Interacting associations 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[6]

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

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

[3]23andMe, Inc., 899 W Evelyn Ave, Mountain View, CA 94041, U.S.A.

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

[5]Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv, Israel

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

†Author for correspondence.

Running head: Ploidy and breeding systems in 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 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 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 observed in the phylogeny.

Introduction

Species accumulate around the tree of life at different rates. The search for traits that explain these differences has been accelerated by dramatic increases in phylogenetic data and, despite some setbacks (**??**), advances in analytical methods (Maddison et al. 2007; **?**; Goldberg and Igić 2012; Beaulieu and O'Meara 2016). In these studies, it is common to identify a single focal trait and investigate its association with rates of speciation and extinction.

This is problematic because the context in which traits occur can lead to complex interactions, perhaps with other traits that make up an organism or its environment. Additionally, a lack of methods that allow for simultaneous inference of effects of other traits, which also influence diversification, can lead to incorrect conclusions about the focal trait (**?**Beaulieu and O'Meara 2016). One approach to this problem is to allow for a 'hidden' or unspecified trait in the analysis (Beaulieu and O'Meara 2016). This can be effective, but it leaves behind the question of whether the hidden trait corresponds to a real trait that should be sought, or whether it is an approximation of some unknowable heterogeneity, as well as whether interactions between the known and hidden trait were modeled appropriately. Alternatively, we can simultaneously consider the effects of more than one known trait, especially in systems where multiple traits are suspected to influence diversification, and where interactions between those traits are well understood. In this paper, we focus on two of the best studied traits in flowering plants, ploidy and breeding system, and investigate their individual and joint effects on diversification.

Polyploidization multiplies the genomic content of cells, and it consequently has the potential to affect many or all traits and a great variety of evolutionary processes. It is also a mutation that occurs commonly in plants, at it is widespread at both population and lineage scales. The prevalence of variation in chromosome number, and especially ploidy, has been broadly considered a salient feature of flowering plants for nearly a century (**?**). The recent dramatic increase in the scale of available genome sequences uncovered ancient rounds of whole-genome duplications, and subsequent diploidization, across angiosperms (**??**). Nearly all lineages of flowering plants are thought to have undergone at least one or two rounds of polyploidization. The prevalence of polyploidy, its variation across clades, and its large effects on genotypes and phenotypes raises the hypothesis that ploidy plays an important role in shaping rates of speciation and extinction. This has led to a debate: Mayrose et al. (2011, 2015) found a lower rate of lineage diversification for polyploids, while Soltis et al. (2014) argued that polyploidy is not an evolutionary dead end because whole genome duplications occurred at the base of many successful angiosperm clades, and Landis et al. (2018) reported higher diversification rates following some whole genome duplications. Our study

revisits this question, additionally incorporating the process of diploidization. Furthermore, we consider the interaction between polyploidy and other traits.

From the beginning, study of this variable cytogenetic property considered correlations with other traits (**?**). Some of these associations are driven by indirect or ecological effects, such as the tendency for polyploidy, and other traits, to be found at higher latitudes or marginal habitats. Other correlations may have a more direct, causal connection. Among the many changes associated with polyploidization, perhaps the most prominent is the association between polyploidy and propensity for self-fertilization (**?**). In some cases, the evidence for a correlated shift in mating system along with polyploidization appears limited and sometimes contradictory (**???**). In other cases, however, polyploidy is not only a suspected correlate of breeding systems but indeed a causal link (**??**). Doubled number of alleles in pollen is thought to effect disruption of the genetic mechanisms in gametophytic self-incompatibility systems, which prevent self-fertilization (**???**). This creates a correlation between polyploidy and self-compatibility by precluding the existence of self-incompatible polyploids. In clades with these systems, it is thus natural to consider the simultanous macroevolution of ploidy and breeding system.

