

**When your favorite trait is not enough to explain diversification: An example of diversification
linked to polyploidy and breeding system in Solanaceae**

Rosana Zenil-Ferguson,¹ J. Gordon Burleigh,² William A. Freyman,³ Boris Igić,⁴ Itay Mayrose,⁵ and Emma
E. Goldberg⁶

¹University of Minnesota Twin Cities

²University of Florida

³23 and me

⁴University of Illinois

⁵Tel Aviv University

⁶University of Minnesota Twin Cities

Author for correspondence: Rosana Zenil-Ferguson

Running head: Polyploidy and breeding systems Solanaceae

Keywords: Polyploidy, Breeding System, Diversification, SSE models

Abstract

The effect of polyploidy in diversification remains a contentious issue. On the one hand, recent studies that found that polyploids have slower speciation rates and higher extinction rates than diploids left scientist wondering if polyploidy is truly an evolutionary dead-end. On the other hand, botanist have found strong
5 molecular support of multiple polyploidy events at the root of highly diverse clades which challenges the evolutionary dead-end conclusions reached by modeling approaches. We re-investigate the role of polyploidy in speciation and extinction from a new modeling perspective considering that patterns found in diversification models can be misleading and incorrectly attributed to polyploidy when other observed and unobserved plant traits are responsible of shaping diversification. Using statistically robust comparative
10 phylogenetic approaches, we show that it is possible to detect whether the contribution of polyploidy to speciation and extinction is significant under the presence of other potential traits also affect diversification. We use the phylogeny, polyploidy, and breeding system data of 595 Solanaceae species to understand the contribution of polyploidy to diversification. We ask if Solanaceae polyploids are evolutionary dead-ends, and whether breeding system or some other unobserved traits are responsible of the patterns of diversification
15 observed in the phylogeny.

Introduction

-Studying diversification linked to trait evolution

Here I need help (perhaps will be the last line I write)

In these context, botanist have been asking how polyploidy shapes patterns of diversification. The prevalence of polyploidy and its detection across multiple and highly diverse clades of angiosperms inevitably lead to the hypothesis of the importance of polyploidy in the speciation and extinction patterns observed across the flowering plant phylogeny. At the same time, a similar question has been asked in the breeding system world, where self-compatibility has evolved multiple times in flowering plants. However, the influence in diversification using both polyploidy and self-compatibility information has not been studied simultaneously.

- Why polyploidy and self-compatibility in particular makes for an interesting study in the context of diversification

In the polyploidy world, an important debate regarding the diversification of angiosperms has been ongoing since the publication of Mayrose et al. (2011). The authors discovered that using the latest diversification model linked to two states diploid and polyploid, the net diversification of polyploids was much slower than the net diversification rate than polyploids. This result was surprising and an sparked a discussion about the long-term evolutionary consequences of polyploidy. Soltis et al. (2014) questioned if polyploidy should be regarded as an evolutionary dead-end since the net diversification of polyploids was negative, despite overwhelming evidence about the incidence of polyploidy especially at the root of highly diverse angiosperm claims (ref here). A year later, diversification models were re-tested and corrected, and still found the same pattern (Mayrose et al. 2015), to the disappointment of plenty of botanists there was no denying of the weak trend of diversification that polyploids left behind. The most recent study by Landis et al. (2018) that used not only the presence of polyploidy in the tips but also the number of whole genome duplications in a taxon lineage found that....

Meanwhile, studies focusing on diversification patterns and breeding system have consistently found that self-incompatible plants often have higher net diversification rates compare to their self-compatible counterparts (Emma and Boris' papers here, what about papers that are not solanaceae?)

- Why other traits need to be considered as well

- What other models and studies have done in the past

- What is lacking from past approaches?

There are two key questions that at the time of the polyploidy debate were difficult to ask. The first is if the models used to measure the diversification of polyploids were correct, and the second question is if the models have potentially included more evidence and potential traits that are not polyploidy or other lines of evidence that could be driving the patterns. At the time, in a different context Beaulieu and O'Meara (2016) were finding an alternative solution to the first question, coming up with a new model that could represent the broad heterogeneity of the diversification process and parse out the signal between the trait of interest and the noise in diversification. Their model, the hidden state speciation and extinction model is a key component to detect whether polyploidy or something else unknown but related to it is driving the speciation and extinction patterns that we see in angiosperms.

-What is our proposal to tackle this problem?

