New Phytologist

Interaction between ploidy, breeding system, and lineage diversification

Journal:	*New Phytologist*
Manuscript ID	Draft
Manuscript Type:	MS - Regular Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Zenil-Ferguson, Rosana; University of Minnesota Twin Cities, Department of Ecology Evolution and Behavior Burleigh, J. Gordon; University of Florida, Biology Freyman, William; University of Minnesota, Department of Ecology, Evolution, and Behavior Igic, Boris; University of Illinois at Chicago, Biological Sciences Mayrose, Itay; Tel Aviv University, School of Plant Sciences and Food Security Goldberg, Emma; University of Minnesota, College of Biological Sciences
Key Words:	Polyploidy, Breeding systems, Self-incompatibility, Diversification, Ploidy, SSE models, Diploidization

INTERACTION BETWEEN PLOIDY, BREEDING SYSTEM, AND LINEAGE DIVERSIFICATION

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

[1]Department of Biology, University of Hawai'i Mānoa, Honolulu, HI 96822, U.S.A.

[2]Department of Biology, University of Florida, Gainesville, FL 32611, U.S.A.

[3]Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, U.S.A.

[4]Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, U.S.A.

[5]School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv 6997801, Israel.

[†]*Author for correspondence*.

Running head: Ploidy and Breeding Systems in Solanaceae

Keywords: Polyploidy, Breeding Systems, Self-incompatibility Diversification, SSE models.

Abstract

If particular traits consistently affect rates of speciation and extinction, broad macroevolutionary patterns can be understood as consequences of selection at high levels of the biological hierarchy. Identifying traits associated with diversification rate differences is tricky, however, because of the the possibly large selection of traits under consideration and the resulting statistical challenge of testing for associations from comparative phylogenetic data. Ploidy (diploid vs. polyploid states) and breeding system (self-incompatible vs self-compatible states) have been repeatedly suggested as possibly drivers of differential diversification. We investigate the role of these traits, including their interaction, on speciation and extinction rates in Solanaceae. We find that the effect of ploidy can largely be explained by its correlation with breeding system, and that additional unknown factors work with breeding system to determine diversification rates. These results are largely robust to assumptions about whether diploidization occurs, but its inclusion generates interesting outcomes. Finally, we show that allowing for state-dependent diversification affects conclusions about the relative contribution of different evolutionary pathways to self-compatible polyploids.

"Among life history traits, reproductive characters that determine mating system patterns are perhaps the most influential in governing macroevolution."

Barrett et al. (1996)

Introduction

Species accumulate across the tree of life at different rates. A possible explanation for this phenomenon is that species possess various traits or character states that differentially affect rates of diversification. Dramatic increases in the available phylogenetic and phenotypic data, as well as methodological advances, have greatly accelerated the search for traits that influence diversification traits. Nevertheless, identifying focal traits associated with rates of speciation and extinction remains a difficult problem (e.g., Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Moore et al. 2016; FitzJohn et al. 2009; Goldberg and Igić 2012; Beaulieu and O'Meara 2016; Rabosky and Goldberg 2017). Problems arise at least in part because speciation and extinction may not depend on a single trait, and the context in which traits occur can lead to complex interactions resulting in a heterogenous speciation and extinction rates (Beaulieu and O'Meara 2016; Caetano et al. 2018; Herrera-Alsina et al. 2018). Consequently, examining the association of only one trait with diversification patterns can be misleading.

Polyploidization is a remarkably common mutation in plants (Husband et al. 2013; Zenil-Ferguson et al. 2017). The widespread prevalence of variation in ploidy has long been considered a salient feature of flowering plant lineages (Stebbins 1938). Increase in ploidy can alter many traits and impact a variety of evolutionary (Ramsey and Schemske 2002) and ecological processes (Sessa 2019). Polyploids are thought to have lower diversification rate than diploids (Mayrose et al. 2011, 2015). Recent studies, however, find common and 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 (WGDs) have played an important macroevolutionary role driving innovation and diversification in plants. Evidence of paleo-polyploidy within the genomes of diploid extant plants also implies pervasiveness of diploidization, the return of polyploids to the diploid state, throughout the angiosperm phylogeny (Soltis et al. 2015; Dodsworth et al. 2015). This could affect the interpretation of different net diversification rates for diploids and polyploids, especially if diploids have a relatively higher net diversification rate if they were ancestrally polyploid than if they had no polyploid ancestry.

Breeding system shifts—changes in the collection of physiological and morphological traits that determine the likelihood that any two gametes unite—are remarkably common and crucially affect the distribution and amount of genetic variation in populations (Stebbins 1974; Barrett 2013). In particular,

self-incompatibility (SI) causes plants to reject their own pollen, and loss of self-incompatibility, yielding self-compatibility (SC), is one of the most replicated transitions 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 (de Vos et al. 2014), as do outcrossing lineages in Onagraceae (Freyman and Höhna 2018). Despite consistent inferences of their macroevolutionary role, breeding systems are clearly not the sole trait determining diversification rates, and may exhibit complex interactions with other traits.

Ploidy, for instance, is widely thought to affect the propensity for self-fertilization (Stebbins 1950). Whole genome duplication may help facilitate the loss of SI and the transition to selfing (Barringer 2007; Barrett 2008; Husband et al. 2008), or it may directly cause the loss of SI (Stout and Chandler 1942; Lewis 1947). Doubled number of alleles in pollen is thought to effect disruption of the genetic mechanisms in gametophytic SI systems, which prevent self-fertilization (Entani et al. 1999; Tsukamoto et al. 2005; Kubo et al. 2010). With exceedingly few exceptions (Hauck et al. 2002; Nunes et al. 2006), this generally precludes the existence of SI polyploids in the most widespread system of SI, which is regarded as ancestral in eudicots (Igić and Kohn 2001; Steinbachs and Holsinger 2002). This system, RNase-based SI, is expressed in all SI species of Solanaceae examined to date. There is consequently a strong correlation between polyploidy and SC in this family (Robertson et al. 2011).

Given that changes in ploidy and breeding systems are causally related, and may have profound effects on the fate of lineages, it seems particularly profitable to examine possible interactions in their macroevolutionary effects. If the effects of ploidy and breeding system on lineage diversification are not additive, studying each trait separately may not reveal the effects of their combinations. As a result, it may be of interest to establish their joint influence on speciation and extinction rates. The two traits may also affect each other's state transitions. For example, losses of SI in diploids could be caused by polyploidization, resulting in a one-step pathway to polyploid SC species. Alternatively, the path from SI diploids to SC polyploids may involve a transition to SC, without an increase in ploidy, and then a subsequent polyploidization. Robertson et al. (2011) found that the pathway from SI diploids to SC polyploids is dominated by the two-step pathway over long timescales, but generally proceeds in one step via polyploidization of SI species over short timescales. It remains to be seen whether these results hold true when allowing for the influence of these traits on lineage diversification.

Two recent methodological developments enable disentangling the complex interactions between two or more traits and diversification. FitzJohn (2012) extended the state-dependent diversification "*SSE" modeling framework (Maddison et al. 2007) to allow estimating the effect of an arbitrary number of states

(enabling multi-state, MuSSE, instead of binary state, BiSSE, speciation and extinction models). This allows multiple states to represent trait combinations, particularly useful if one is interested in explicitly testing whether combinations of traits non-additively affect speciation or extinction. Soon thereafter, Beaulieu and O'Meara (2016) proposed an extension of the *SSE framework by introducing a model with unspecified, or 'hidden' traits (HiSSE model), which can affect the diversification process. Because variation in speciation and extinction rates may not be solely due to the focal trait, HiSSE simultaneously enables hidden states to explain such variation, and it reduces the elevated false-positive error rates. A joint application of these advances allows simultaneous consideration of the effects of multiple known traits on the diversification process, along with the presence of other unobserved states that can affect speciation and extinction rates.

Here, we examine the roles of ploidy and breeding system, as well as their interaction, on the diversification process in Solanaceae. Using extensive trait data, and considering other possible sources of heterogeneity in the diversification process, we assess the influence of these traits on speciation and extinction rates. First, we present models in which the effect of ploidy and breeding system are considered separately, and we compare the inferences to previously published results. Second, we add a hidden trait to each of those models and investigate whether the focal trait still explains differences in diversification. Third, we model ploidy and breeding system jointly to assess their combined relationship to diversification, with or without an additional hidden trait. With this joint trait model, we also quantify the relative contributions of the two pathways from SI diploids to SC polyploids. For all of the models involving ploidy, we investigate the potential effects of including diploidization. Our results highlight the importance of considering non-additive effects of traits on net diversification rates, especially when there are strong biologically driven correlations among them.

Methods

Ploidy and breeding system data

Chromosome number data were obtained for all Solanaceae taxa in the Chromosome Counts Database (CCDB; Rice et al. 2015), and the ca. 14,000 records were curated semi-automatically using the CCDBcurator R package (Rivero et al. 2019). CCDB contains records from original sources that have multiple complex symbol patterns denoting multivalence, or irregularites of chromosome counts. After a first full round of curation via CCDBcurator, we contrasted results by hand and corrected chromosomes where CCDBcurator's output was not able to adequately correct records. Our hand-curated records were also contrasted against the ploidy dataset from Robertson et al. (2011) and against ploidy data in the C-value DNA dataset from Bennett and Leitch (2005). 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 and Leitch 2005). For taxa without ploidy information but with information about chromosome number, we assigned ploidy based on the multiplicity of chromosomes within the genus or based on SI/SC classification. For example, *Solanum betaceum* did not have information about ploidy level but it has 24 chromosomes, so since $x = 12$ is the base chromosome number of the genus *Solanum* (Olmstead and Bohs 2007), we assigned *S. betaceum* as diploid. Additionally, because of the absence of SI polyploids (discussed further 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 was scored as self-incompatible (I) or self-compatible (C) based on results curated from the literature and original experimental crosses (as compiled in Igić et al. 2006; Goldberg et al. 2010; Robertson et al. 2011; Goldberg and Igić 2012). Many 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.

The Supplementary Information contains citations for the numerous original data sources. Resolution of taxonomic synonymy followed the conventions provided in 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. The single instance of an SI polyploid individual appears to be an allopentaploid hybrid of *Solanum oplocense* Hawkes x *Solanum gourlayii* Hawkes, reported by Camadro and Peloquin (1981). Under exceedingly rare circumstances, it is possible for polyploids containing multiple copies of S-loci to remain SI, so long as they express a single allele at the S-locus (discussed in Robertson et al. 2011). Because of the resulting absence of SI and polyploid populations, as well as the linked functional explanation for disabling of gametophytic self-incompatibility systems with non-self recognition, following whole genome duplication (reviewed in Ramsey and Schemske 1998; Stone 2002), we consider only three observed character states: self-incompatible diploids (ID), self-compatible diploids (CD), and polyploids which are always self-compatible (CP).

Matching our character state data to the largest time-calibrated phylogeny of Solanaceae (Särkinen et al. 2013) yielded 595 species with ploidy and/or breeding system information on the tree. Binary or three-

state classification of ploidy and breeding system for the 595 taxa is summarized in Fig. 1. We retained all of these 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. A total of 405 taxa without any information about breeding system or ploidy were excluded.

Models for ploidy and diversification

In order to investigate the association between ploidy level and diversification, we first defined a binary state speciation and extinction model (BiSSE, Maddison et al. 2007) in which taxa were classified as diploid (D) or polyploid (P). We call this the D/P ploidy model. The model parameters are speciation rates (λ_D, λ_P) and extinction rates (μ_D, μ_P) for each state, and a rate of transition from D to P (ρ). This is the same model fitted by Mayrose et al. (2011) but applied to the larger Solanaceae dataset.

Our second model addressed diversification due to ploidy differences while also parsing out the heterogeneity of diversification rates due a possible unobserved trait. Diversification rate differences explained by some trait other than ploidy were accommodated by adding a hidden state (HiSSE model; Beaulieu and O'Meara 2016). In this model, each of the observed diploid and polyploid states were subdivided by a binary hidden trait with states *A* and *B*. This model was called the D/P+A/B ploidy and hidden state model. Model parameters include speciation rates (λ_{D_A}, λ_{D_B}, λ_{P_A}, λ_{P_B}) and extinction rates (μ_{D_A}, μ_{D_B}, μ_{P_A}, μ_{P_B}) for each subdivided state. The polyploidization rate is again represented by ρ, and the rates of transition between hidden states are symmetrical with parameter α.

Models for breeding system and diversification

To assess the effects of breeding system in the diversification process, we first fit a model in which the states are self-incompatible (I) or self-compatible (C). This is the same as the analysis of Goldberg et al. (2010), but with an updated phylogeny (Särkinen et al. 2013). We call this BiSSE model the I/C breeding system model. It has parameters for state-specific speciation (λ_I, λ_C) and extinction (μ_I, μ_C), and for transitions from I to C (q_{IC}).

To parse out the effect of breeding system on diversification, while allowing for the possibility of heterogeneous diversification rates unrelated to breeding system, we subdivided each of those states into hidden states *A* and *B*. We call this HiSSE model the I/C+A/B breeding system and hidden state model. It also has parameters for state-specific speciation (λ_{I_A}, λ_{I_B}, λ_{C_A}, λ_{C_B}) and extinction (μ_{I_A}, μ_{I_B}, μ_{C_A}, μ_{C_B}), for loss of SI (q_{IC}), and for transitions between hidden states (α). An I/C+A/B model was used by Freyman and Höhna (2017) used to study diversification in Onagraceae, but here we restricted the hidden state transitions to be symmetrical.

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 non-self recognition, RNase-based gametophytic self-incompatibility (shared even with other euasterid families; Ramanauskas and 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). This represents very strong evidence that the I state is ancestral in Solanaceae, and that the SI mechanism was not regained within this family. Therefore, for all breeding system models, we estimated a transition rate from I to C but not the reverse ($q_{CI} = 0$).

Models for ploidy, breeding system, and diversification

Ploidy and breeding system might influence lineage diversification individually, but these two traits also have complex interactions in Solanaceae species. Therefore, we considered a multi-state model that investigates the contribution of both traits and the possible transitions between them. Self-incompatible diploids (ID), self-compatible diploids (CD), and polyploids, which are always self-compatible (CP) were used in a MuSSE model (FitzJohn 2012). We did not include a state for self-incompatible polyploids because they are not observed in the data, and that trait combination state is mechanistically predicted not to occur. We call this the ID/CD/CP ploidy and breeding system model. It has six parameters for diversification in each state (λ_{ID}, λ_{CD}, λ_{CP} for speciation, μ_{ID}, μ_{CD}, μ_{CP} for extinction), and three for transitions between states (ρ_I, ρ_C for polyploidization transitions from ID and CD to CP, respectively; q_{IC} for loss of self-incompatibility without polyploidization, from ID to CD). The total rate of loss of self-incompatibility, i.e., transitions out of ID, is $q_{IC} + \rho_I$.

The ID/CD/CP model could also suffer from a high type I error, like the two-state BiSSE model. We therefore added a hidden trait layer on top of our three-state model (analogous to Caetano et al. 2018; Huang et al. 2018). We refer to this as the ID/CD/CP+A/B model. A fully parameterized version of this model would have 26 rate parameters. Because our goal was to look for diversification rate differences associated with ploidy and breeding system rather than the specific effects of the hidden states, we fit a simplified version with 16 parameters. The reduction in parameter space was achieved by fixing the rates for transitions among hidden states to be equal with rate α, and fixing the transition rates between observed states to be independent of the hidden state (rates ρ_I, ρ_C, q_{IC} as defined for the ID/CD/CP model). There are additionally twelve diversification rate parameters ($\lambda_{ID_A}, \lambda_{ID_B}, \lambda_{CD_A}, \lambda_{CD_B}, \lambda_{CP_A}, \lambda_{CP_B}, \mu_{ID_A}, \mu_{ID_B}, \mu_{CD_A}, \mu_{CD_B}, \mu_{CP_A}, \mu_{CP_B}$).