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 affect the distribution and amount of genetic variation in populations (**??**). In particular, self-incompatibility (SI) systems cause a plant to reject its own pollen, and their loss, yielding self-compatibility (SC), is one of the most replicated transitions in flowering plant evolution. Previous phylogenetic analyses have reported higher rates of diversification for SI than for SC lineages (Goldberg et al. 2010; **?**), but they have not considered the possibility of other correlated traits driving this pattern. Given that changes in ploidy and breeding systems may be causally related and have profound affects on the fate of lineages, it seems particularly profitable to examine possible interactions in their macroevolutionary effects. This includes their joint influence on lineage diversification, and also potential patterns in the order of their transitions. For example, do losses of SI more commonly occur tied to polyploidization, or without a ploidy shift? Do polyploids arise more commonly from SI or SC diploids? Robertson et al. (2011) found that the pathway from SI diploids to SC polyploids is dominated by loss of SI followed later by polyploidization over long timescales, but proceeds in one step via polyploidization of SI species over short timescales. We revisit this question with a greatly improved phylogeny and methods that allow for diversification rate differences.

Here, we use extensive data on ploidy and breeding system in Solanaceae to investigate the associations of these two traits with lineage diversification. Considering each trait separately indicates that each is connected to diversification differences. Considering them jointly, however, reveals that the ploidy con-

nection is removed by incorporating breeding system. We further show that the general results are robust to allowing for diploidization and a hidden trait, and something about pathways. Our results emphasize the importance of considering traits not only in isolation, especially when there are strong correlations between them.

Methods

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 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 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). 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 sources for the added data. 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 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 a convincing case of a naturally occurring SI and polyploid population. The single instance of an SI and polyploid indi-

vidual appears to be an allopentaploid hybrid of *Solanum oplocense* Hawkes x *Solanum gourlayii* Hawkes, reported by **?**. 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 polyploidy were excluded. Tips without trait data are much less informative for diversification parameters linked to trait values. Including this many more species would have prohibitively slowed our analyses, especially those implementing 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 posterior probability distributions of speciation rates (λ_D, λ_P), extinction rates (μ_D, μ_P), net diversification rates ($r_D = \lambda_D - \mu_D$, $r_P = \lambda_P - \mu_P$), and relative extinction rates ($\nu_D = \mu_D/\lambda_D$, $\nu_D = \mu_D/\lambda_D$) 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 signal of diversification due to ploidy differences while also parsing out the 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 (**?**Beaulieu and O'Meara 2016). Diversification rate differences explained by something (trait) other than ploidy, are accommodated by adding a hidden state (HiSSE model; Beaulieu and O'Meara 2016). In this model, each of the observed diploid and polyploid states is

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 (λ_{D_A}, λ_{D_B}, λ_{P_A}, λ_{P_B}), extinction rates (μ_{D_A}, μ_{D_B}, μ_{P_A}, μ_{P_B}), net diversification rates (r_{D_A}, r_{D_B}, r_{P_A}, r_{P_B}), and relative extinction rates (ν_{D_A}, ν_{D_B}, ν_{P_A}, ν_{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.

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), save for 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 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*.

For all breeding system models, we allow transitions from *I* to *C* (at rate q_{IC}) but not the reverse. Within Solanaceae, self-incompatibility is homologous in all species in which S-alleles were cloned, and controlled crosses performed. All species sampled to date, possess a non-self recognition, RNase-based, gametophytic self-incompatibility (shared even shared even with other euasterid families; **?**). Furthermore, species that are distantly related within this family carry closely-related alleles, with deep trans-specific polymorphism, at the S-locus, which controls the SI response (**?**Igić et al. 2006). This represents very strong evidence that the SI mechanism, and our *I* state is ancestral to the Solanaceae, and did not arise independently within the family ($q_{CI} = 0$).

Models for ploidy, breeding system, and diversification

If ploidy and breeding system each influence lineage diversification individually, it is logical to examine their possible joint effects. We thus fit a multi-state 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). As explained above, 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. The model has 10 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 represent a simultaneous regain of SI.

The ID/CD/CP model could potentially capture similar dynamics as earlier models, if the effects of the hidden state in D/P–A/B were effectively caused by breeding system (or its correlates), and the hidden state in I/C–A/B was effectively caused by ploidy. There is also the potential, however, for a hidden factor to be influencing diversification 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; **?**). 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_{ID_A}, λ_{ID_B}, λ_{CD_A}, λ_{CD_B}, λ_{CP_A}, λ_{CP_B}, μ_{ID_A}, μ_{ID_B}, μ_{CD_A}, μ_{CD_B}, μ_{CP_A}, μ_{CP_B}).