- How this paper is structured

The second question (talk here about breeding system and polyploidy, and how this could be different across clades and this is one of the reasons why we focused on Solanaceae). -Diversification and breeding systems

Goldberg and Igic 2012, different perspectives

=

Methods

Data

Chromosome number data were obtained for all Solanaceae taxa in the the Chromosome Counts Database (CCDB; Rice et al. 2015), and the ca. 14,000 records were cleaned semi-automatically using the CCDBcurator R package (Rivero et al. 2019). This large dataset includes the compilation of Solanaceae ploidy states from Robertson et al. (2011). Species were coded as either diploid (D) or polyploid (P). For the majority of species, ploidy was assigned according to information from the original publications and the Kew Royal Botanic Gardens C-value DNA resource (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. For example, *Solanum betaceum* did not include information about ploidy level but it has

24 chromosomes, so because $x = 12$ is the base chromosome number of the *Solanum* genus (Olmstead and Bohs 2007), we assigned *S. betaceum* as diploid. Species with more than one ploidy level were assigned the smallest and most frequent ploidy level recorded. Breeding system was scored as self-incompatible (I) or self-compatible (C) based on results hand-curated from the literature (as compiled in Igić et al. 2006; Goldberg et al. 2010; Robertson et al. 2011; Goldberg and Igić 2012). Most species could unambiguously be coded as either I or C (Raduski et al. 2012). Following previous work, we coded as I any species with functional I systems, even if C or dioecy was also reported. Dioecious species without functional I were coded as C.

To those existing data sets, we added some additional records for chromosome number and breeding system. The Supplementary Information contains citations for the numerous original sources for all of the data. Synonymy followed Solanaceae Source (PBI *Solanum* Project 2012). Hybrids and cultivars were excluded because ploidy and breeding system are likely to be altered by domestication. As in Robertson et al. (2011), we examined closely 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 individuals that were both SI and polyploid. Because of this empirical observation and the functional explanation for whole genome duplication disabling gametophytic self-incompatibility (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 polyploidy were excluded. Tips without trait data are much less informative for diversification parameters linked to trait values, and including this many more species would have prohibitively slowed our analyses, especially for the most complex models.

Models for ploidy and diversification

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) (Fig. 1). We call this the *D/P* ploidy model. In a Bayesian framework, we obtained poste-

rior probability distributions of speciation rates (λ_D, λ_P), extinction rates (μ_D, μ_P), net diversification rates
 110 ($r_D = \lambda_D - \mu_D, r_P = \lambda_P - \mu_P$), and relative extinction rates ($v_D = \mu_D/\lambda_D, v_P = \mu_P/\lambda_P$) associated with
 each state. This analysis explores the same question as Mayrose et al. (2011, 2015), but our analyses differ
 because we include not only polyploidization (parameter ρ , the transition rate from D to P), but also
 diploidization (parameter δ , the transition rate from P to D).

Our second model assesses the the signal of diversification due to ploidy while also parsing out the
 115 heterogeneity of diversification rates due a possible unobserved trait. BiSSE-like models can suffer from a
 large false discovery rate because they fail to account for diversification rate changes that do not directly
 depend on the trait of interest (Rabosky and Goldberg 2015; Beaulieu and O'Meara 2016). To allow for
 diversification rate differences explained by something other than ploidy, we added a hidden state (HiSSE
 model; Beaulieu and O'Meara 2016). In this model, each of the observed diploid and polyploid states is
 120 subdivided by a binary hidden trait with states A and B . We call this the $D/P - A/B$ ploidy and hidden
 state model. We estimated the posterior probability distributions of speciation rates ($\lambda_{D_A}, \lambda_{D_B}, \lambda_{P_A}, \lambda_{P_B}$),
 extinction rates ($\mu_{D_A}, \mu_{D_B}, \mu_{P_A}, \mu_{P_B}$), net diversification rates ($r_{D_A}, r_{D_B}, r_{P_A}, r_{P_B}$), and relative extinction
 rates ($v_{D_A}, v_{D_B}, v_{P_A}, v_{P_B}$). In this model polyploidization rate ρ and diploidization rate δ are also included,
 and changes between hidden states are symmetrical with rate α .

125 ***Models for breeding system and diversification***

To assess the effects of breeding system in the diversification process, we first fit 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),
 though with an updated phylogeny (Särkinen et al. 2013). We call this BiSSE model the I/C breeding
 system model. To parse out the effect of breeding system in diversification while allowing for the possibility
 130 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.