Pathways to polyploidy

Considering ploidy and breeding system together, there are two evolutionary pathways from SI diploid to SC polyploid (Brunet and 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 maximum *a posteriori* (MAP) estimates of rates from the IC/CD/CP model.

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 only 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 (see below). First, we consider only the rates of transitions between these states, placing them in the propagation matrix Q. The matrix $\mathsf{P} = \exp(\mathsf{Q}t)$ then provides the 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). Our results will differ from those of Robertson et al. (2011) because our transition rate estimates come from a dated phylogeny and a model that allows for state-dependent diversification. 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. In particular, 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 propogation matrix. As before, matrix exponentiation provides the relative chance of changing from one state to any other state after time t. In this case, however, these are not probabilities because diversification changes the number of lineages as time passes. We can still use their ratios to consider the relative contribution of each pathway, though, analogous to the normalized age structure in a growing population.

Diploidization as an exploratory hypothesis

For the models with ploidy changes, we allowed diploidization by adding a parameter δ for transitions from P back to D (in the D/P and D/P+A/B models) or from CP back to CD (in the I/C and I/C+A/B models). We did not allow transitions from CP back to ID because that would include regain of SI. The models are denoted with $+\delta$ in Table 1. Previous modeling approaches (Mayrose et al. 2011) have argued against inferring diploidization rates when using ploidy data that comes from classifications based on chromosome

number multiplicity or chromosome number change models like chromEvol (Mayrose et al. 2010; Glick and Mayrose 2014; Mayrose et al. 2015; Freyman and Höhna 2017). These types of classifications do not allow for a ploidy reversion. Where indicated, the classification of ploidy for the data used in our models was based on chromosome multiplicity at the genus level (see `https://github.com/roszenil/solploidy` for archived dataset). However, the majority of the ploidy classifications were adopted from original studies with alternative sources of information (e.g., geographic distribution, genus ploidy distribution) where ploidy was defined by authors that found evidence for it. Since it is not clear whether diploidization can be detected under alternative ploidy classifications or even classifications based on chromosome number multiplicity at the genus level, we fit models with and without diploidization in order to test whether the conclusions about diversification are sensitive to including diploidization. As discussed by Servedio et al. (2014), the presence or absence of a process can have an exploratory goal. In our models, the diploidization parameter (or its absence, $\delta = 0$), represents an opportunity to explore a possibly significant assumption that may not decisively affect interactions among ploidy, breeding system, and diversification.

Statistical inference under the models

Parameters for each of the ten models described were estimated using RevBayes (Höhna et al. 2016). Code for analyses and key results are available at `https://github.com/roszenil/solploidy`. We used the method of FitzJohn et al. (2009) to account for incomplete sampling in all analyses by setting the probability of sampling a species at the present to $595/3000$ since the Solanaceae family has approximately 3,000 species as estimated by the Solanaceae Source project (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 $\left(\text{number of taxa}/[2 \times \text{root age}]\right)$ 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 for 96 hours in the high-performance computational cluster at the Minnesota Supercomputing Institute, which allowed for 5,000 generations of burn-in and a minimum of 200,000 generations of MCMC for each of the 10 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 model parameters, and also for the compound parameters of net diversification $(\lambda - \mu)$ and extinction fraction (μ/λ) for each state. Additionally, ancestral states at each node in the phylogeny were sampled jointly during the MCMC analyses. Ancestral state reconstructions for all models can be found in the Supplementary Information and show the maximum posterior estimates of

the marginal probability distributions for each of the 594 internal nodes for each of the 10 models.

Model selection

We calculated the marginal likelihood for each of the ten models in RevBayes (Höhna et al. 2016). Marginal likelihoods were calculated using 50 stepping stone steps under the methodology of Xie et al. (2010). Each stepping stone step was found by calculating 500 generations of burn-in followed by a total of 1,000 MCMC steps (Table 1). The calculation of each marginal likelihood ran for 24 hours on a high-performance computational cluster.

Using the marginal likelihood values, we calculated thirteen different Bayes factors. Six compared the models of ploidy against one other (D/P and D/P+A/B, each with or without diploidization), one compared the breeding system models (I/C and I/C+A/B), and six compared the models with both traits (ID/CD/CP and ID/CD/CP+A/B, each with or without diploidization) (Table 2). Other comparisons between these models are not valid because the input data are different under the different state space codings (Fig. 1). In mathematical terms, the D/P, I/C, and ID/CD/CP state spaces are not 'lumpable' with respect to one another (Tarasov 2018). Each model comparison is reported with a Bayes factor on the natural log scale: the comparison between models M_0 and M_1 is $BF(M_0,\ M_1) = \ln[P(\mathbf{X}|M_0) - P(\mathbf{X}|M_1)]$. There is 'positive' support for M_0 when this value is more than 2, 'strong' support when it is more than 6, and 'very strong' support when it is more than 10 (Kass and Raftery 1995).

Results

Traits and diversification

When analyzed separately, ploidy and breeding system each showed significant associations with net diversification differences in Solanaceae. When analyzed together, however, the effect of breeding system dominated the effect of ploidy, although hidden factors played an important role as well.

When considering ploidy alone (D/P model), we found a greater net diversification rate for diploids than for polyploids, in agreement with (Mayrose et al. 2011, 2015). This result holds with (Fig. 2A) or without (Fig. 3A) the diploidization parameter, although including diploidization shifts the net diversification rate of polyploids to be non-negative. Incorporating a hidden state in this model, however, removes the clear separation in diversification rate estimates between diploids and polyploids (D/P+A/B model; Fig. 2B, Fig. 3B). Thus, differences in net diversification rates are better explained by an unknown factor than by ploidy. Statistical model comparisons show very strong support for including the hidden state and strong support for including diploidization (Table 2).

When considering breeding system alone (I/C model, 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 (I/C+A/B model), the large net diversification rate difference persists for one of the hidden states but is removed for the other (Fig. 2D). Thus, differences in net diversification rates are best explained by both breeding system and an unknown factor (as in Onagraceae; Freyman and Höhna 2018). The statistical model comparison shows very strong support for including the hidden state (Table 2).

When considering ploidy and breeding system together (ID/CD/CP model), the net diversification rate for SI diploids was greater than for either SC diploids or SC polyploids, with or without diploidization (Fig. 2E, Fig. 3C). The difference in net diversification with breeding system persists when ploidy is included in the model, but not the reverse. That is, the association of ploidy with net diversification in the D/P model (Fig. 2A, Fig. 3A) appears to be driven by the subset of diploids that are SI; among SC species, net diversification rates for diploids and polyploids are similar. When a hidden state is included (IC/CD/CP+A/B model), the same general pattern remains when diploidization is prevented (Fig. 3D), although the higher net diversification rate of ID is less clear within one of the hidden states. With diploidization, the net diversification rate of ID is still greater than CD within each hidden state, but diversification estimate for P is highly uncertain and perhaps bimodal. Statistical model comparisons show very strong support for including the hidden state and at most positive support for including diploidization (Table 2).

Pathways to polyploidy

There are two pathways by which SI diploid lineages eventually—given enough time—become a SC polyploid. In the one-step pathway, polyploidization directly disables SI. In the two-step pathway, SI is first lost within the diploid state, followed by polyploidization. Determining the relative contribution of these pathways based on our estimated transition rates from the ID/CD/CP model, 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 (ID) lineage, when not much time has elapsed, the one-step pathway is more likely because it only necessitates a single event to effect transition to 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. S13). These conclusions, however, ignore the changes in numbers of lineages in a particular state from speciation and extinction events. By analogy, frequent sinking of a middle stepping stone in a path would lower the value of such a path. We therefore examined the effects of diversification rate on relative importance of each pathway.

Indeed, incorporation of lineage diversification rates changes conclusions regarding the relative contribution of each pathway. Beginning with a single ID lineage, assessment of relative pathway importance

necessitates consideration of proliferation in the ID state, as well as the CD state for the two-step pathway, and to track the transitions to the CP state. In this situation, even over long timescales, the two-step pathway contributes less to generation of CP (Fig. 4, right panels). 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.

Diploidization as an exploratory hypothesis

We considered models both with and without diploidization in order to explore its effects on the estimates of state-dependent diversification. For the two models that only include diploid and polyploid states (D/P and D/P+A/B), the net diversification rate of polyploid lineages is likely negative when diploidization (parameter δ) is excluded but positive when δ is included (Fig. 2A versus Fig. 3A). For the two models that also included breeding system (ID/CD/CP and ID/CD/CP+A/B), the main effect of including diploidization is greatly increasing the uncertainty of the estimated polyploid net diversification rate (Fig. 2E,F versus Fig. 3C,D). For all models, there is greater uncertainty in the estimate of diploidization rate than polyploidization rate, as judged by the width of the credibility intervals (see Figs. S3, S7, S15 and S19).

Discussion

Although species are composed of vast assemblages of variable traits, many traits are heritable, and they may affect the propensity of species to perish and multiply (Lewontin 1970), examining such macroevolutionary effects of many traits is rare, and remains challenging. We investigated the influence of two traits, ploidy and breeding system, in diversification of Solanaceae. Most significantly, we found that the effects of ploidy, inferred as significant when it is the only trait under consideration, are overshadowed by hidden traits, or breeding system when added to models. Interestingly, the difference in diversification rates between the diploids and polyploids is negligible on the background of self-compatibility. Furthermore, we recovered a smaller difference in diversification rates between self-incompatible and self-compatible lineages against the diploid genomic background. Motivated by the widespread recent findings of polyploidy in the ancestry of angiosperms, we also examined and found modest support for diploidization in the history of the family, but without significant effects on inferences of diversification dynamics. Disentangling the patterns of diversification linked to multiple traits is difficult, but our strategy allows tests of intricate hypotheses regarding a heterogenous diversification process. Below, we discuss key findings and highlight the most interesting and paradoxical implications. We provide an outline for broader application of phylogenetic comparative methods in the context of interactions among traits and diversification process.

Disentangling trait interactions in the context of diversification

Previous analyses of the effects of ploidy on diversification showed that diploids had greater net diversification rates than polyploids when exploring across multiple clades of the angiosperm phylogeny Mayrose et al. (2011, 2015). We obtain a broadly similar outcome, when considering only the effects of ploidy, but the results change significantly with joint consideration of hidden states, and breeding systems (Fig. 2A and B). The addition of a hidden trait results in blurring of differences in diversification rates between diploids and polyploids. Simultaneous consideration of breeding system better explains the variation of the inferred net diversification rates among ploidy states. SI, which is itself associated with a faster diversification rate, is the breeding system context in which many diploids occur. Among SC species, diploids and polyploids are associated with similar diversification rates (Fig. 2E). These results indicate that some of the effects of a hidden trait can be explained by each of the two traits under consideration. The finding is unique, and implications may be of wide interest, because it evaluates the relative influence of these two exceptionally interesting traits in a comparative phylogenetic framework.

Earlier work has shown that SI species in Solanaceae are associated with higher diversification rates compared to their SC counterparts (Goldberg et al. 2010). When we only considered the role of breeding system in the diversification process, we were able to replicate the slower net diversification for SC (Fig. 2C). Incorporating hidden states again resulted in SI associated with greater net diversification rates than SC in both hidden states (Fig. 2D). However, the overall picture is a bit more complex. Depending on the background hidden state, it is possible that a class of SC species may experience higher diversification rates than SI species. Moreover, these SC lineages may not constitute a dead end (with negative diversification rates).

Pathways to polyploidy

We find that SI diploid lineages are much more likely to take a one-step ID→CP pathway to polyploidy rather than a two-step ID→CD→CP pathway, in Solanaceae. This result is significant because it has a number of implications for the inference of history of breeding and sexual systems in flowering plants. First, support for the direct pathway lends credence to the idea that breakdown of SI by WGD—by generation of diploid 'heteroalleleic' pollen—may often trigger the evolution of gender dimorphism across angiosperms (Miller and Venable 2000). A more direct and definitive test of this hypothesis would additionally examine the propensity of polyploids generated through either pathway to become dioecious (Robertson et al. 2011). Second, statistical phylogenetic approaches measure lineage transition rates. Transitions depend on a combination of availability of individual mutation rates, which generate trait variation, and selection, which may

cause disproportionate loss or fixation of these mutants. Estimates of mutation rates are highly uncertain, but the total rate of breakdown of SI within diploids ranges is estimated to be on the order of $\mu_{IC} = 10^{-5}$ per pollen grain, and includes breakdown by tetraploidization (Lewis 1979), which is itself estimated to occur on the order of $\mu_{DP} = 10^{-5}$ (Ramsey and Schemske 1998). Seemingly, then, the simple genic mutation rate that leads to loss of SI is at best equal to tetraploidization mutation rate, and possibly far lower. We nevertheless recover transition rates from ID to CD (q_{IC}) threefold greater than ρ_I, indicating that selection restricts the fixation rate of new polyploids compared with new SC mutants (Robertson et al. 2011). Overall, our findings suggest that the effect of ploidy is largely explained explained by its correlation with breeding system, and that other unknown factors work with breeding system to determine diversification rates.

We implicitly modeled trait changes in both of these pathways as anagenetic, transitions occurring within lineages, without associated speciation events. However, losses of SI (I→C transitions) are often associated with cladogenesis in Solanaceae (Goldberg and Igić 2012), and polyploidization (D→P transition) also seem to co-occur with speciation events in a number of other families (Zhan et al. 2016; Freyman and Höhna 2017). Models that fail to consider such cladogenetic changes may make misleading estimates of anagenetic transition rates. In the only available data bearing on this problem, Goldberg and Igić (2012) found that extending the *SSE framework to allow cladogenetic changes did not substantially affect inference of net diversification rates. Nevertheless, a distinct possibility remains that allowing for cladogenetic transitions could tip the scales further in favor of the one-step pathway to SC polyploidy (CP), but this remains to be tested in future work. Specifically, the multi/hidden state SSE-based framework we employed here could be extended to incorporate cladogenetic transitions. Another useful line of inquiry would examine the path-dependent diversification rates, asking whether diversification rates of polyploid SC lineages differ depending on whether they were generated via a one- or two-step pathway.

Diploidization

The implications of the results of our diploidization analyses, as well as much of the other recent work on WGDs, are interesting because they conflict with other lines of evidence, especially those concerning the observed simple synteny of genomes in Solanaceae, and patterns of evolution as self-incompatibility loci. Flowering plant lineages are thought to have experienced at least one round of polyploidization in their evolutionary history (Soltis et al. 2015). Following polyploidization, it is possible for genome reorganization, downsizing, and loss to occur (Dodsworth et al. 2015; Zenil-Ferguson et al. 2016; Mandakova and Lysak 2018). As a consequence, nearly all extant species classified as “diploid" in our analyses, based on cytogenetic data, are possibly secondary diploids, having undergone both polyploidization and re-diploidization.

Unduly ignoring secondary diploidization would necessarily underestimate the rate of transitions from polyploids to diploids. It would then likely cause inflated net diversification rates for diploids, because species considered diploid may have been ancestrally polyploid, instead. However, we find only modest support for diploidization in the joint model (Table 2), and conclusions regarding the relative effects of ploidy and breeding system on diversification are robust to its inclusion. Furthermore, estimating the rate of diploidization, based on our ploidy level classifications, is highly uncertain (see parameter δ in Figs. S3, S7, S15 and S19), and yields dramatic changes in ancestral reconstructions (Figs. S2, S4, S6, S8, S14, S16, S18 and S20).