Diploidization as an exploratory hypothesis

For all four models that consider ploidy changes, we allowed diploidization. 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 (**?**). 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 (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 also fit the models 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 hypothesis can have an exploratory goal. In our case the diploidization parameter (or its absence, $\delta = 0$) in our models is an opportunity to explore an assumption that might be important but that 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 10 diversification models were performed using custom code in the RevBayes (Höhna et al. 2016) environment. Code for analyses and key results is available at `https://github.com/roszenil/solploidy`. We included a correction for incomplete sampling in all analyses, based on assuming that the Solanaceae family has approximately 3,000 species ($s = 595/3000$) as estimated by the Solanaceae Source project (PBI *Solanum* Project 2012). For all 10 models, we assumed that speciation and extinction parameters had log-normal prior distributions with means equal to the expected net diversification rate $(\text{number of taxa}/[2 \times \text{root age}])$ and standard deviation 0.5. Priors for parameters defining trait changes were assumed to be gamma distributed with parameters $k = 0.5$ and $\theta = 1$. For each model, an MCMC chain 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 6 models. For each model, convergence and mixing of the MCMC was tested using the R library `coda` and the software package Tracer (see supplementary information for convergence plots).

Model selection

We calculated the marginal likelihood for each of the 10 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 polyploidy against each 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 essence, the D/P, I/C, and ID/CD/CP state spaces are not lumpable with respect to one another (Tarasov 2018).

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 0.87 (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 diploididization 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)).

Polyploidy and Breeding System models

When breeding system is considered simultaneously from polyploidy, we found that in the ID/CD/CP model self-incompatible and diploid state (ID) has significantly larger net diversification rate compare to both self-compatible diploid (SD) and polyploid rates(CP). Meanwhile, both self-compatible diploid and polyploid net diversification rates have overlapping posterior distributions (Figure 2(E)). When hidden state was added, the ID/P/CD+A/B model, significant differences between self-compatible and self-incompatible diploids continue to be the truth for both A and B values of the hidden state. 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). This result indicated that polyploid state was not significantly different from diploid from the net diversification perspective. In the absence of diploidization, models ID/CD/CP no $/delta$ and ID/P/CD no $delta$ +A/B showed that the net diversification of polyploids is equivalent to the net diversification of self-compatible diploids. In the absence of diploidization, the net diversification rates of self-incompatible diploids are always faster than their self-compatible diploid and polyploid counterparts (Figures 3C and 3D).

Polyploidization rates from self-compatible diploid ρ_{CD} and from self-incompatible diploid ρ_{ID} are not significantly different (see supplementary information figures). However, the distribution of ρ_{ID} seems to always consider faster values compare to the distribution of ρ_{CD}. This result is highlighted in ancestral re-

constructions of the $ID/CD/CP$ model where transitions from ID to CP are more probable than transitions from CD to CP.

Diploidization as an exploratory hypothesis

For the two models that only consider diploid and polyploid state (D/P and D/P+A/B) the effect of removing parameter δ was that net diversification rate of polyploids can be negative with high probability (Figures 3A and 3B) . However, in the models that simultaneously considered polyploidy and breeding system, removing the δ parameter did not have much of an effect on the location of the posterior distribution of polyploid net diversification rate but on the credible interval width and overlap with net diversification of diploids. When diploidization was absent, the net diversification of polyploids overlaps with the net diversification of self-compatible diploids but not with the one of self-incompatible diploids (Figures 3C and 3D). For all models, the posterior distribution of diploidization rate is more uncertain than the one of polyploidization rates (See supplementary information figures).

Modeling breeding system by itself

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. Furthermore, the net diversification of self-compatible taxa on hidden state B overlaps with the hidden state rate of self-incompatible taxa on state A (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. Transition rate between incompatible and compatible state are frequent with a rate $q_{IC} = 0.3$ for both the I/C and the I/C+A/B models.