For these and the subsequent models including breeding system, we allow transitions from I to
 C (at rate q_{IC}) but not the reverse. Within Solanaceae, self-incompatibility in species that possess it is a
 homologous form of gametophytic SI (shared even shared even with other dicot families; Igić and Kohn
 135 2001; Steinbachs and Holsinger 2002). Furthermore, species that are distantly related within the family
 carry several closely-related alleles at the locus that controls the SI response (Ioerger et al. 1990; Igić et al.
 2006). This is very strong evidence that the SI mechanism is ancestral in Solanaceae and has not re-evolved
 within the family.

Models for ploidy, breeding system, and diversification

If both ploidy and breeding system have the potential to influence lineage diversification, we should logically consider their effects jointly. We thus fit a model that includes both traits (MuSSE, FitzJohn 2012). The three states in this model are self-incompatible diploids (ID), self-compatible diploids (CD), and polyploids, which are always self-compatible (CP). We did not include a state for self-incompatible polyploids because they are not present in the data, and because that trait combination is logically expected not to occur (as explained above). We call this the *ID/CD/CP* ploidy and breeding system model. It has ten parameters, six for diversification in each state (λ_{ID} , λ_{CD} , λ_{CP} for speciation, μ_{ID} , μ_{CD} , μ_{CP} for extinction) and four for transitions between states (ρ_I , ρ_C for polyploidization transitions from *ID* and *CD* to *CP*, respectively; δ for diploidization from *CP* to *CD*; 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 $q_{IC} + \rho_I$. Diploidization from *CP* to *ID* is not allowed because it would be a regain of self-incompatibility.

The *ID/CD/CP* model could potentially capture similar dynamics as earlier models, if the hidden state in *D/P* – *A/B* was effectively breeding system, and the hidden state in *I/C* – *A/B* was effectively ploidy. There is also the potential, however, for a hidden factor to be influencing diversification even beyond both of our focal traits, and this could again mislead inferences. We therefore added a hidden trait layer on top of our three-state model (analogous to Caetano et al. 2018; Herrera-Alsina 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 (Herrera-Alsina et al. 2018). 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 fitted a simplified version with 16 parameters. The reduction in parameter space is 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 , δ , q_{IC} as defined for the *ID/CD/CP* model). There are additionally twelve diversification rate parameters (λ_{IDA} , λ_{IDB} , λ_{CDA} , λ_{CDB} , λ_{CPA} , λ_{CPB} , μ_{IDA} , μ_{IDB} , μ_{CDA} , μ_{CDB} , μ_{CPA} , μ_{CPB}).

Diploidization as an exploratory hypothesis

For all the four models that consider ploidy changes we assumed that diploidization could happen. 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). 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. However, the majority of the ploidy classifications were adopted from original studies with alternative sources of information (i.e. 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 also fitted the models without diploidization in order to test whether including an assumption of diploidization is sensitive. As discussed by Servedio et al. (2014) the presence or absence of a hypothesis can have an exploratory goal. In our case the diploidization parameter (or the lack of denoted by no δ) in our models is an opportunity to explore an assumption that might be important but it is not the single definitive process to understand the interactions among polyploidy, breeding system, and diversification.

Statistical Inference under the models

Parameters for each of the ten diversification models described previously were performed using custom code in RevBayes (Höhna et al. 2016) environment. Code and results for all analyses is freely available at <https://github.com/roszenil/solploidy>. For all analyses, a correction for sampling bias is included assuming that Solanaceae family has approximately 3,000 species ($s = 595/3000$) as estimated by the Solanaceae Source project (PBI *Solanum* Project 2012). For all ten models, we assumed that speciation and extinction parameters for all models had log-normal prior distributions with means equal the expected net diversification rate ($\frac{(\text{number of taxa})/2}{\text{root age}}$) and standard deviation 0.5. Prior distributions for parameters defining trait changes were assumed to be gamma distributed with parameters $k = 0.5$ and $\theta = 1$. For each model, an MCMC was performed for 96 hours in the high-performance computational cluster at Minnesota Super-

computing Institute which allowed for 5,000 generation of burn-in and a minimum of 200,000 generations of MCMC for each of the 6 models. For each model, convergence and mixing of the MCMC was tested using R package coda and software Tracer (see supplementary information for convergence plots).

Model selection

We calculated the marginal likelihoods for each of the ten models using custom code in RevBayes Höhna et al. (2016). Marginal likelihoods were calculated using 50 stepping stone steps using 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 1000 MCMC steps Table 1. The calculation of each marginal likelihood ran for 24 hours in the high-performance computational cluster at Minnesota Supercomputing Institute.