Some lines of evidence indicate that polyploidy may have occurred prior to the origin of Solanaceae, rendering all extant 'diploids' secondarily derived. First, Ku et al. (2000) and Blanc and Wolfe (2004) posited that the lineage leading to tomato, *Solanum lycopersicum*, may have experienced one or more of WGDs. A subsequent analysis of synteny between grape and *Solanum* genomes, as well as the distribution 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 (Tomato Genome Consortium 2012). The age of the peak of paralog Ks distances, is approximately 71$\pm$19 My (Tomato Genome Consortium 2012). If this is the case, then all of the genomes may have been subsequently re-diploidized, yielding the widespread and common chromosome numbers in this and related families, n=11-12, presently considered to be diploid (Robertson et al. 2011).

On the other hand, there is little evidence for the occurrence of re-diploidization within Solanaceae, since the origin of the family—the time frame we considered. First, the Ks-inferred duplication likely predates the origin of the family (49 My, HPD 46-53 My Särkinen et al. 2013). The best-supported, most-recent WGD Tomato Genome Consortium (2012) is older than the lineage under consideration in this study. At best then, the species we consider are ancestrally diploid, having re-diploidized after a suspected WGD, but likely before the origin of the family. Second, studies comparing map-based genome synteny within the family find no evidence for recent diploidization (Wu and Tanksley 2010). Instead, simple genome rearrangements appear sufficient to explain chromosomal evolution between a number of species, including all of those in the relatively cytogenetically conserved 'x=12' group, which includes tomato, potato, eggplant, pepper, and tobacco. Despite the limitations of assignment of ploidy levels, we found some support for models that include parameter δ, meaning that further study may be warranted, especially in other groups.

Regardless of whether genome polypoidization, followed by widespread diploidization, is a dominant mechanism of genome evolution in Solanaceae, it is clear that our present understanding of the evolution of ploidy and breeding systems is incomplete. An increasingly forceful weight of evidence seems

to support WGDs in the ancestry of many angiosperms. It is then easily inferred that many species have undergone diploidization, alongside genome size and chromosome number reductions. Inference of such a common genome upheaval in eudicot history seems to clash with the data indicating that a homologous mechanism of SI, which breaks down nearly invariably in natural and induced tetraploids (Stone 2002; McClure 2009), has been present continually in many families, including Solanaceae (Igić et al. 2006). Most problematically in this context, it is unclear how to explain the maintenance of trans-generic polymorphism at the orthologous S-loci in this and other families. As well, chromosome numbers and genome comparisons within the family (esp. 'x=12' clade, containing *Solanum* and *Nicotiana*) reveal strong conservation. Wu and Tanksley (2010) review the evidence from map-based genome comparisons and find that tomato and potato differ six inversions, tomato and eggplant by 24 inversions and five translocations, tomato and pepper by 19 inversions and six translocations, and tomato and tobacco by ca. 10 inversions and 11 translocations (likely underestimated). Recovery of such simple relationships would require outstanding convergent loss of duplicated segments. In either case, all of the approaches make ample assumptions, and it seems that at least some of them will necessitate deep revisions. Our study, for example, made numerous assumptions in coding trait states (e.g. what constitutes a polyploid or diploid, SI or SC species), in models used for analyses (e.g. irreversibility of SI loss underlain by trans-specific orthologous S-loci), and many more, which could each mislead inference.

Estimating rates of trait linked diversification models is not only a problem of difficult parametric inference (as discussed in Rabosky 2010; Beaulieu and O'Meara 2015), but also, a problem of inadequate "*SSE" model specification that can result in misleading inferences when the presence of a second trait has a complex interaction with the focus trait first used to model diversification. We presented a roadmap with a series of analyses that can identify whether unknown or unobservable traits are worth pursuing. In this roadmap, we have proposed fitting HiSSE models that have effectively shown that some binary states associated with diversification under BiSSE analyses are actually not different in net diversification terms, and instead a hidden state is linked to differences in the diversification rate (Beaulieu and O'Meara 2016). If that is the case in a study system, raising the question of the identity of the hidden state should be considered as the next step in the modeling process. When considering a second candidate trait, one can model the complex interactions via a MuSSE model, especially in systems where multiple traits are suspected to influence diversification, and where interactions between those traits are well understood. Since MuSSE might also suffer from high type I errors, considering more heterogeneity using a hidden state model on top of the multi-state diversification model is necessary to adequately infer the effect of the complex

interactions in the speciation and extinction process. Investigating the process of diversification linked to traits can be done via thorough statistical inferences that carefully evaluate multiple evolutionary histories. Background heterogeneity in lineage of diversification is the rule, rather than the exception. Inferences that question whether other complex trait interactions are hidden in this background heterogeneity are necessary because they critically contribute to the reconstruction and understanding of diversification.

Acknowledgements

This work was supported by the National Science Foundation (NSF DEB-1655478 to EEG and NSF DEB-1655692 to BI). The computing resources were provided by the Minnesota Supercomputing Institute (MSI) at the University of Minnesota.

Literature Cited

Barrett, S. C., 2008. Major evolutionary transitions in flowering plant reproduction: an overview.

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

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

Beaulieu, J. M. and B. C. O'Meara, 2015. Extinction can be estimated from moderately sized molecular phylogenies. Evolution 69:1036–1043.

———, 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst Biol 65:583–601.

Bennett, M. D. and I. J. Leitch, 2005. Plant DNA C-values database.

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

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

Caetano, D. S., B. C. O'Meara, and J. M. Beaulieu, 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution 72:2308–2324.

Camadro, E. and S. J. Peloquin, 1981. Cross-incompatibility between two sympatric polyploid *Solanum* species. Theoretical and Applied Genetics 60:65–70.

Dodsworth, S., M. W. Chase, and A. R. Leitch, 2015. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? Botanical Journal of the Linnean Society 180:1–5.

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

FitzJohn, R. G., 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

FitzJohn, R. G., W. P. Maddison, and S. P. Otto, 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Systematic Biology 58:595–611.

Freyman, W. and S. Höhna, 2018. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. Systematic Biology (in press; syy078).

Freyman, W. A. and S. Höhna, 2017. Cladogenetic and anagenetic models of chromosome number evolution: A Bayesian model averaging approach. Systematic Biology 67:195–215.

Glick, L. and I. Mayrose, 2014. Chromevol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Molecular Biology and Evolution 31:1914–1922.

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

Goldberg, E. E., J. R. Kohn, R. Lande, K. A. Robertson, S. A. Smith, and B. Igić, 2010. Species selection maintains self-incompatibility. Science 330:493–495.

Hastings, W. K., 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika 57:97–109.

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

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

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ron-

quist, 2016. Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology 65:726–736.

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

Husband, B. C., S. J. Baldwin, and J. Suda, 2013. The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. Pp. 255–276, *in* Plant Genome Diversity Volume 2. Springer.

Husband, B. C., B. Ozimec, S. L. Martin, and L. Pollock, 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., L. Bohs, and J. R. Kohn, 2006. Ancient polymorphism reveals unidirectional breeding system shifts. Proceedings of the National Academy of Sciences USA 103:1359–1363.

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

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

Ioerger, T. R., A. G. Clark, and T. H. Kao, 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proceedings of the National Academy of Sciences USA 87:9732–9735.

Kass, R. E. and A. E. Raftery, 1995. Bayes factors. Journal of the American Statistical Association 90:773–795.

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

Kubo, K.-i., T. Entani, A. Takara, N. Wang, A. M. Fields, Z. Hua, M. Toyoda, S.-i. Kawashima, T. Ando, A. Isogai, et al., 2010. Collaborative non-self recognition system in S-RNase-based self-incompatibility. Science 330:796–799.

Landis, J. B., D. E. Soltis, Z. Li, H. E. Marx, M. S. Barker, D. C. Tank, and P. S. Soltis, 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. American Journal of Botany 105:348–363.

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

———, 1979. Sexual Incompatibility in Plants, *The Institute of Biology's Studies in Biology*, vol. 110. Edward Arnold Limited.

Lewontin, R. C., 1970. The units of selection. Annual Review of Ecology and Systematics 1:1–18.

Maddison, W. P. and R. G. FitzJohn, 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Systematic Biology 64:127–136.

Maddison, W. P., P. E. Midford, and S. P. Otto, 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56:701–710.

Mandakova, T. and M. A. Lysak, 2018. Post-polyploid diploidization and diversification through dysploid changes. Current opinion in plant biology 42:55–65.

Mayrose, I., M. S. Barker, and S. P. Otto, 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. Systematic Biology 59:132–144.

Mayrose, I., S. H. Zhan, C. J. Rothfels, N. Arrigo, M. S. Barker, L. H. Rieseberg, and S. P. Otto, 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., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto, 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., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953. Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21:1087–1092.

Miller, J. S. and D. L. Venable, 2000. Polyploidy and the evolution of gender dimorphism in plants. Science 289:2335–2338.

Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck, 2016. Critically evaluating the theory and performance of bayesian analysis of macroevolutionary mixtures. Proceedings of the National Academy of Sciences USA 113:9569–9574.

Nunes, M. D., R. A. Santos, S. M. Ferreira, J. Vieira, and C. P. Vieira, 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, R. G. and L. Bohs, 2007. A summary of molecular systematic research in Solanaceae: 1982–2006. Acta Horticulturae Pp. 255–268.

PBI *Solanum* Project, 2012. Solanaceae Source: a global taxonomic resource for the nightshade family.

Rabosky, D. L., 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64:1816–1824.

Rabosky, D. L. and E. E. Goldberg, 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Systematic Biology 64:340–355.

———, 2017. FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. Evolution 71:1432–1442.

Raduski, A. R., E. B. Haney, and B. Igić, 2012. The expression of self-incompatibility in angiosperms is bimodal. Evolution 66:1275–1283.

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

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

———, 2002. Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics 33:589–639.

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

Rivero, R., E. B. Sessa, and R. Zenil-Ferguson, 2019. EyeChrom and CCDB curator: Visualizing chromosome count data from plants. Applications in Plant Sciences P. e01207.

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

Särkinen, T., L. Bohs, R. G. Olmstead, and S. Knapp, 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol 13:214.

Servedio, M. R., Y. Brandvain, S. Dhole, C. L. Fitzpatrick, E. E. Goldberg, C. A. Stern, J. Van Cleve, and D. J. Yeh, 2014. Not just a theory? The utility of mathematical models in evolutionary biology. PLoS Biology 12:e1002017.

Sessa, E. B., 2019. Polyploidy as a mechanism for surviving global change. New Phytologist 221:5–6.

Soltis, D. E., M. C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N. M. Miles, E. V. Mavrodiev, W. Mei, M. B. Cortez, P. S. Soltis, and M. A. Gitzendanner, 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of mayrose et al.(2011). New Phytologist 202:1105–1117.

Soltis, P. S., D. B. Marchant, Y. Van de Peer, and D. E. Soltis, 2015. Polyploidy and genome evolution in plants. Current Opinion in Genetics & Development 35:119–125.

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

———, 1950. Variation and Evolution in Plants. Columbia University Press, New York.

———, 1974. Flowering plants: evolution above the species level. Belknap Press of Harvard University Press, Cambridge, Mass.

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

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

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

Tarasov, S., 2018. Integration of anatomy ontologies and evo-devo using structured markov models suggests a new framework for modeling discrete phenotypic traits. BioRxiv P. 188672.

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

Tsukamoto, T., T. Ando, H. Watanabe, E. Marchesi, and T.-h. Kao, 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, J. M., C. E. Hughes, G. M. Schneeweiss, B. R. Moore, and E. Conti, 2014. Heterostyly accelerates diversification via reduced extinction in primroses. Proceedings of the Royal Society B: Biological Sciences 281:20140075.

Wu, F. and S. D. Tanksley, 2010. Chromosomal evolution in the plant family solanaceae. BMC Genomics 11:182.

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

Zenil-Ferguson, R., J. M. Ponciano, and J. G. Burleigh, 2016. Evaluating the role of genome downsizing and size thresholds from genome size distributions in angiosperms. American journal of botany 103:1175–1186.

———, 2017. Testing the association of phenotypes with polyploidy: An example using herbaceous and woody eudicots. Evolution 71:1138–1148.

Zhan, S. H., M. Drori, E. E. Goldberg, S. P. Otto, and I. Mayrose, 2016. Phylogenetic evidence for cladogenetic polyploidization in land plants. American Journal of Botany 103:1252–1258.

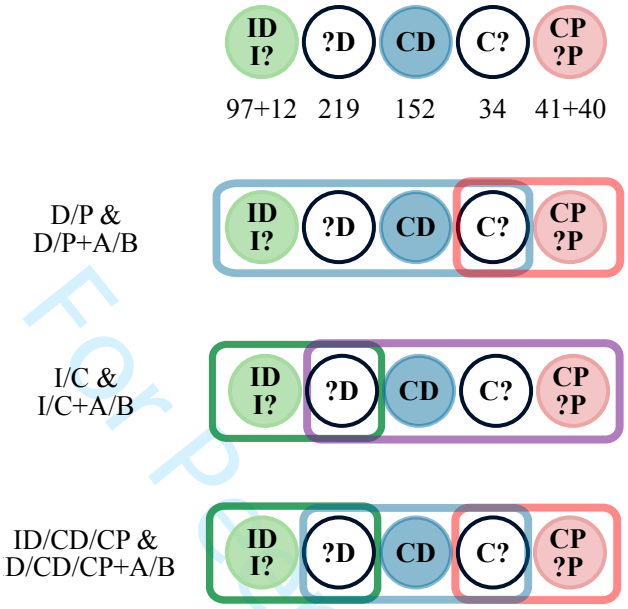

Figure 1: Character states used for each of the models. Each species retained on the tree belonged to one of five possible categories, depending on whether ploidy and/or breeding system were known. The number of species in each is shown under the corresponding circles in the top row. These categories were then grouped in a manner appropriate to the states of each model. For example, there are 34 species that are self-compatible (*C*) with unknown ploidy, and 40 species known to be polyploid(*P*), but without breeding system data; the latter can be inferred, because all polyploid species in the dataset are self-compatible (see Robertson et al. 2011, and text). In the D/P models, these are coded as either *D* or *P* (uncertain/consistent with either state); in the I/C models, they are coded as *C*; and in the ID/CD/CP models, they are coded as either *CD* or *CP*. In the hidden state models, species were coded as either *A* or *B*. The model-specific coding is shown in different rows. They each use state spaces that are not directly comparable with one another. For example, we cannot test whether the D/P model fits better than the I/C model because they use states that are not the same, and are therefore not 'lumpable' (Tarasov 2018).

