevolutionary pathways of ploidy and breeding system. First, our support for the direct pathway is consistent with the idea that breakdown of SI by whole genome duplication – via diploid 'heteroallelic' pollen – may trigger the evolution of gender dimorphism as a different mechanism of inbreeding avoidance (Miller & Venable, 2000). A further test of this hypothesis would additionally examine the propensity of polyploids generated through either pathway to become dioecious (Robertson *et al.*, 2011). Second, we might wonder whether the propensity for a polyploid species to diversify depends on whether it arose via the one-step or two-step pathway. This could be tested with a different form of a hidden state model, in which the polyploid state is subdivided into parts, reflecting the arriving path taken to that state. Such an estimate of path dependence could also include the possibility of different diversification rates for the subdivided CP substates. Third, the generality of our findings in other families remains to be assessed. An identical procedure could be used in other families with gametophytic SI. In clades with sporophytic SI systems, however, SI is not disabled by whole genome duplication, so there is no one-step pathway (Miller & Venable, 2000; Mable, 2004). The correlation between breeding system and ploidy may therefore be different in sporophytic systems, and it is unclear whether one of the two-step pathways might predominate.

Diploidization

Polyploidy is known to be common in plants, but the pace and relative frequency of the reverse process – diploidization – remains unclear, and it is under active investigation (Dodsworth *et al.*, 2015; Mandakova & Lysak, 2018). Ignoring diploidization, if it is common, could cause underestimation and increase of uncertainty in the polyploidy rate. Therefore, we included a slate of models with a diploidization parameter, and show that our main conclusions are robust to this process (Fig. S18). These models also suggest modest statistical support for diploidization (Table S6), although our estimates of its rate were highly uncertain. Furthermore, additional lines of evidence for classifying species as diploid or polyploid (beyond the genus-level chromosome multiplicity that we primarily relied on) are needed for more reliable conclusions.

Other lines of evidence about the prevalence of diploidization within Solanaceae or its ancestors are mixed or even conflicting. On the one hand, polyploidy may have occurred before the origin of Solanaceae, rendering all extant 'diploids' secondarily derived. Ku *et al.* (2000) and Blanc & Wolfe (2004) posited that the lineage leading to cultivated tomato, *Solanum lycopersicum*, may have experienced one or more whole genome duplications. A subsequent analysis of synteny between grape and *Solanum* genomes, as well as genetic distances (Ks) between inferred paralogs within genomes of *Solanum* (tomato and potato), each suggested that this lineage experienced a likely round of ancient genome duplication or triplication (71 ± 19 Myr; Tomato Genome Consortium, 2012) likely pre-dating the origin of the family (49 ± 3 Myr; Särkinen *et al.*, 2013).

On the other hand, there is little evidence for the occurrence of diploidization after the origin of Solanaceae itself. Studies comparing genetic map-based genome synteny within a number of species in this family find no evidence for diploidiztion (Wu & Tanksley, 2010). Instead, simple genome re-arrangements appear sufficient to explain chromosomal evolution within species in the cytogenetically conserved '$x=12$' group, which includes tomato, potato, eggplant, pepper and tobacco. Recovery of comparatively few rearrangements would require outstanding convergent loss of duplicated segments. Furthermore, whole genome duplication in a eudicot lineage ancestral to Solanaceae would clash with the evidence that the homologous mechanism of SI, which has been present continually in many families (Igić *et al.*, 2006), breaks down nearly invariably in natural and induced tetraploids (Stone, 2002; McClure, 2009). Most problematically in this context, it is unclear how to explain the maintenance of trans-generic polymorphism at the orthologous S-loci in Solanaceae and other families if SI was previously broken down by polyploidy. Regardless of whether genome polyploidization, followed by widespread diploidization, is a dominant mechanism of genome evolution in Solanaceae, it is clear that more work is needed for a complete understanding of the joint evolution of ploidy and breeding systems.

Conclusion

Heterogeneity in lineage diversification across time and clades is the rule, rather than the exception. This background heterogeneity makes it difficult to test for the association of any one, isolated trait with different rates of speciation or extinction. Our study provides

an example of how diversification linked to a particular trait can be better assessed by a suite of more inclusive models that allow for alternative explanations, whether other traits or unknown factors. Additionally, our analysis of evolutionary pathways to polyploidy shows the importance of including diversification effects even when addressing questions that focus on trait evolution. Finally, although a growing number of plant traits has been studied to date, breeding systems indeed seem to be among the most influential in governing macroevolution (Barrett *et al.*, 1996).