Using the marginal likelihood values we calculated fourteen different Bayes factors in log-scale.

The first six Bayes factors compared the four models of polyploidy against each other, followed by one single comparison between the breeding system models, and finalizing with four comparisons between each polyploidy and breeding system models (see Table 2 for all comparisons). Comparison between polyploidy models vs. polyploidy and breeding system models are not valid for two reasons: first, the input data in each model is numerically different (see ?? for data classifications) ; second, even when it is possible to re-write the polyploidy and diversification models to include a hidden state that would account for breeding system allowing to use the same input data, the re-written model would still not be comparable to the multivariate SSE models proposed for polyploidy and breeding system due to the hierarchical structure imposed by the interaction between breeding system and polyploidy. This second point has been recently denoted by Tarasov (2018) as the lack of lumpability in Markov models for discrete traits. In our particular case, polyploids cannot be self compatible, and the irreversibility of self-compatible to self-incompatible state makes the three-state Markov models non-lumpable into two-state models ultimately making multivariate and the bivariate models incomparable.

Results

Polyploidy and Diversification Models

Similarly to the results obtained by Mayrose et al. (2011) and Mayrose et al. (2015), we found that in the D/P polyploidy model the net diversification of diploids is larger than the the net diversification of polyploids since the net diversification distributions do not overlap (Figure 2(A)). This result holds true whether or not the diploidization parameter is present. However in the presence of the diploidization parameter the net diversification rate of polyploids is nonnegative with probability 1 (Figure 2(A)), whereas in the absence of diploidization the net diversification rate of polyploids can be negative with a probability (HERE verify the quantile) (FIGURE 3(A)). In terms of the relative extinction, when the diploidization parameter is present both polyploids and diploids have posterior distributions that overlap, but that pattern changes in the absence of the diploidization parameter leading to a significant difference between relative extinction where polyploids have a significant higher relative extinction rate (see Supplementary Information).

For the D/P- A/B model with diploidization the diploid and polyploid net diversification rates are overlapping for both state A and B of the hidden trait (Figure 2(B)). In this model, the differences in net diversification are due to the presence of a hidden trait and not to the differences in ploidy. When diploidization parameter is absent the hidden state is still driving the differences in diversification rates (Figure 3(B)).

Breeding System and Diversification models

In the I/C breeding system model we found that the net diversification rate for self-incompatible state is larger than the net diversification rate for self-compatible state (Figure 2(C)). The net diversification rate for self-compatible has a probability distribution centered at zero.

When a hidden state is added in the I/C-A/B model, we found that under the hidden state A the self-compatible and self-incompatible net diversification rates are different. Under the hidden state B, those two rates overlap with a probability (HERE CALCULATE THAT) meaning that they are different with probability (CALCULATE) (Figure 2(D)). These results agree with previous results found by Goldberg and Igić (2012). However, Goldberg and Igić (2012) used a ClaSSE approach since they were interested in anagenetic and cladogenetic changes for self-incompatibility using a smaller subset of the data presented in the current work.

Polyploidy and Breeding Sytem models

In the ID/P/CD model we found that self-incompatible and diploid state has a significantly larger net diversification rate compare to both self-compatible diploid and polyploid rates. Meanwhile, both self-compatible diploid and polyploid posterior distributions of net diversification rates completely overlap (Figure 2(E)). When hidden state was added in the ID/P/CD-A/B model, we observed significant differences between self-compatible and self-incompatible diploids for both A and B values of the hidden state. Self-incompatible state had a larger net diversification rate than self compatible for both A and B states. However, the posterior distribution for the net diversification rate of polyploids overlaps with both the self-compatible and self-incompatible posterior distributions for each value of the hidden state(Figure 2F) meaning that polyploidy state is not significantly different from diploid in net diversification terms. The resulting effect of adding the hidden state values is significant

Diploidization as an exploratory hypothesis

In the D/P model the diploidization rate δ and polyploidization rate ρ are different from zero with probability 1. Diploidization rate is more uncertain than polyploidization. For the D/P no δ model, the polyploidization rate is still different from zero but has a wider 95% credible interval (see Supplementary Information). In the D/P-A/B model the diploidization rate remains positive but there rate of polyploidization becomes really uncertain. Whereas, in the absence of diploidzation (D/P no δ A/B) the rate of polyploidy has a small credible interval

For the model ID/P/CD containing both polyploidy and breeding system traits, we found that the diploidization rate is really uncertain with a MAP that is (CALCULATE VALUE HERE- close to zero). The
265 polyploidization rate from self-incompatible diploid ρ_I is slightly faster than the polyploidization rate from self-compatible diploid ρ_C and that pattern remains the same for the ID/P no δ /CD model(see supplementary information).