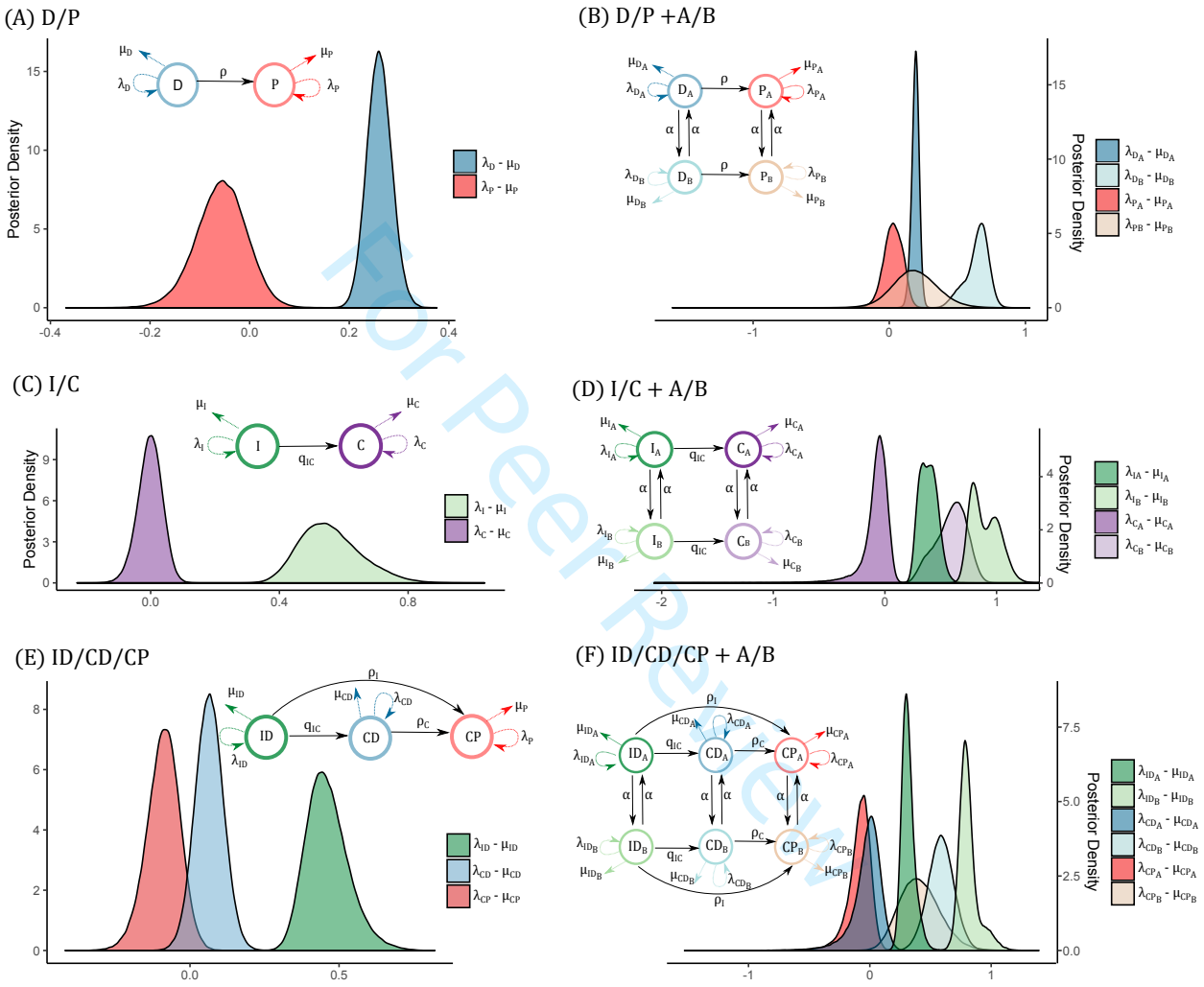

Figure 2: Net diversification rates for all models without diploidization rate estimate. Each panel contains a graphical summary of estimated model parameters and displays posterior distributions for net diversification estimates.

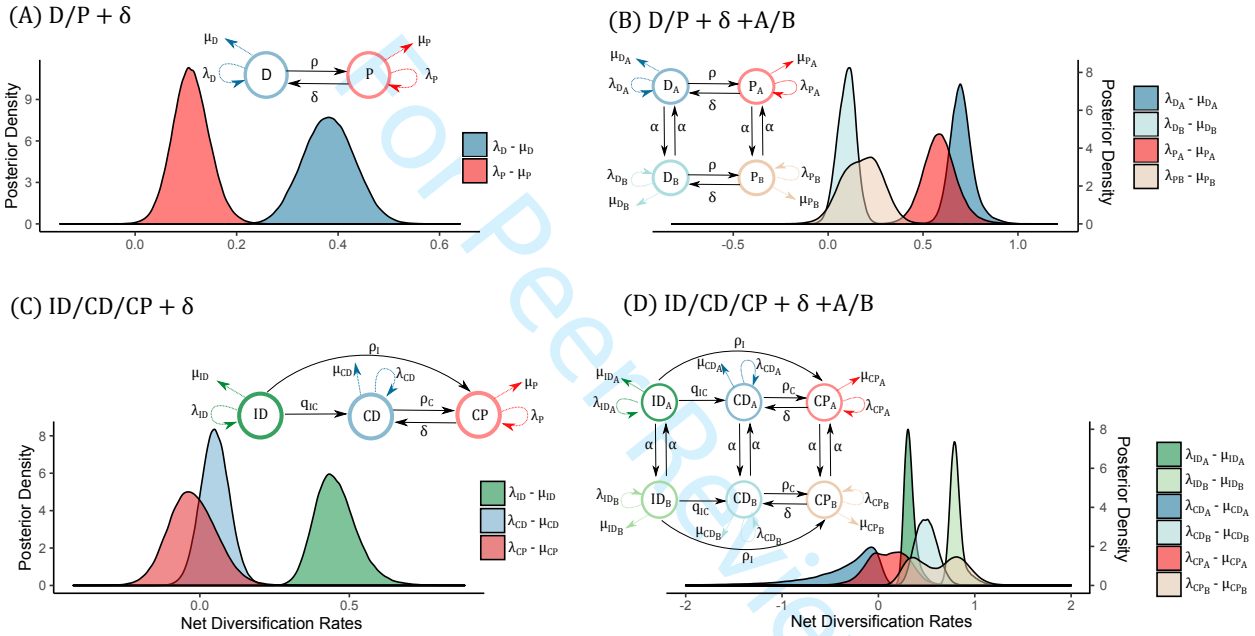

Figure 3: Net diversification rates for all models that include diploidization (δ). Each panel contains a graphical summary of estimated model parameters and displays posterior distributions for net diversification estimates.

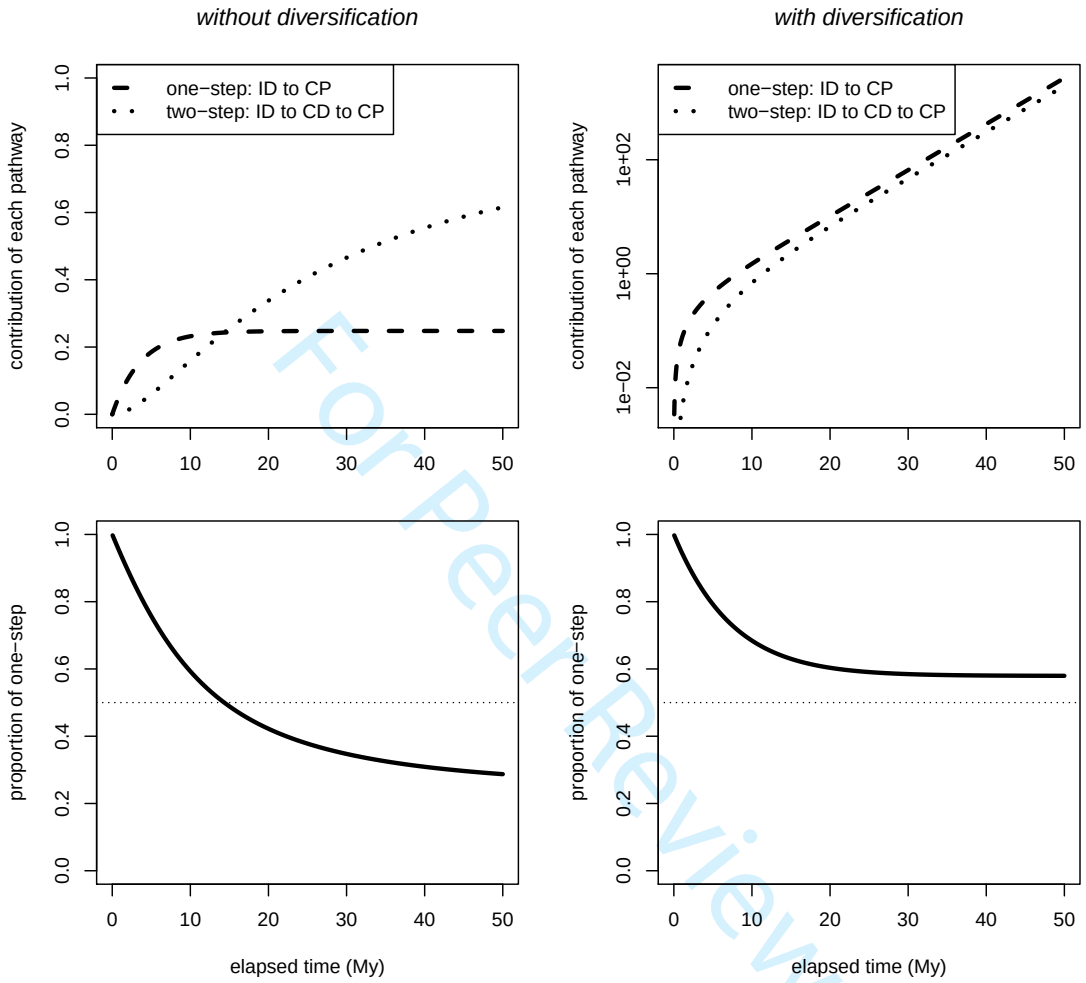

Figure 4: Contributions of the two pathways to polyploidy. The one-step pathway is direct ID→CP transitions. The two-step pathway consists of ID→CD→CP transitions. When considering only rates of transitions among the states (ignoring the diversification rate parameters), the one-step pathway dominates 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 polyploidization breaks down SI, dominates over any timescale (right panels). The top panels show the separate contributions of each pathway. The bottom panels show the proportional contribution of the one-step pathway (i.e., one-step / [one-step + two-step]).

Model	Ploidy	Diploidization	Breeding System	Hidden State	Num Parameters	Marginal Log-Likelihood
1. D/P	Yes	No	No	No	5	-1193.66
2. D/P +δ	Yes	Yes	No	No	6	-1182.93
3. D/P+A/B	Yes	No	No	Yes	10	-1150.99
4. D/P+A/B+δ	Yes	Yes	No	Yes	11	**-1145.69**
5. I/C	No	No	Yes	No	5	-1194.80
6. I/C+A/B	No	No	Yes	Yes	10	**-1155.37**
7. ID/CD/CP	Yes	No	Yes	No	9	-1345.87
8. ID/CD/CP+δ	Yes	Yes	Yes	No	10	-1344.50
9. ID/CD/CP+A/B	Yes	No	Yes	Yes	15	-1303.55
10. ID/CD/CP+A/B+δ	Yes	Yes	Yes	Yes	16	**-1300.35**

Table 1: The ten models and their marginal likelihoods. Values in bold are for the best models within each class that are comparable (see Table 2). Abbreviations are D: diploid, P: polyploid, I: self-incompatible, C: self-compatible, A: one state of hidden trait, B: other state of hidden trait, δ: diploidization.

Ploidy Models					Breeding System Models			Ploidy and Breeding System Models				
	1	2	3	4		5	6		7	8	9	10
1. D/P+δ	·	10.72	-37.24	-31.94	5. I/C	·	-39.43	7. ID/CD/CP+δ	·	1.36	-44.15	-40.95
2. D/P	·	·	-47.97	-42.66	**6. I/C+A/B**	·	·	8. ID/CD/CP	·	·	-45.51	-42.31
3. D/P+A/B+δ	·	·	·	5.30				**9. ID/CD/CP+A/B+δ**	·	·	·	3.2
4. D/P+A/B	·	·	·	·				10. ID/P/CD+A/B	·	·	·	·

Table 2: Bayes factors for model comparisons. Each of the three boxes contains models that can be compared with one another, based on the character states they include (see Fig. 1). Models are numbered in as Table 1. Bayes factors are reported on the natural log scale, so numbers greater than $+2$ mean that the model in the row label has 'positive' support relative to the model in the column label; numbers less than -2 mean that model in the column label is the preferred one. Conventional thresholds for 'strong' and 'very strong' support are 6 and 10, respectively. The best model in each set is written in bold. In each case, it is the most complex model of the set.

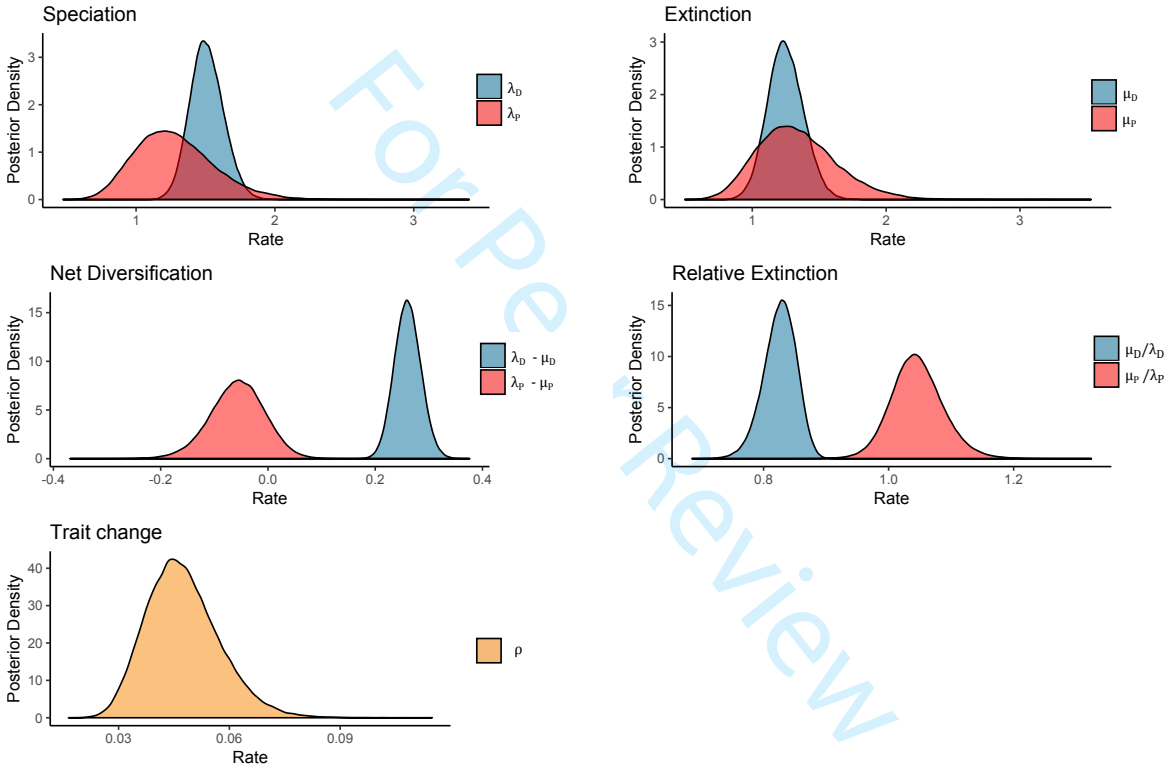

Figure S1: Posterior distribution for each of the parameters in the D/P ploidy model