Model selection

In Table 1 we listed the marginal likelihood values calculated in log scale ($log(P(\mathbf{X}|M))$ for each of the models tested. The different components included and excluded for each model as a a summary of the diagrams from Fig. 2 are indicated in each model. From the marginal likelihoods in log scale, the Bayes factors in log-scale($BF(M_0,M_1) = log(P(\mathbf{X}|M_0) - P(\mathbf{X}|M_0))$ were calculated as shown in table Table 2. After testing every single pair of polyploidy models (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, that is Bayes factors were always greater than 2 when D/P+A/B was the first model in the Bayes factor equation or negative when D/P+A/B was the second model in Bayes factor equation.

For the two models following the evolution of breeding system, the I/C+A/B is the best choice between models 3 and 4 (BF(IC, IC+A/B)=-39.5<0).

The four models that follow the diversification linked to both polyploidy and breeding system (models 5-8) where compared in pairs. Bayes factors show decisive evidence in favor of the IC/P/CD+A/B over the rest, meaning that the model that has a hidden state and diploidization is chosen over the ones that lack either of both of those options Table 2 .

Therefore, the models that were chosen were always the ones containing the hidden trait, or diploidization parameter δ when ploidy state was considered.

Discussion

Acknowledgements

NSF DEB-1655478. The Minnesota Supercomputing Institute (MSI) at the University of Minnesota provided computing resources for this project.

Literature Cited

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.

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.

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 .

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.

Igić, B. and J. W. Busch, 2013. Is self-fertilization an evolutionary dead end? New Phytol 198:386–397.

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., 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.

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.

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.

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 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.

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.

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 mayroseetăal . (2011). New Phytol 202:1105–1117.

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 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.

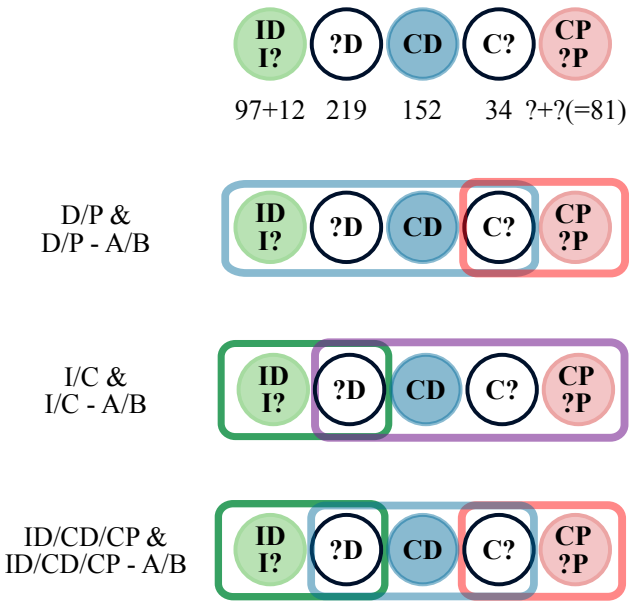

Figure 1: Character states used for each of the models. Each species retained on the tree belonged to one of five possible categories, depending on whether ploidy and/or breeding system were known. The number of species in each is shown under the corresponding circles in the top row. These categories were then grouped in a manner appropriate to the states of each model. For example, there are 34 species that are self-compatible and of unknown ploidy; these are coded as either *D* or *P* in the D/P models (uncertain, or consistent with either state), as *C* in the I/C models, and as either *CD* or *CP* in the ID/CD/CP models. In all cases, species were coded as either *A* or *B* in the hidden state models.