Model selection

In Table 1 we list the marginal likelihood in log scale for each of the models tested. In the table we show
270 what are the different components included and excluded for each model as a summary of the diagrams from Fig. 2. From the marginal likelihoods in log scale, the Bayes factors in log-scale were calculated as shown in table Table 2. After testing every single pair of polyploidy modes (1-4) we found overwhelming evidence that the best polyploidy model is always the D/P-A/B, that is the model with hidden state and diploidization.

275 For the two models following the evolution of breeding system, the I/C-A/B is the best choice between models 3 and 4.

The models that follow the diversification linked to both polyploidy and breeding system are the last 4 (models 5-8). When comparing using Bayes factors every two models we found that the IC/P/CD-A/B is always preferred over the rest, meaning that the model that has a hidden state and diploidization is chosen
280 over the ones that lack either of both of those options.

Therefore, the models that were chosen were always the ones containing a hidden trait, and in the case of polyploidy models, the ones containing a diploidization parameter δ are preferable.

Discussion

Acknowledgements

Literature Cited

285

- Beaulieu, J. M. and B. C. O'Meara, 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.
- 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.
- 290 FitzJohn, R. G., 2012. Diversitree : comparative phylogenetic analyses of diversification in r. *Methods Ecol Evol* 3:1084–1092.
- Goldberg, E. E. and B. Igić, 2012. Tempo and mode in plant breeding system evolution. *Evolution* 66:3701–3709.
- 295 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.
- 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* .
- 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.
- 300 Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist, 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol* 65:726–736.
- Igić, B., L. Bohs, and J. R. Kohn, 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proc Natl Acad Sci USA* 103:1359–1363.
- 305 Igić, B. and J. R. Kohn, 2001. Evolutionary relationships among self-incompatibility RNases. *Proceedings of the National Academy of Sciences USA* 98:13167–13171.
- 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.
- 310 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. *Am J Bot* 105:348–363.
- Maddison, W. P., P. E. Midford, and S. P. Otto, 2007. Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56:701–710.
- 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.
- 315 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 Phytol* 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.
- 320 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. and E. E. Goldberg, 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology* 64:340–355.
- 325 Raduski, A. R., E. B. Haney, and B. Igić, 2012. The expression of self-incompatibility in angiosperms is bimodal. *Evolution* 66:1275–1283.
- 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.
- 330 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 chromo-

some numbers. *New Phytol* 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.

335 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
340 D. J. Yeh, 2014. Not just a theory? the utility of mathematical models in evolutionary biology. *PLoS biology* 12:e1002017.

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 Phytol* 202:1105–1117.

345 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.

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

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.

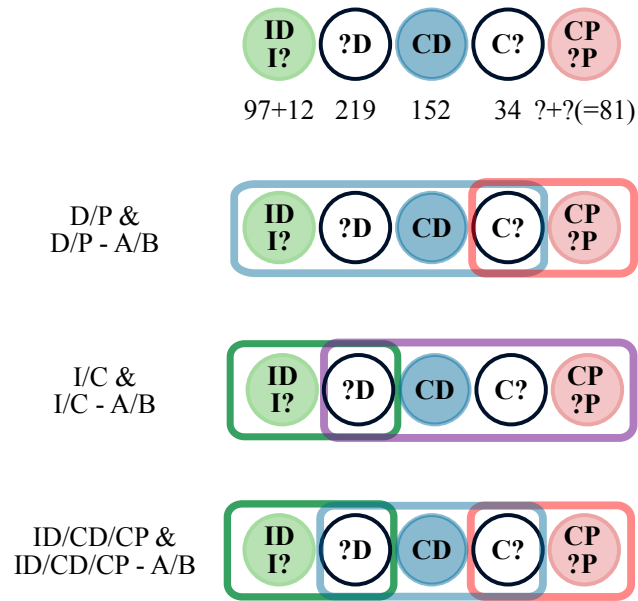


Figure 1: Binary and three state classifications for 595 taxa with ploidy and/or breeding system data. The number of taxa in the sample was maximized by including tips with only ploidy or only breeding system and assigned them as uncertain in some of the models. For example, the 34 self-compatible taxa with unknown ploidy were assigned as uncertain (diploid or polyploid) in the D/P models but as fully compatible in the I/C models.