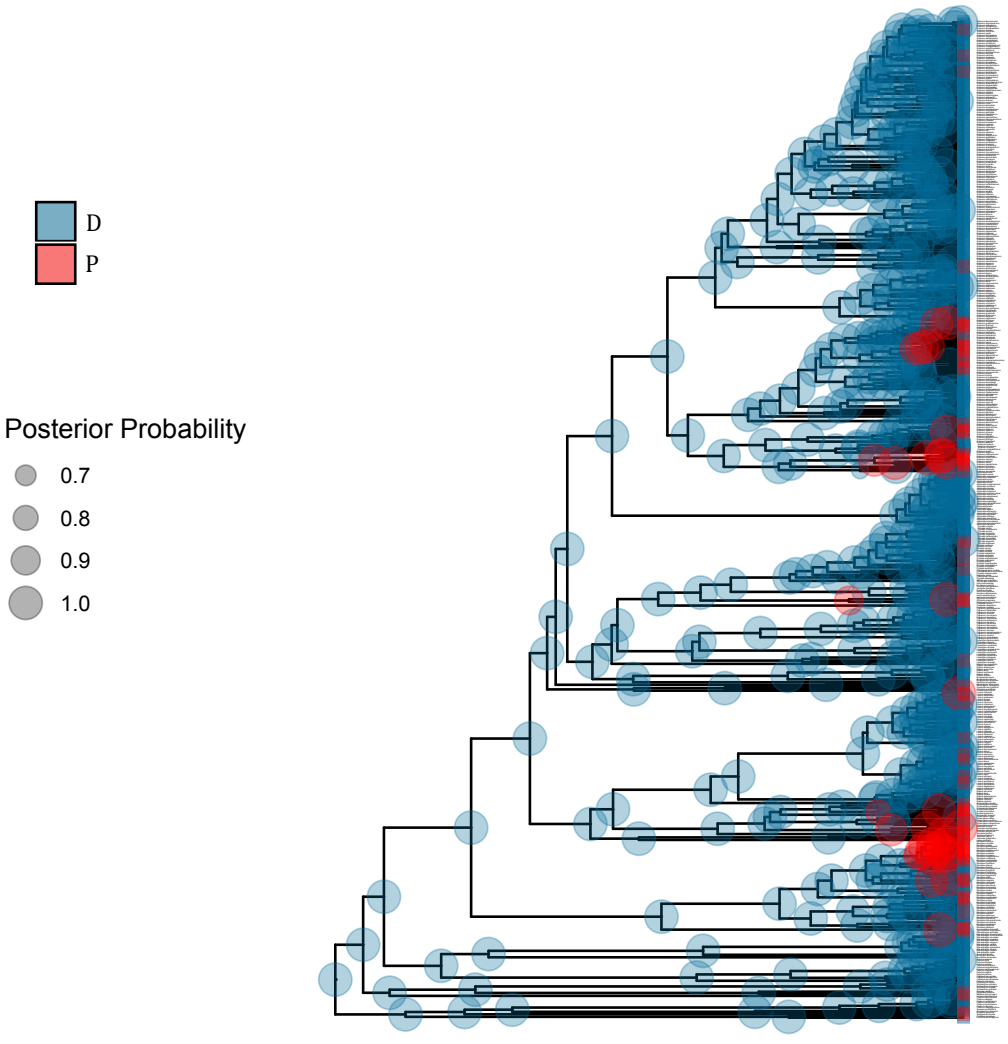

Figure S2: Ancestral state reconstruction showing the maximum a posteriori for each node of the D/P ploidy model

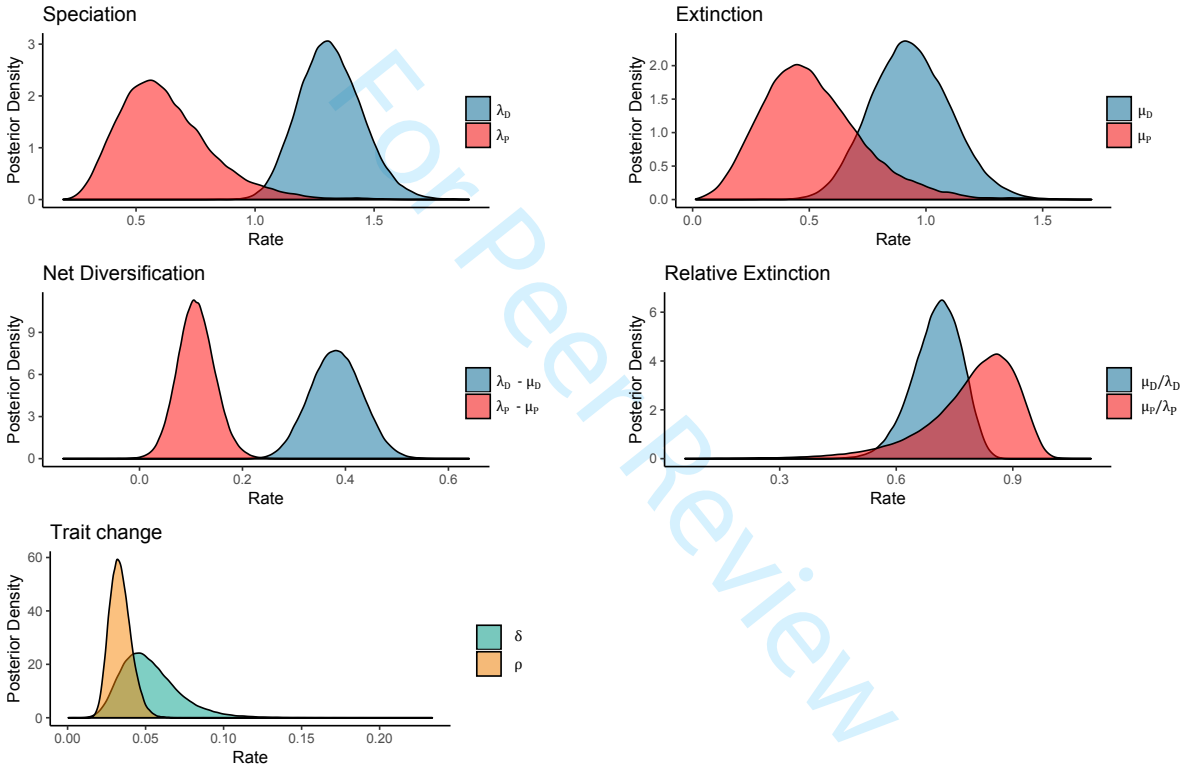

Figure S3: Posterior distribution for each of the parameters in the D/P+ δ ploidy model

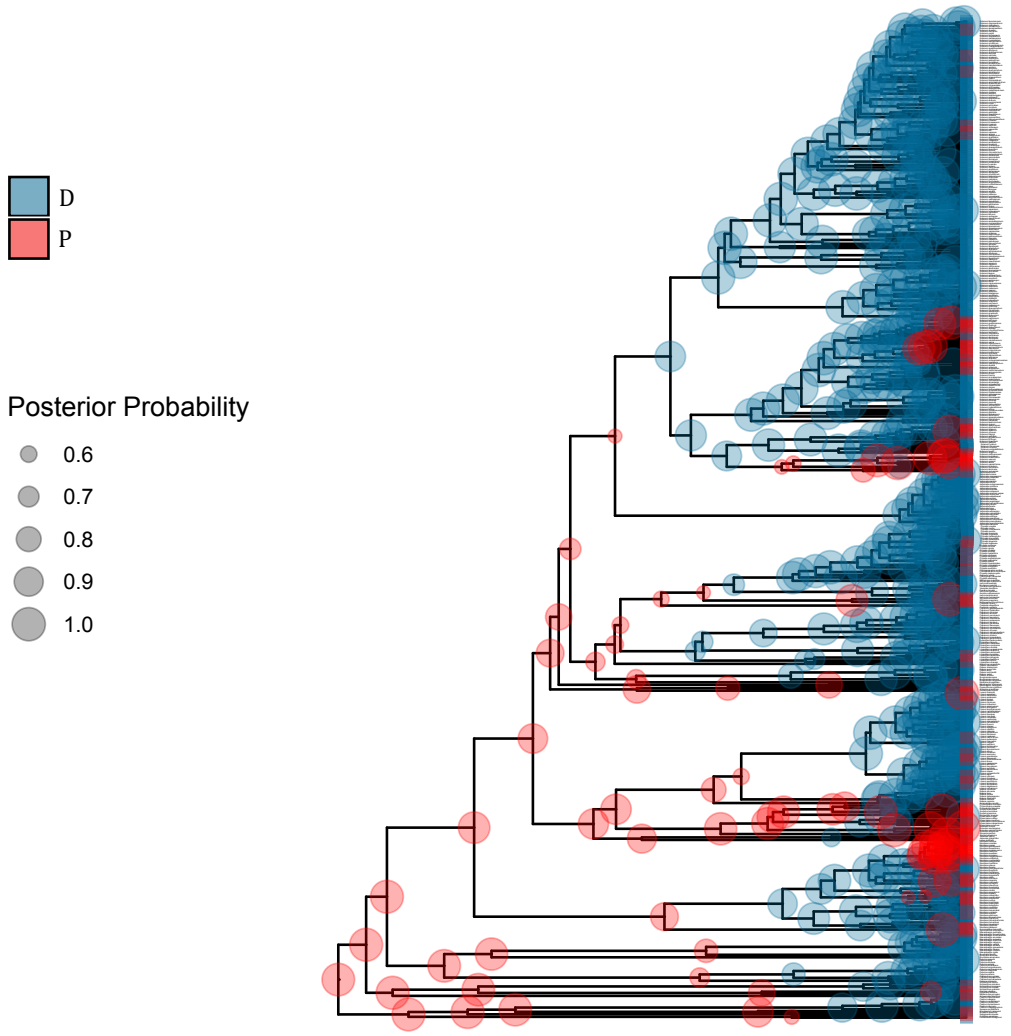

Figure S4: Ancestral state reconstruction showing the maximum a posteriori for each node in the D/P+ δ, ploidy model

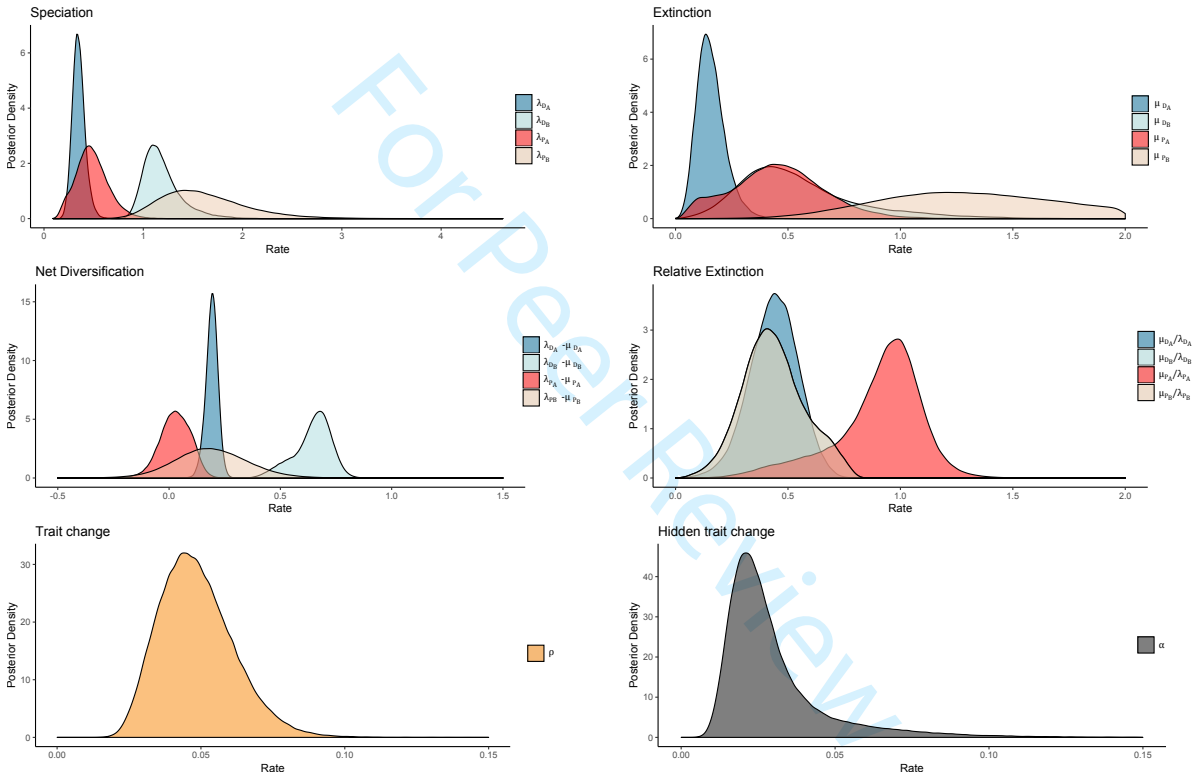

Figure S5: Posterior distribution for each of the parameters in the D/P+A/B ploidy model

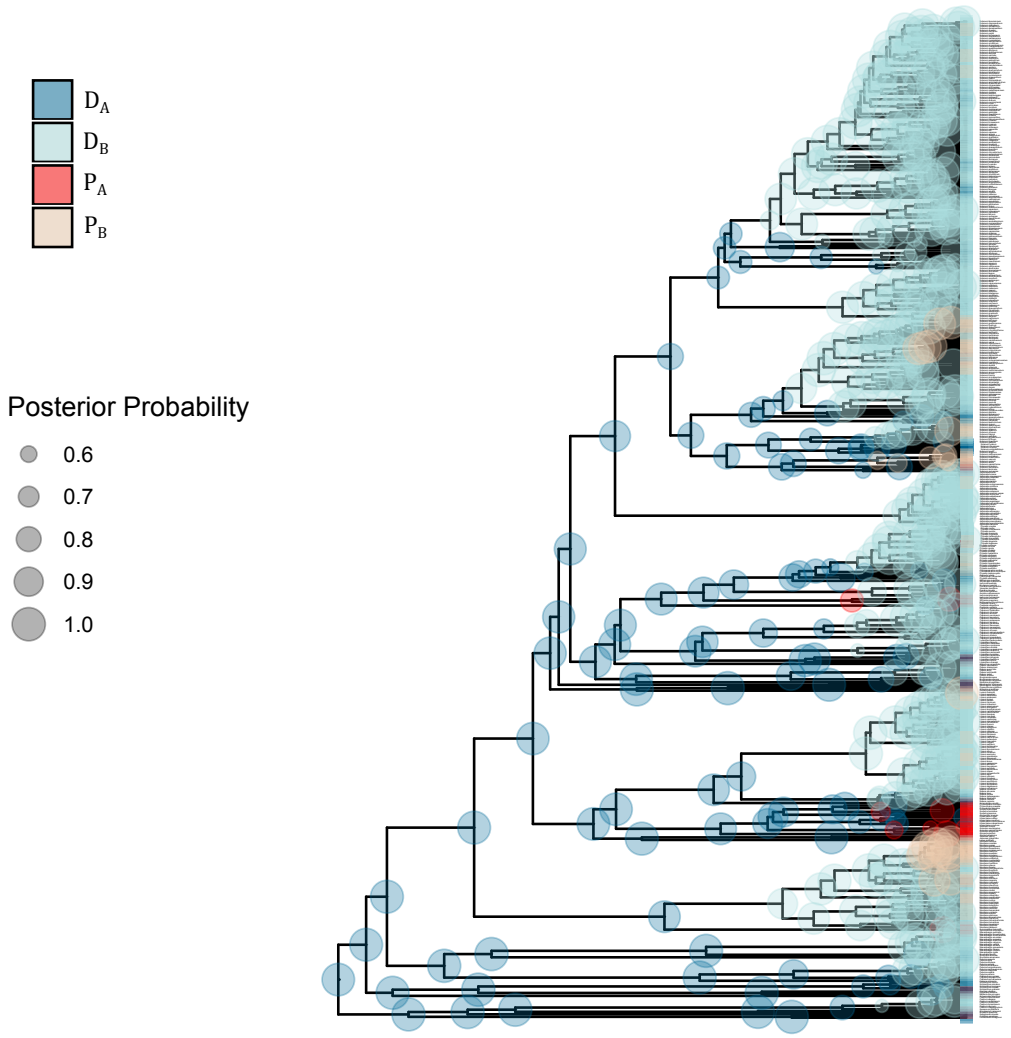

Figure S6: Ancestral state reconstruction showing the maximum a posteriori for each node in the D/P D/P+A/B ploidy model