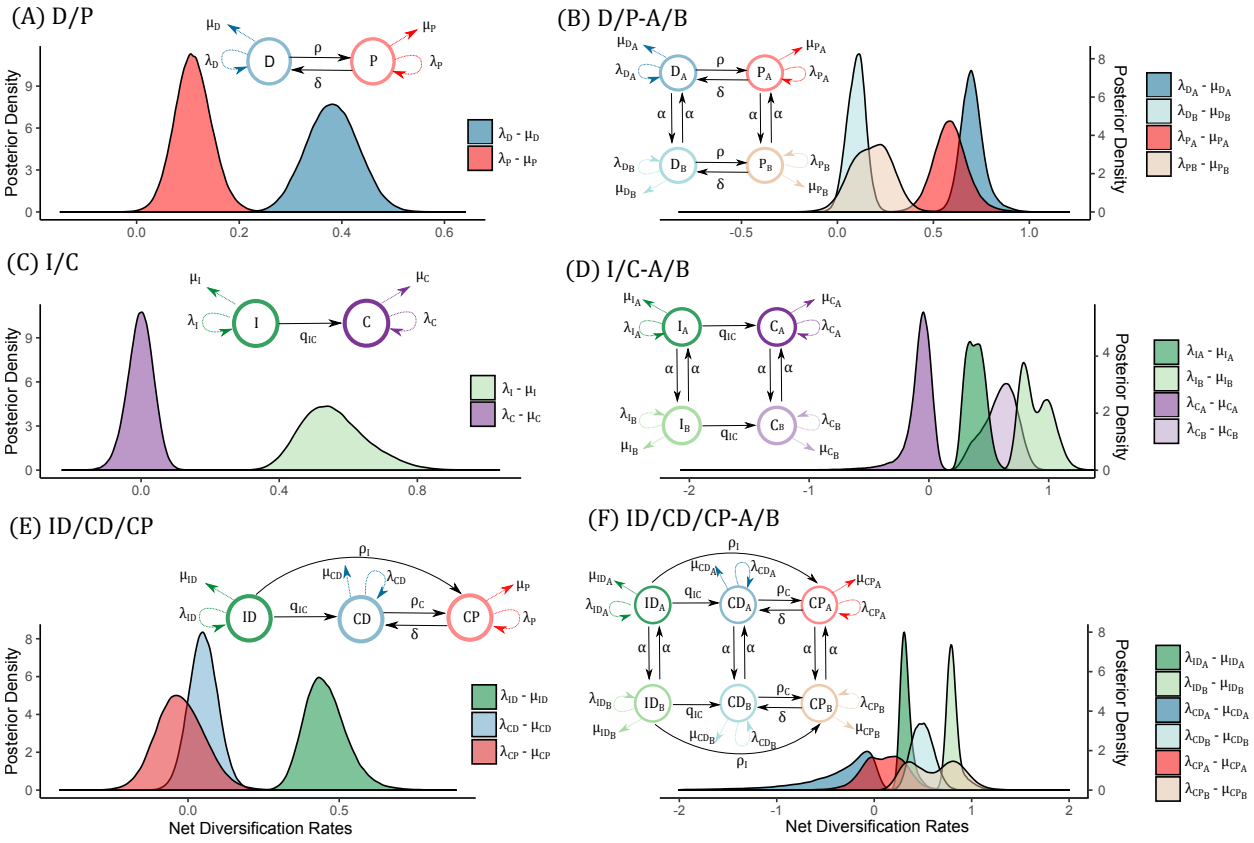

Figure 2: Net diversification rates for all models that include diploidization.