Acknowledgements

Author contributions

RZ-F, BI and EEG designed the research; RZ-F and EEG performed the research; RZ-F conducted data analyses; RZ-F, BI and EEG collected the data; RZ-F, JGB, WAF, BI, EEG were involved in data interpretation; RZ-F, JGB, WAF, BI, IM and EEG wrote the manuscript.

ORCID

Boris Igić (iD) https://orcid.org/0000-0002-6302-8030
Rosana Zenil-Ferguson (iD) https://orcid.org/0000-0002-9083-2972

References

Barrett SCH. 2008. Major evolutionary transitions in flowering plant reproduction: an overview. *International Journal of Plant Sciences* **169**: 1–5.

Barrett SCH. 2013. The evolution of plant reproductive systems: how often are transitions irreversible? *Proceedings of the Royal Society B: Biological Sciences* **280**: 20130913.

Barrett SCH, Harder LD, Worley AC. 1996. The comparative biology of pollination and mating in flowering plants. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **351**: 1271–1280.

Barringer BC. 2007. Polyploidy and self-fertilization in flowering plants. *American Journal of Botany* **94**: 1527–1533.

Beaulieu JM, O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* **65**: 583–601.

Bennett M, Leitch I. 2012. *Angiosperm dna c-values database* (release 8.0, Dec. 2012). [WWW document] URL https://data.kew.org/cvalues/ [accessed 23 March 2018].

Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**: 1667–1678.

Brunet J, Liston A. 2001. Technical comment on polyploidy and gender dimorphism. *Science* **291**: 1441–1442.

Caetano DS, O'Meara BC, Beaulieu JM. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* **72**: 2308–2324.

Dodsworth S, Chase MW, Leitch AR. 2015. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Botanical Journal of the Linnean Society* **180**: 1–5.

Entani T, Takayama S, Iwano M, Shiba H, Che F-S, Isogai A. 1999. Relationship between polyploidy and pollen self-incompatibility phenotype in *Petunia hybrida* Vilm. *Bioscience, Biotechnology, and Biochemistry* **63**: 1882–1888.

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.

Freyman WA, Höhna S. 2019. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. *Systematic Biology* **68**: 505–519.

Goldberg EE, Igić B. 2012. Tempo and mode in plant breeding system evolution. *Evolution* **66**: 3701–3709.

Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igić B. 2010. Species selection maintains self-incompatibility. *Science* **330**: 493–495.

Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.

Hauck NR, Yamane H, Tao R, Iezzoni AF. 2002. Self-compatibility and incompatibility in tetraploid sour cherry (*Prunus cerasus* L.). *Sexual Plant Reproduction* **15**: 39–46.

Herrera-Alsina L, van Els P, Etienne RS. 2018. Detecting the dependence of diversification on multiple traits from phylogenetic trees and trait data. *Systematic Biology* **68**: 317–328.

Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model specification language. *Systematic Biology* **65**: 726–736.

Huang D, Goldberg EE, Chou L, Roy K. 2018. The origin and evolution of coral species richness in a marine biodiversity hotspot. *Evolution* **72**: 288–302.

Husband BC, Baldwin SJ, Suda J. 2013. The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: Leitch IJ, Greilhuber J, Dolezel J, Wendel J, eds. *Plant genome diversity, vol. 2.* Vienna, Austria: Springer-Verlag, 255–276.

Husband BC, Ozimec B, Martin SL, Pollock L. 2008. Mating consequences of polyploid evolution in flowering plants: current trends and insights from synthetic polyploids. *International Journal of Plant Sciences* **169**: 195–206.

Igić B, Bohs L, Kohn JR. 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proceedings of the National Academy of Sciences, USA* **103**: 1359–1363.

Igić B, Kohn JR. 2001. Evolutionary relationships among self-incompatibility RNases. *Proceedings of the National Academy of Sciences, USA* **98**: 13167–13171.