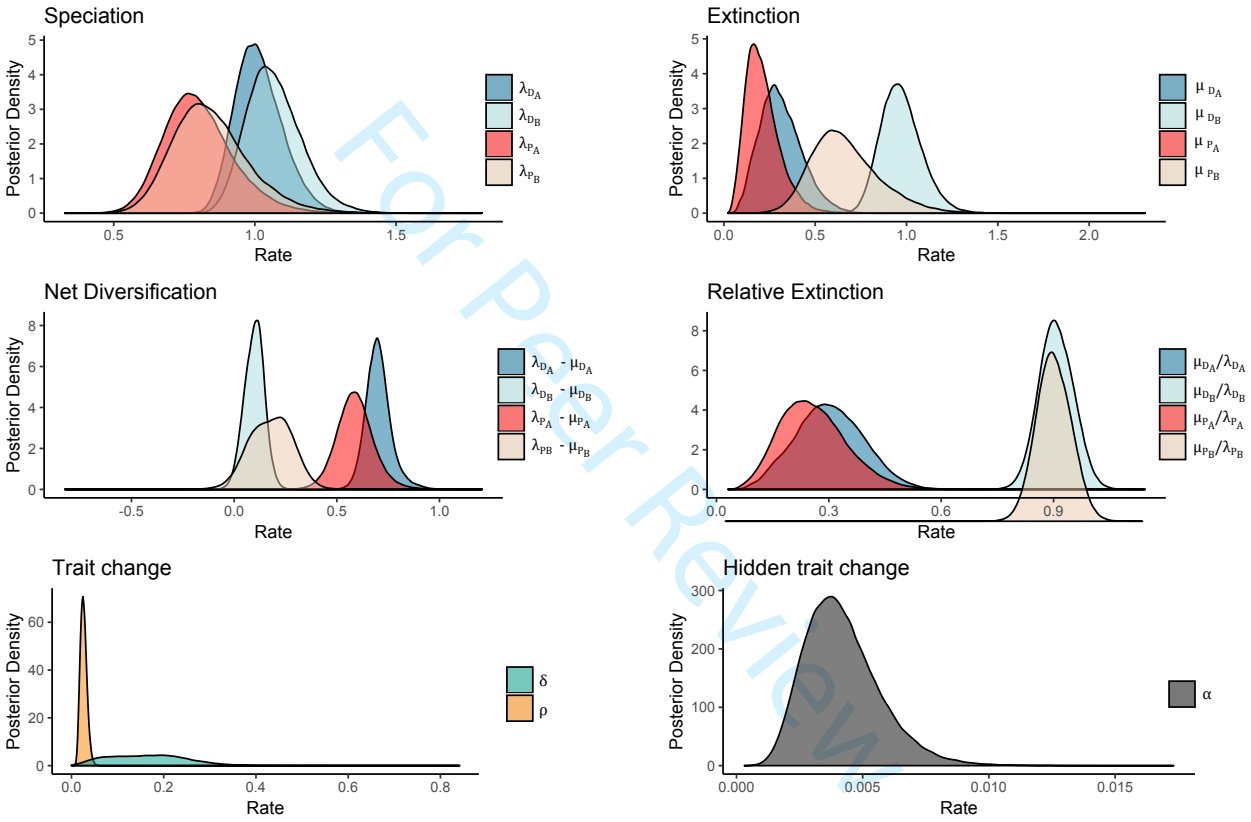

Figure S7: Posterior distribution for each of the parameters in the D/P+δ+A/B ploidy model. The axis is offset in one location so that the two overlapping distributions can be seen.

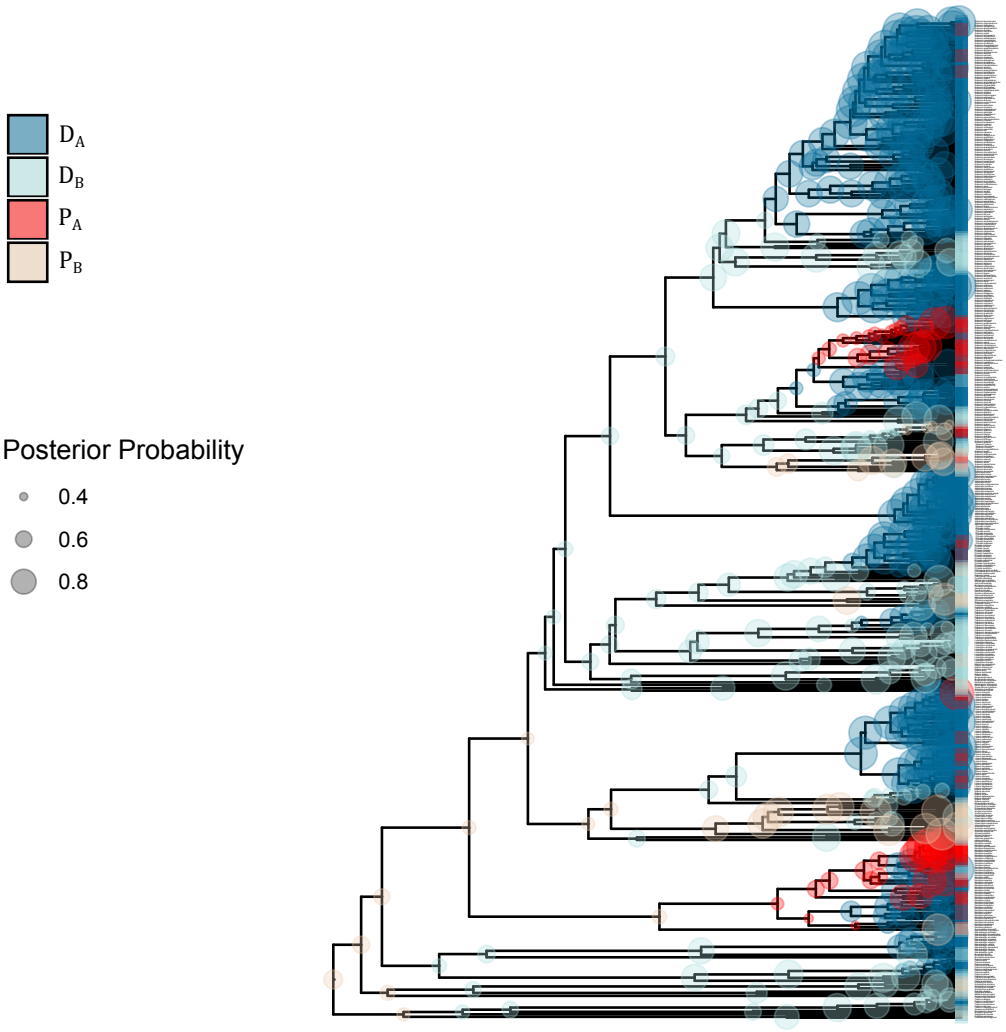

Figure S8: Ancestral state reconstruction showing the maximum a posteriori for each node in the D/P+δ+A/B ploidy model.

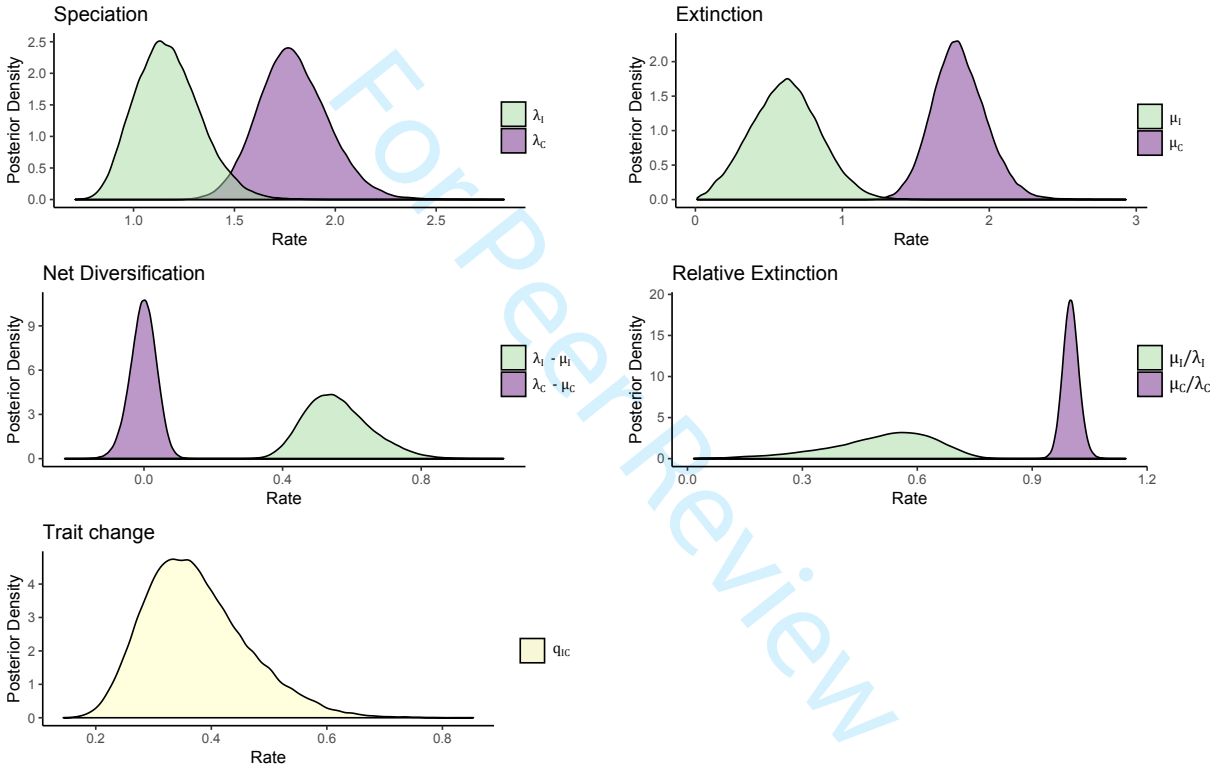

Figure S9: Posterior distribution for each of the parameters in the I/C breeding system model

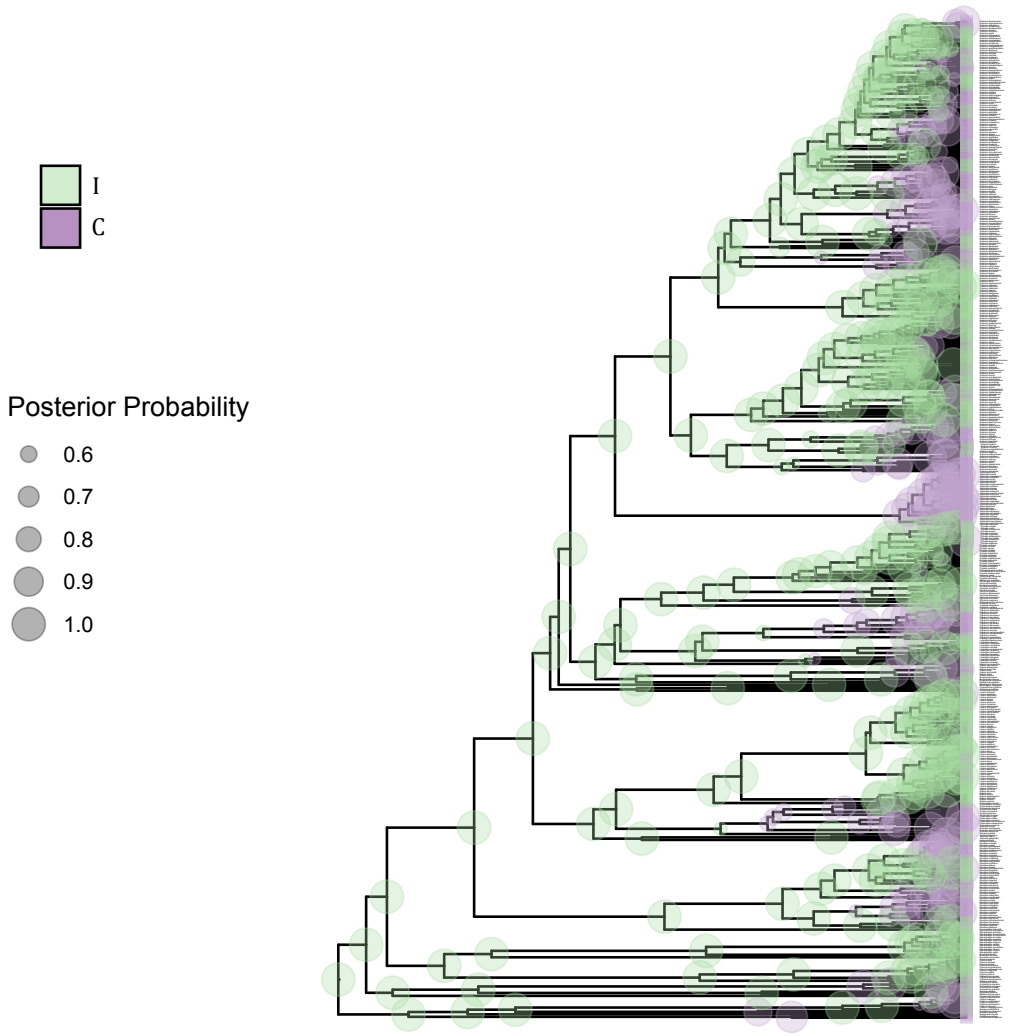

Figure S10: Ancestral state reconstruction showing the maximum a posteriori for each node in the I/C breeding system model

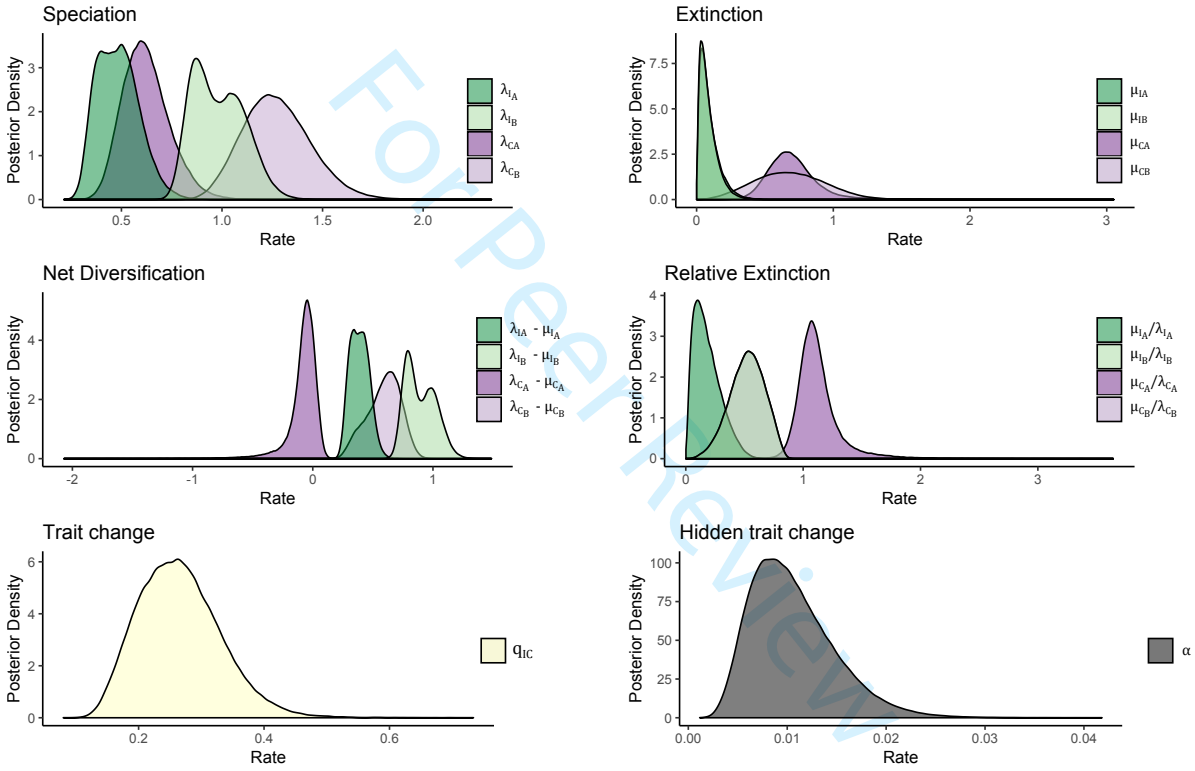

Figure S11: Posterior distribution for each of the parameters in the I/C+A/B, breeding system model