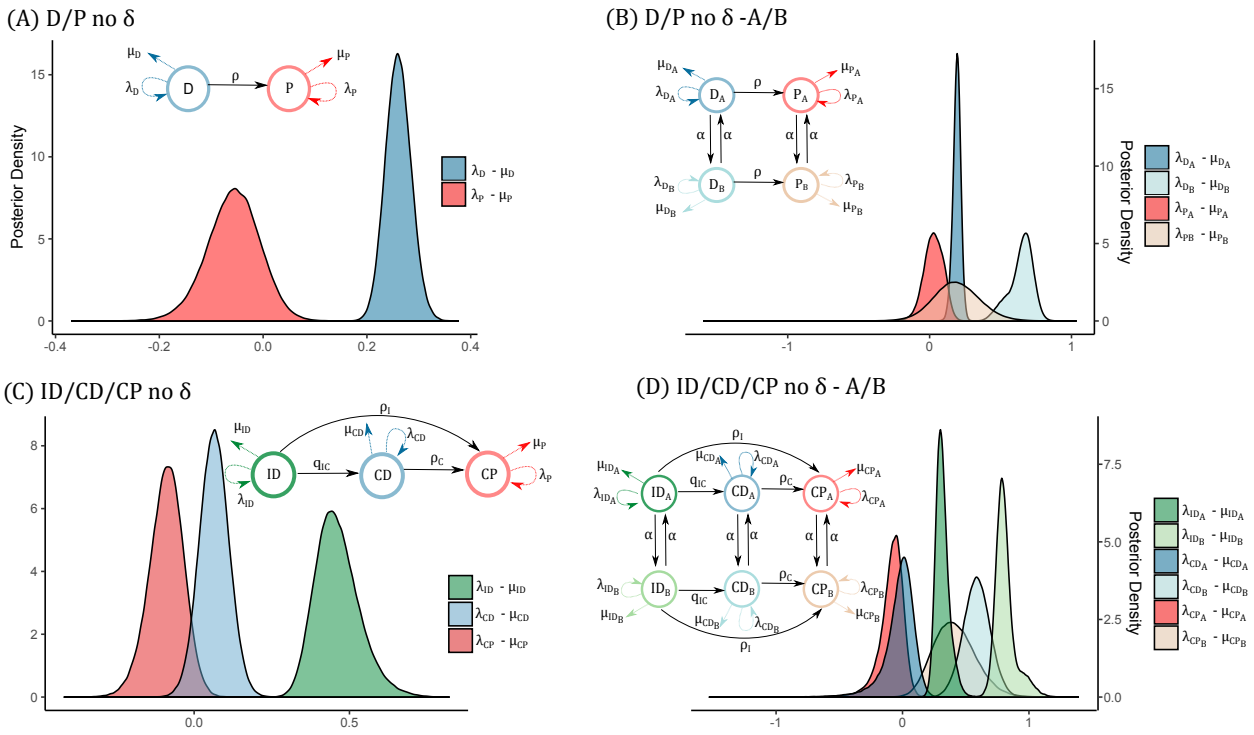

Figure 3: Net diversification rates for all models that do not include diploidization.

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 models examined. States encoded with letter D: diploid, P: polyploid, C: self-compatible, I: self-incompatible, A: first state of hidden trait, B: second state of hidden trait, δ: re-diploidization.

Ploidy Models					Breeding System Models			Ploidy 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.