Igić B, Lande R, Kohn JR. 2008. Loss of self-incompatibility and its evolutionary consequences. *International Journal of Plant Sciences* **169**: 93–104.

Ioerger TR, Clark AG, Kao TH. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences, USA* **87**: 9732–9735.

Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* **90**: 773–795.

Ku HM, Vision T, Liu J, Tanksley SD. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences, USA* **97**: 9121–9126.

Kubo K-i, Entani T, Takara A, Wang N, Fields AM, Hua Z, Toyoda M, Kawashima S-i, Ando T, Isogai A et al. 2010. Collaborative non-self

recognition system in S-RNase-based self-incompatibility. *Science* 330: 796–799.

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* 105: 348–363.

Leslie PH. 1945. On the use of matrices in certain population mathematics. *Biometrika* 33: 183–212.

Levin DA. 1975. Minority cytotype exclusion in local plant populations. *Taxon* 24: 35–43.

Lewis D. 1947. Competition and dominance of incompatibility alleles in diploid pollen. *Heredity* 1: 85–108.

Lewis D. 1979. *Sexual incompatibility in plants, the institute of biology's studies in biology, vol. 110.* London, UK: Edward Arnold Limited.

Lewontin RC. 1970. The units of selection. *Annual Review of Ecology and Systematics* 1: 1–18.

Mable BK. 2004. Polyploidy and self-compatibility: is there an association? *New Phytologist* 162: 803–811.

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.

Magnuson-Ford K, Otto SP. 2012. Linking the investigations of character evolution and species diversification. *The American Naturalist* 180: 225–245.

Mandakova T, Lysak MA. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* 42: 55–65.

Mayrose I, Zhan SH, Rothfels CJ, Arrigo N, Barker MS, Rieseberg LH, Otto SP. 2015. Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis *et al.* (2014). *New Phytologist* 206: 27–35.

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.

McClure B. 2009. Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for S-RNase-based SI. *Journal of Experimental Botany* 60: 1069–1081.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21: 1087–1092.

Miller JS, Venable DL. 2000. Polyploidy and the evolution of gender dimorphism in plants. *Science* 289: 2335–2338.

Nunes MD, Santos RA, Ferreira SM, Vieira J, Vieira CP. 2006. Variability patterns and positively selected sites at the gametophytic self-incompatibility pollen SFB gene in a wild self-incompatible *Prunus spinosa* (Rosaceae) population. *New Phytologist* 172: 577–587.

Olmstead RG, Bohs L. 2007. A summary of molecular systematic research in Solanaceae: 1982–2006. *Acta Horticulturae* 745: 255–268.

PBI *Solanum* Project. 2012. *Solanaceae Source: a global taxonomic resource for the nightshade family.* [WWW document] URL http://solanaceaesource.org/content/citing-us [accessed 23 March 2018].

Rabosky DL, Goldberg EE. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology* 64: 340–355.

Rabosky DL, Goldberg EE. 2017. FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution* 71: 1432–1442.

Raduski AR, Haney EB, Igić B. 2012. The expression of self-incompatibility in angiosperms is bimodal. *Evolution* 66: 1275–1283.

Ramanauskas K, Igić B. 2017. The evolutionary history of plant T2/S-type ribonucleases. *PeerJ* 5: e3790.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467–501.

Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 33: 589–639.

Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. 2015. The chromosome counts database (CCDB) – a community resource of plant chromosome numbers. *New Phytologist* 206: 19–26.

Rivero R, Sessa EB, Zenil-Ferguson R. 2019. EyeChrom and CCDB curator: visualizing chromosome count data from plants. *Applications in Plant Sciences* 7: e01207.

Robertson K, Goldberg EE, Igić B. 2011. Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. *Evolution* 65: 139–155.

Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* 13: 214.

Sessa EB. 2019. Polyploidy as a mechanism for surviving global change. *New Phytologist* 221: 5–6.

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development* 35: 119–125.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al.* (2011). *New Phytologist* 202: 1105–1117.

Stebbins GL. 1938. Cytological characteristics associated with the different growth habits in the dicotyledons. *American Journal of Botany* 25: 189–198.

Stebbins GL. 1950. *Variation and evolution in plants.* New York, NY, USA: Columbia University Press.