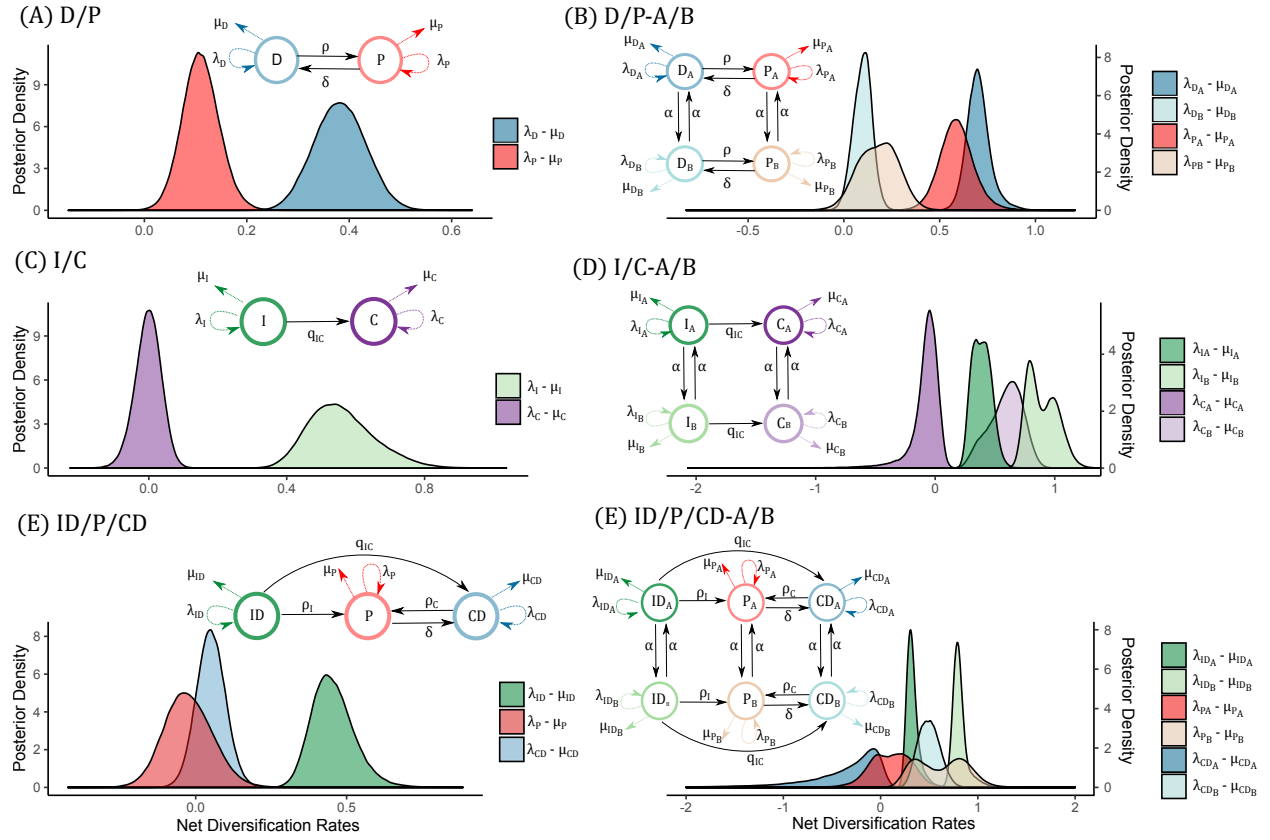


Figure 2: Net diversification rates for all models that include diploidization.

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

Table 1: Marginal likelihoods for the ten diversification models proposed.

Polyploidy Models					Breeding System Models		Polyploidy and Breeding System Models					
	1	2	3	4		5	6		7	8	9	10
1. D/P	·	10.72	-37.24	-31.94	5. I/C	·	-39.43	7. ID/P/CD	·	1.36	-44.15	-40.95
2. D/P no δ	·	·	-47.97	-42.66	6. I/C-A/B	·	·	8. ID/P/CD no δ	·	·	-45.51	-42.31
3. D/P- A/B	·	·	·	5.30				9. ID/P/CD-A/B	·	·	·	3.2
4. D/P no δ -A/B	·	·	·	·				10. ID/P/CD no δ -A/B	·	·	·	·

Table 2: Bayes factors in log scale. We compare every possible pair. Number models as indicated in Table 1.