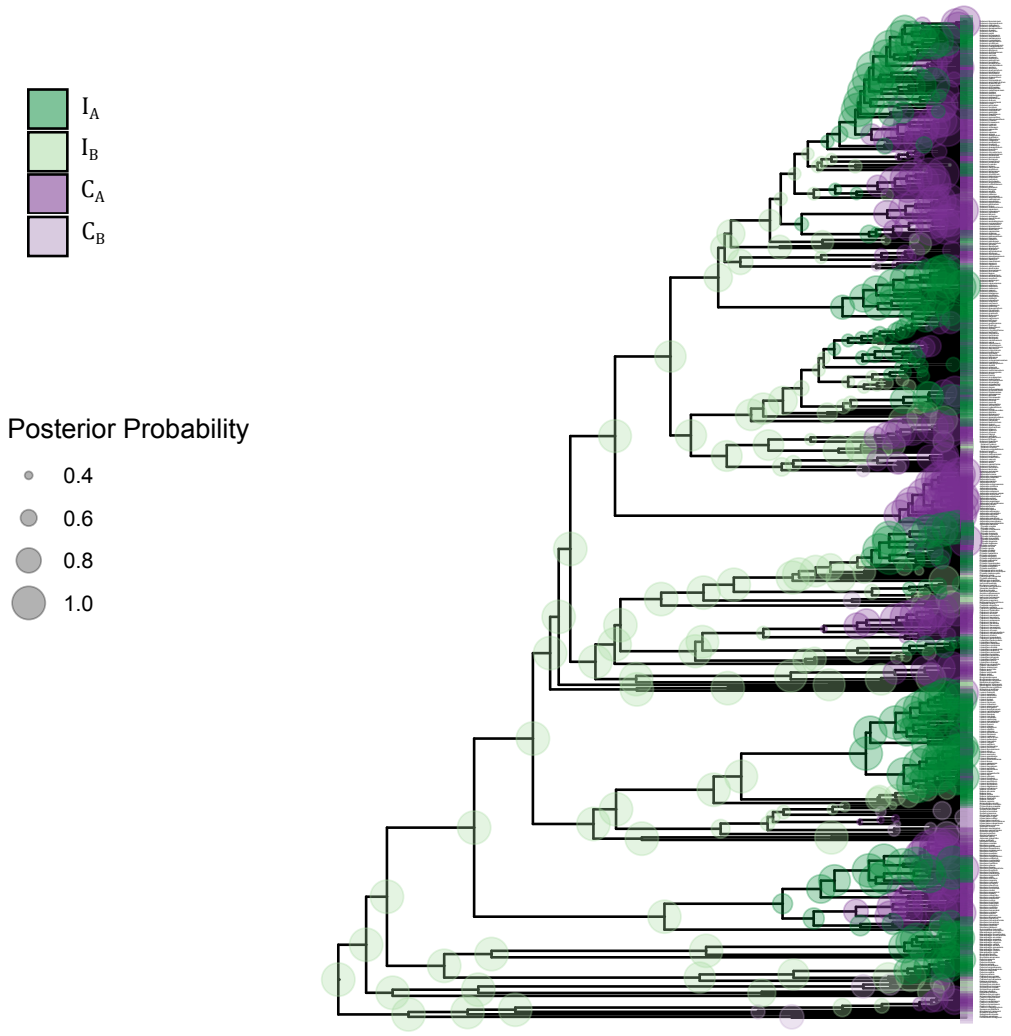

Figure S12: Ancestral state reconstruction showing the maximum a posteriori for each node in the I/C+A/B breeding system model

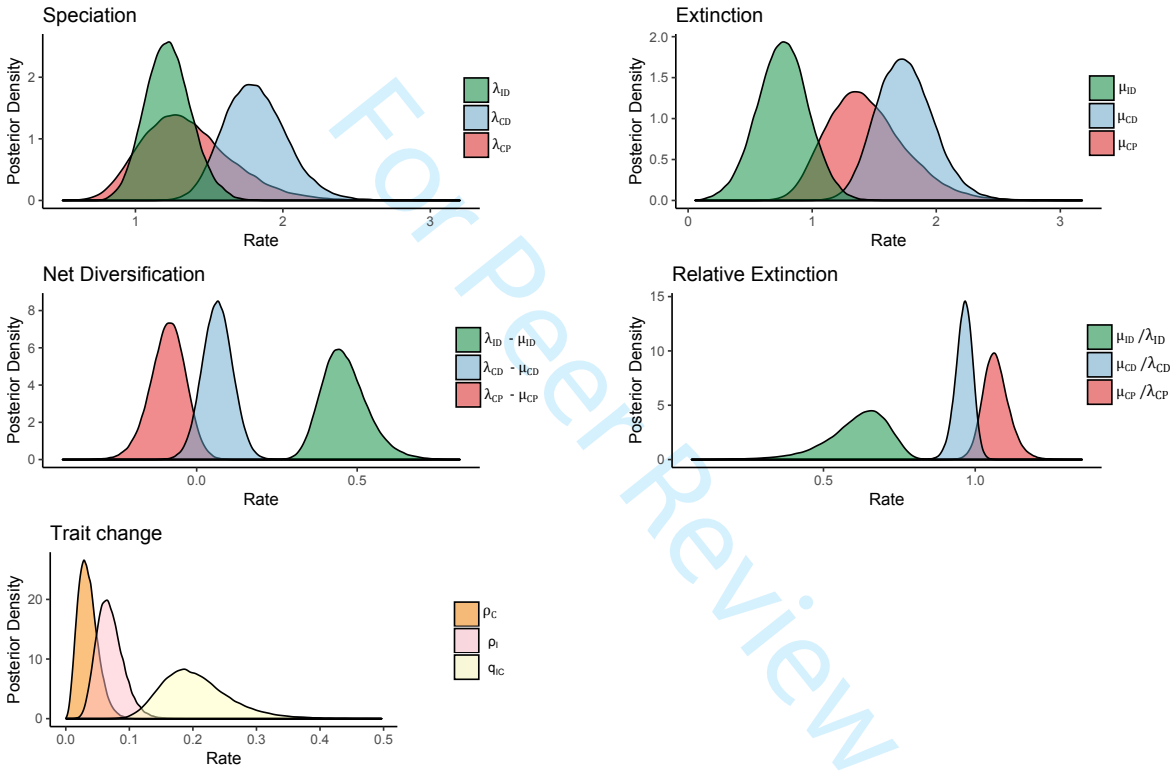

Figure S13: Posterior distribution for each of the parameters in the ID/CD/CP ploidy and breeding system model

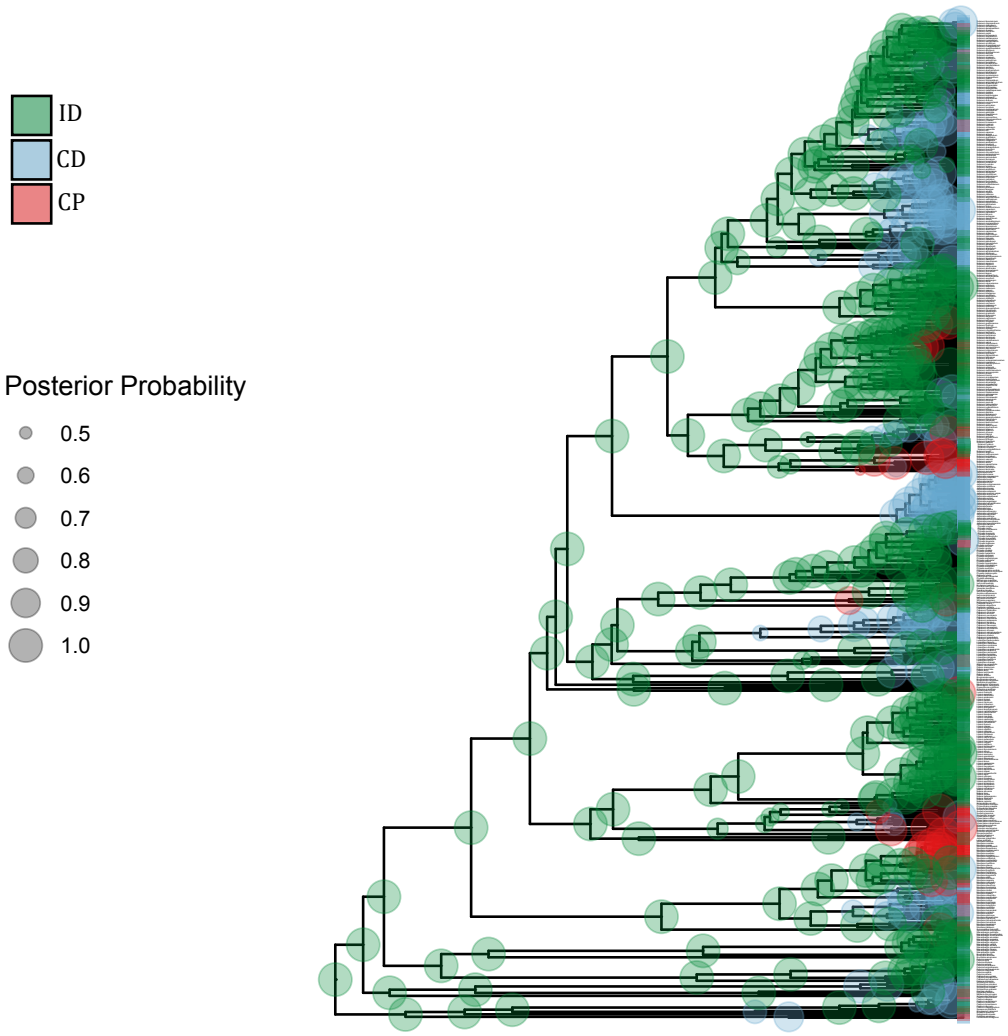

Figure S14: Ancestral state reconstruction showing the maximum a posteriori for each node in the ID/CD/CP ploidy and breeding system model

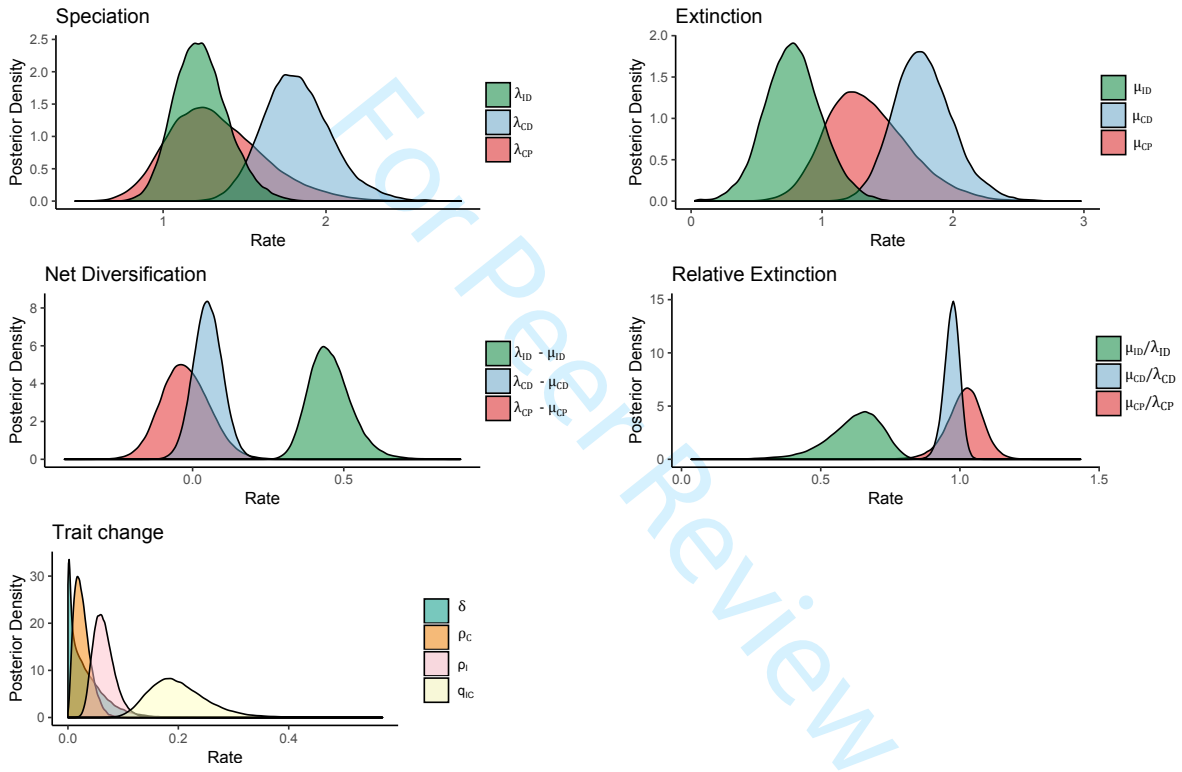

Figure S15: Posterior distribution for each of the parameters in the ID/CD/CP+δ, polyploidy and breeding system model

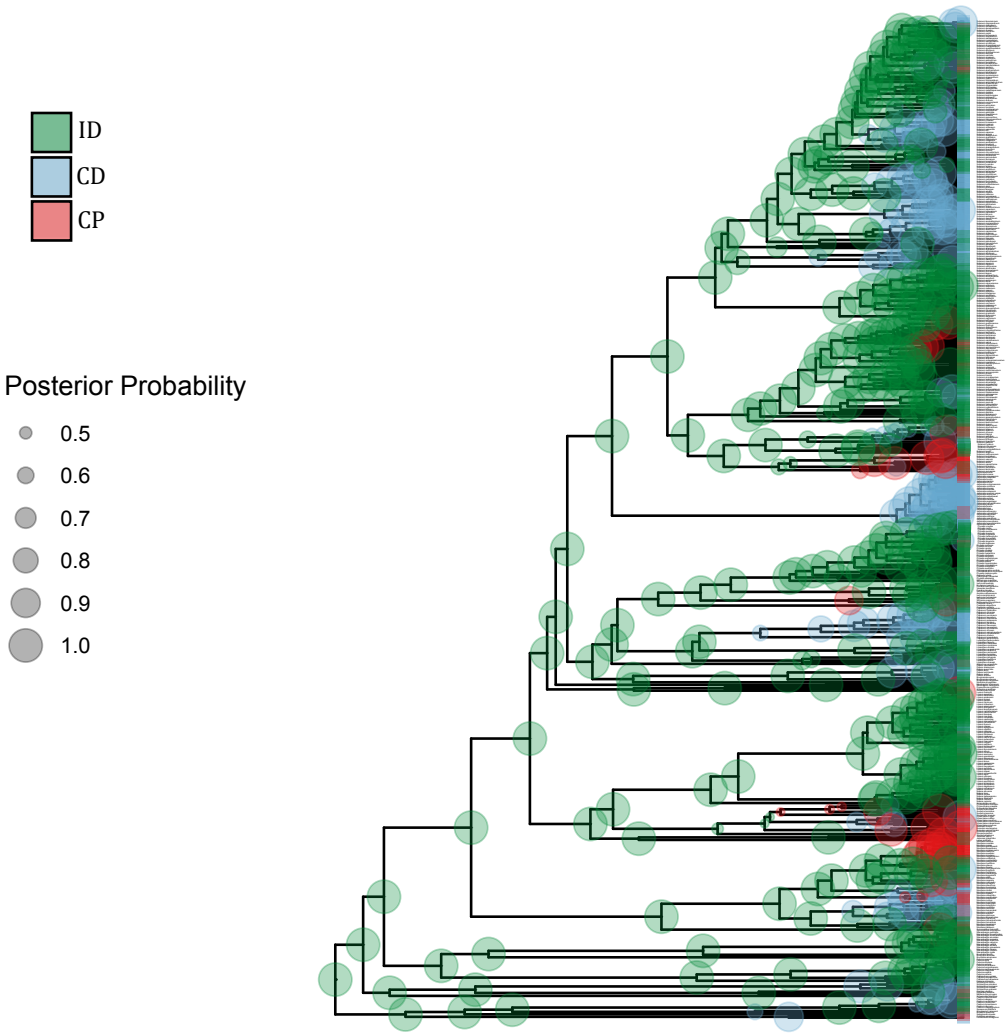

Figure S16: Ancestral state reconstruction showing the maximum a posteriori for each node in the ID/CD/CP+δ ploidy and breeding system model