Stebbins GL. 1974. *Flowering plants: evolution above the species level.* Cambridge, MA, USA: Belknap Press of Harvard University Press.

Steinbachs JE, Holsinger KE. 2002. S-RNase-mediated gametophytic self-incompatibility is ancestral in eudicots. *Molecular Biology and Evolution* 19: 825–829.

Stone JL. 2002. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *The Quarterly Review of Biology* 77: 17–30.

Stout AB, Chandler C. 1942. Hereditary transmission of induced tetraploidy and compatibility in fertilization. *Science* 96: 257–258.

Tarasov S. 2019. Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits. *Systematic Biology* 68: 698–716.

Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635.

Tsukamoto T, Ando T, Watanabe H, Marchesi E, Kao T-h. 2005. Duplication of the S-locus F-box gene is associated with breakdown of pollen function in an S-haplotype identified in a natural population of self-incompatible *Petunia axillaris*. *Plant Molecular Biology* 57: 141–153.

de Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E. 2014. Heterostyly accelerates diversification via reduced extinction in primroses. *Proceedings of the Royal Society B: Biological Sciences* 281: 20140075.

Wu F, Tanksley SD. 2010. Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* 11: 182.

Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H. 2010. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* 60: 150–160.

Zenil-Ferguson R, Ponciano JM, Burleigh JG. 2017. Testing the association of phenotypes with polyploidy: an example using herbaceous and woody eudicots. *Evolution* 71: 1138–1148.

Zhan SH, Drori M, Goldberg EE, Otto SP, Mayrose I. 2016. Phylogenetic evidence for cladogenetic polyploidization in land plants. *American Journal of Botany* 103: 1252–1258.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Ploidy and breeding system data according to three different classifications.

Fig. S2 Twenty-nine models of diversification are proposed for the study of ploidy, breeding systems and hidden states linked to the process of diversification.

Fig. S3 Posterior distributions for each of the parameters in the ploidy-only model (M1).

Fig. S4 Posterior distributions for each of the parameters in the ploidy and hidden-trait model (M4).

Fig. S5 Posterior distributions for each of the parameters in the breeding system-only model (M11).

Fig. S6 Posterior distributions for each of the parameters in the breeding system and hidden-trait model (M14).

Fig. S7 Posterior distribution for each of the parameters in the ploidy and breeding system model (M16).

Fig. S8 Posterior distribution for each of the parameters in the ploidy, breeding system and hidden-trait model (M19).

Fig. S9 Ancestral state estimation showing the maximum *a posteriori* estimates of the marginal probability distributions for each of the 650 internal nodes under the ploidy-only model (M1).

Fig. S10 Ancestral state estimation showing the maximum *a posteriori* estimates of the marginal probability distributions for each of the 650 internal nodes under the ploidy and hidden states model (M4).

Fig. S11 Ancestral state estimation showing the maximum *a posteriori* estimates of the marginal probability distributions for each of the 650 internal nodes under the breeding system-only model (M11).

Fig. S12 Ancestral state estimation showing the maximum *a posteriori* estimates of the marginal probability distributions for each of the 650 internal nodes under the breeding system and hidden states model (M14).

Fig. S13 Ancestral state estimation showing the maximum *a posteriori* estimates of the marginal probability distributions for

each of the 650 internal nodes under the ploidy and breeding system model (M16).

Fig. S14 Ancestral state estimation showing the maximum *a posteriori* estimates of the marginal probability distributions for each of the 650 internal nodes under the ploidy, breeding systems and hidden states model (M19).

Fig. S15 Testing the addition of breeding system to ploidy models.

Fig. S16 Testing the addition of ploidy to breeding system models.

Fig. S17 Effect of asymmetric rates in hidden models.

Fig. S18 Posterior distributions for the net diversification rates of the preferred models with diploidization.

Table S1 Bayes factors in log-scale of ploidy-only models with diploidization.

Table S2 Bayes factors of ploidy and breeding system without diploidization.

Table S3 Bayes factors of ploidy and breeding system with diploidization.

Table S4 Test for addition of hidden states in models via Bayes factors.

Table S5 Test for asymmetry of the hidden-trait transition rates via Bayes factors.

Table S6 Test for inclusion of a diploidization rate via Bayes factors.