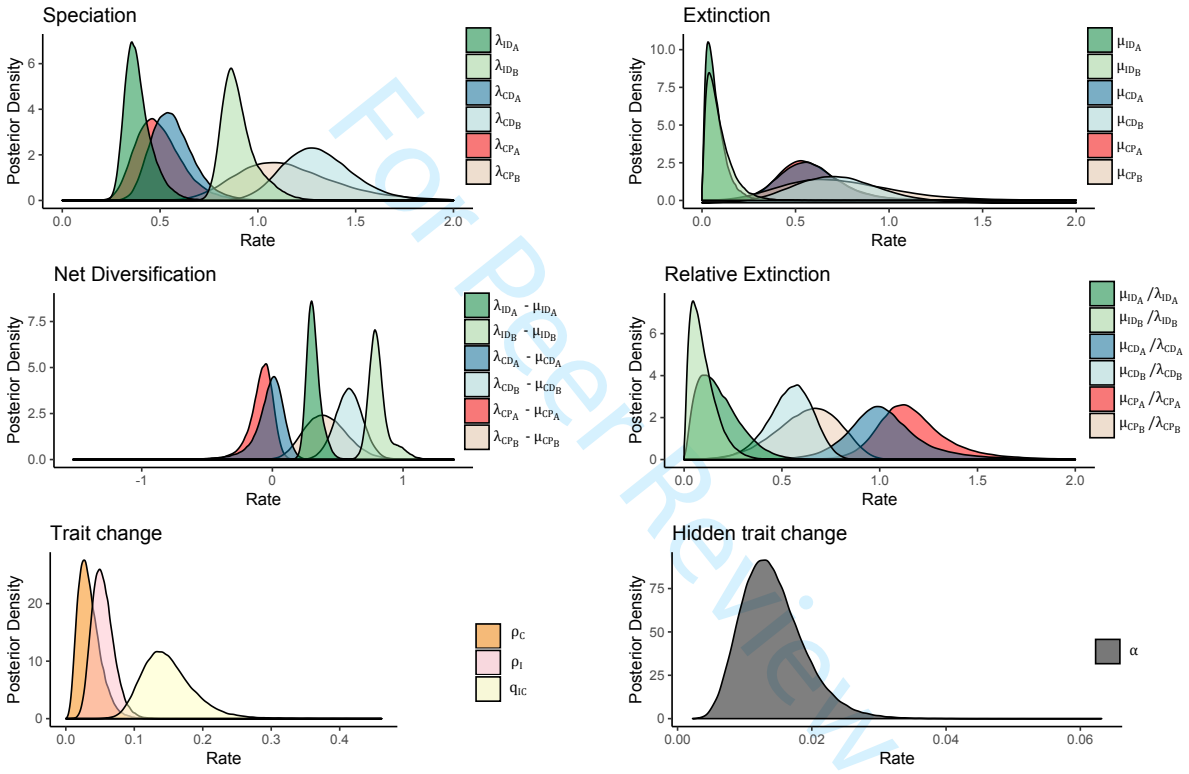

Figure S17: Posterior distribution for each of the parameters in the ID/CD/CP+A/B polyploidy and breeding system model

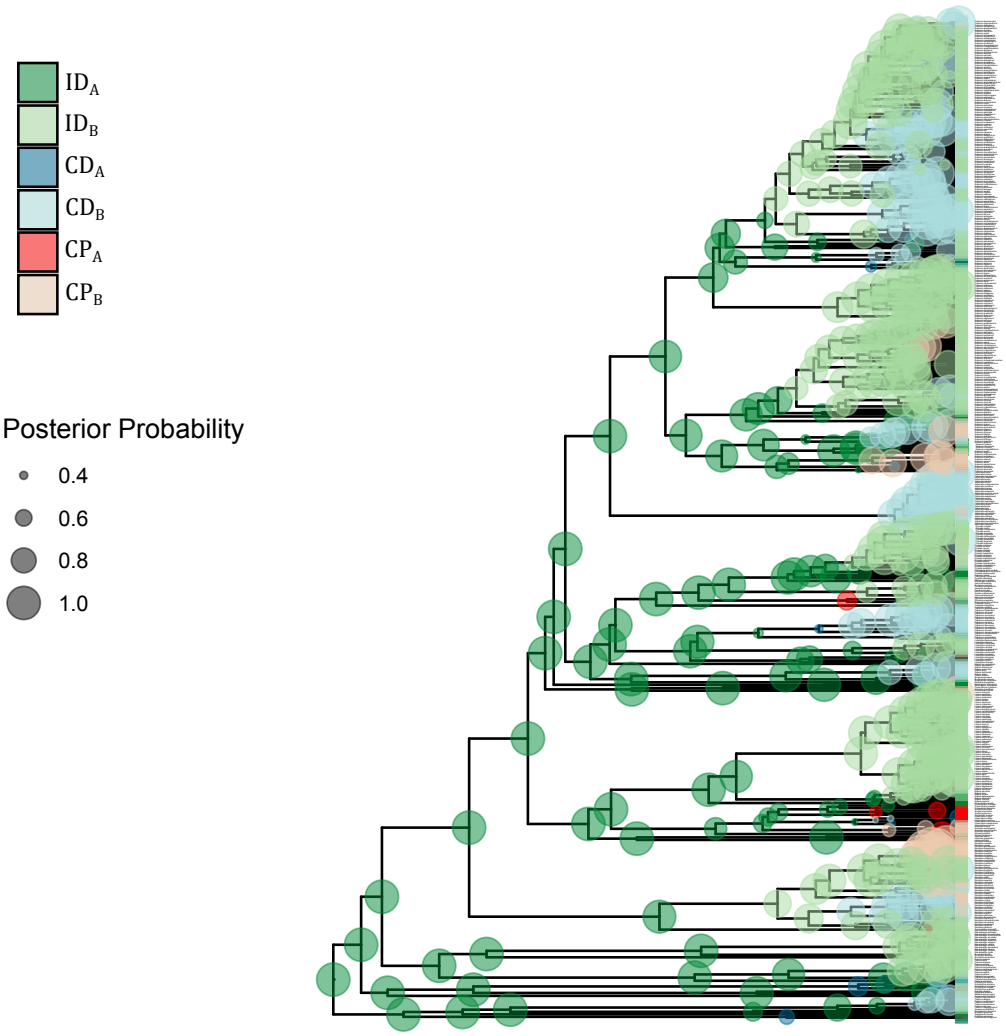

Figure S18: Ancestral state reconstruction showing the maximum a posteriori for each node in the ID/CD/CP+A/B polyploidy and breeding system model

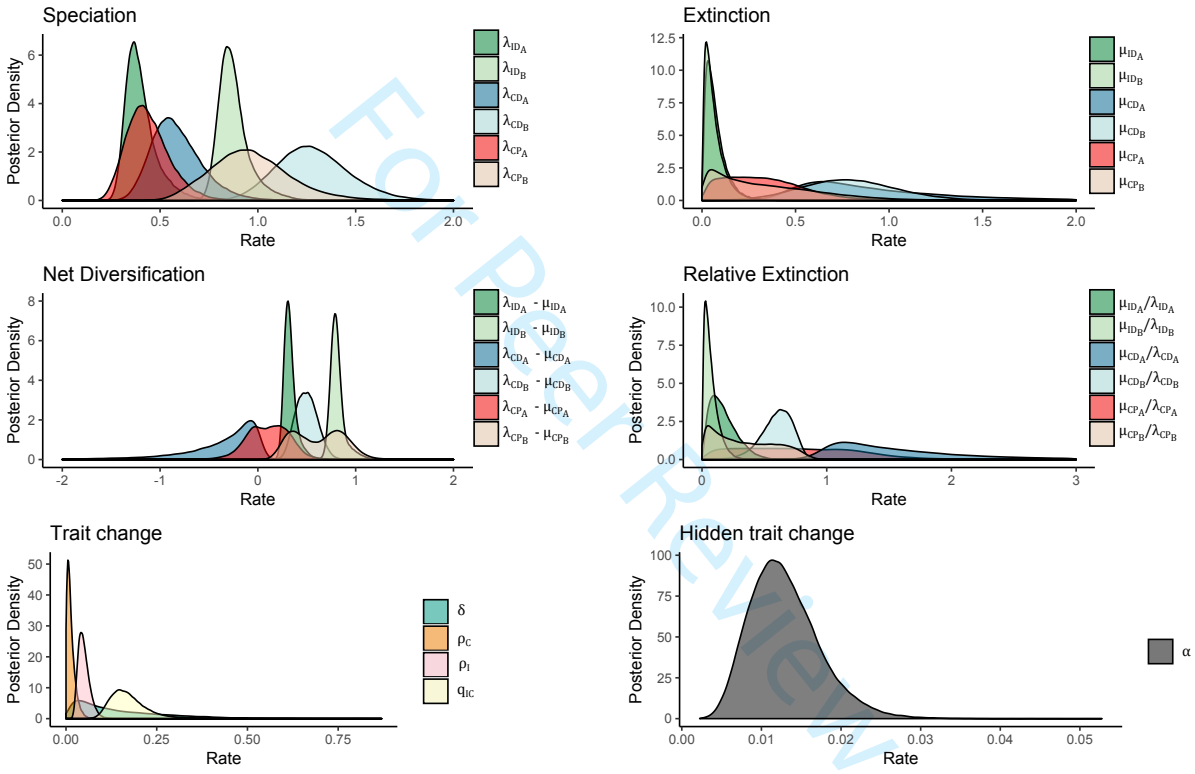

Figure S19: Posterior distribution for each of the parameters in the ID/CD/CP+δ+A/B, polyploidy and breeding system model

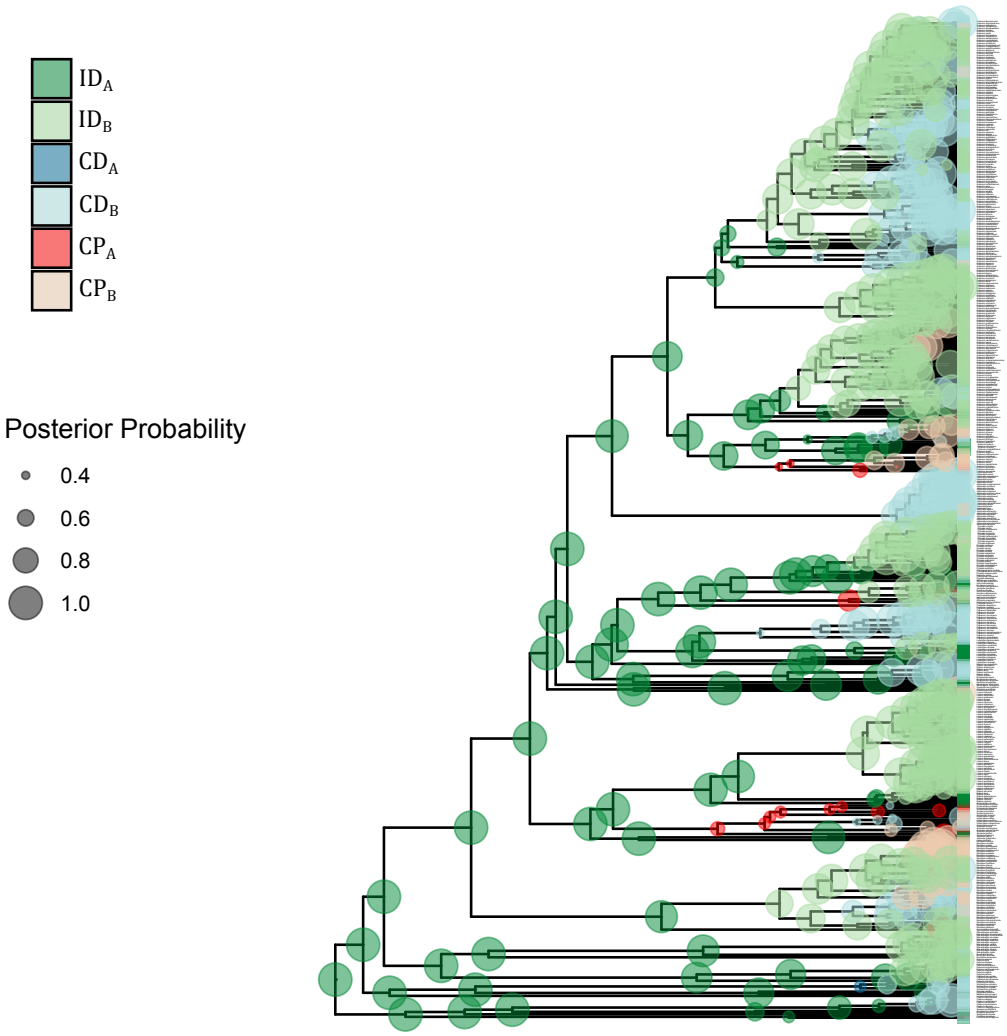

Figure S20: Ancestral state reconstruction showing the maximum a posteriori for each node in the ID/CD/CP+δ+A/B, polyploidy and breeding system model

	D/P	D/P+A/B	I/C	I/C+A/B	ID/CD/CP	ID/CD/CP+A/B
r_D	0.260	0.193, 0.658	—	—	—	—
r_P	-0.056	0.030, 0.187	—	—	—	—
r_I	—	—	0.550	0.386, 0.877	—	—
r_C	—	—	-0.001	-0.059, 0.606	—	—
r_{ID}	—	—	—	—	0.455	0.309, 0.797
r_{CD}	—	—	—	—	0.065	-0.006, 0.587
r_{CP}	—	—	—	—	-0.088	-0.074, 0.403
ρ	0.047	0.047	—	—	—	—
ρ_I	—	—	—	—	0.067	0.053
ρ_C	—	—	—	—	0.033	0.032
q_{IC}	—	—	0.364	0.261	0.198	0.145

	D/P	D/P+A/B	ID/CD/CP	ID/CD/CP+A/B
r_D	0.382	0.698, 0.100	—	—
r_P	0.109	0.587, 0.182	—	—
r_I	—	—	—	—
r_C	—	—	—	—
r_{ID}	—	—	0.449	0.318, 0.789
r_{CD}	—	—	0.050	-0.248, 0.494
r_{CP}	—	—	-0.027	0.110, 0.634
ρ	0.033	0.026	—	—
ρ_I	—	—	0.063	0.047
ρ_C	—	—	0.024	0.011
δ	0.050	0.162	0.022	0.107
q_{IC}	—	—	0.194	0.164

Table S1: Median rate estimates for all fitted models. Units are per million years. Two comma-separated numbers refer to the *A* and *B* hidden states, and — means the parameter was not present in the model. Net diversification rates (r) are subscripted with trait state initials (Diploid, Polyploid, Incompatible, Compatible). Transition rates are ρ (polyploidization), subscripted with background breeding system state; δ (diploidization); and q_{IC} (loss of self-incompatibility). The upper section is for models without diploidization, and the lower section is for models with diploidization. The supplemental figures show the corresponding distributions of parameter estimates.