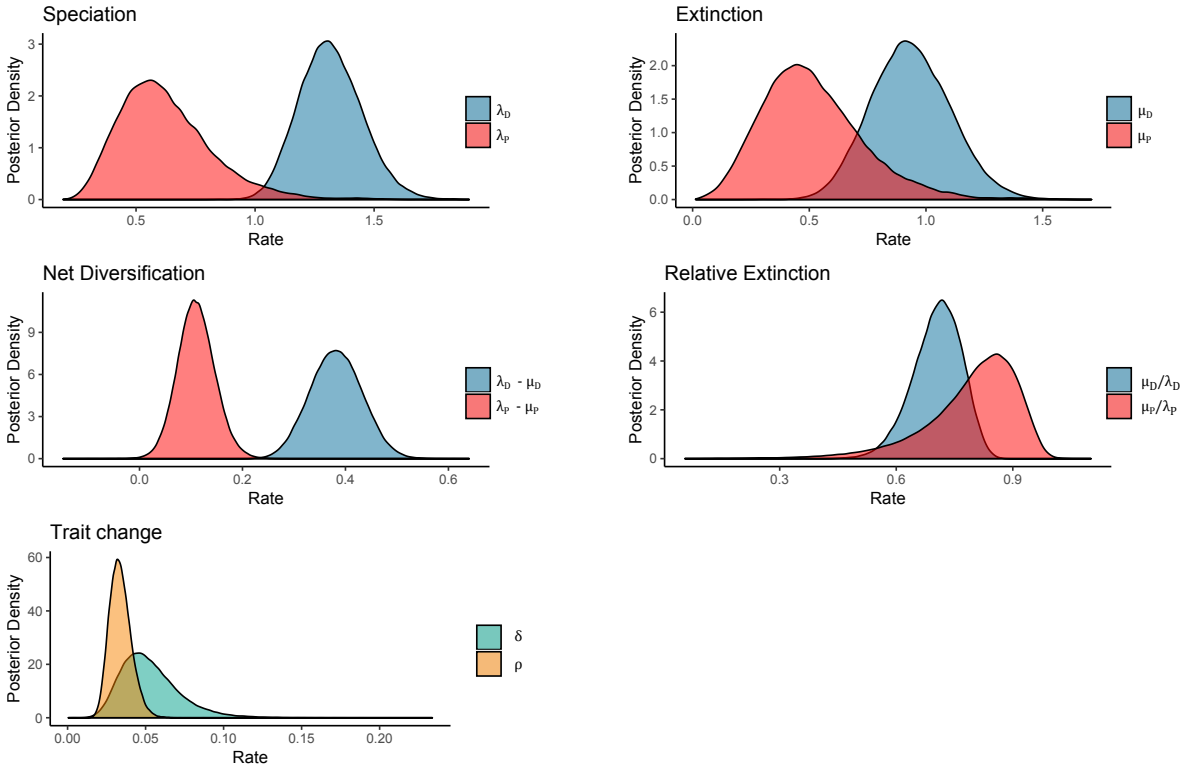

Figure S1: Posterior distribution for each of the parameters in the D/P polyploidy model

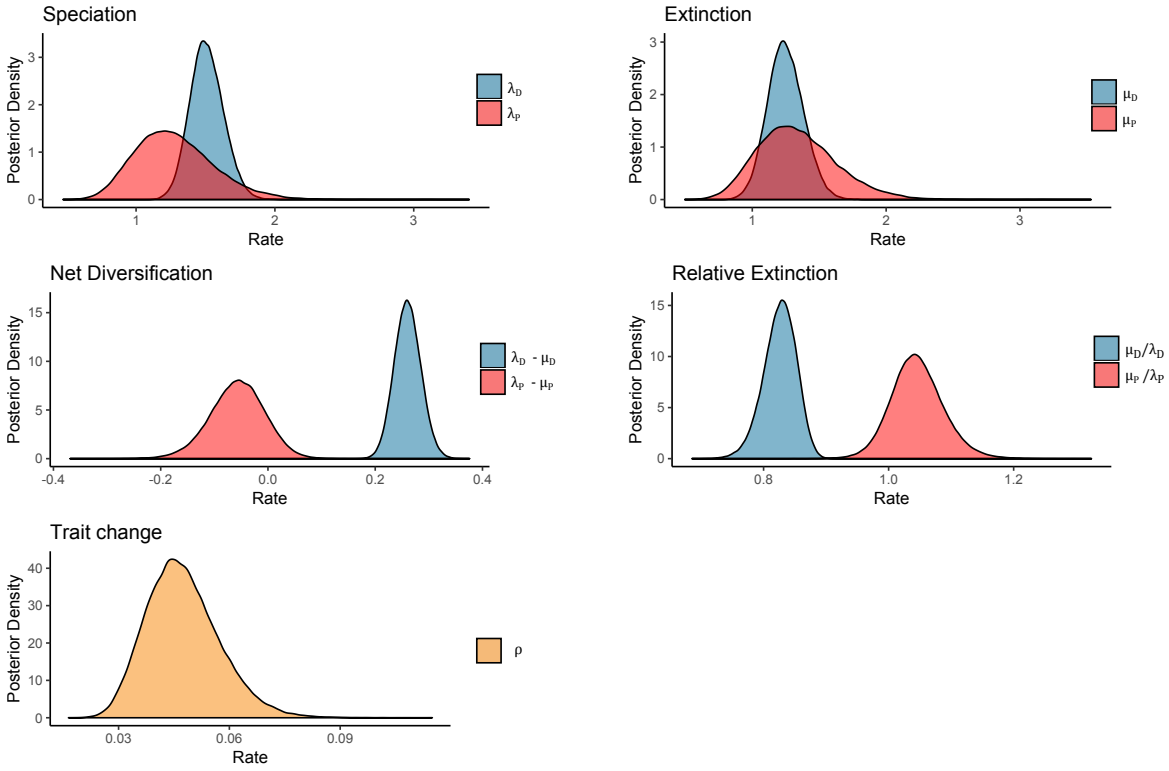

Figure S2: Posterior distribution for each of the parameters in the D/P no δ polyploidy model

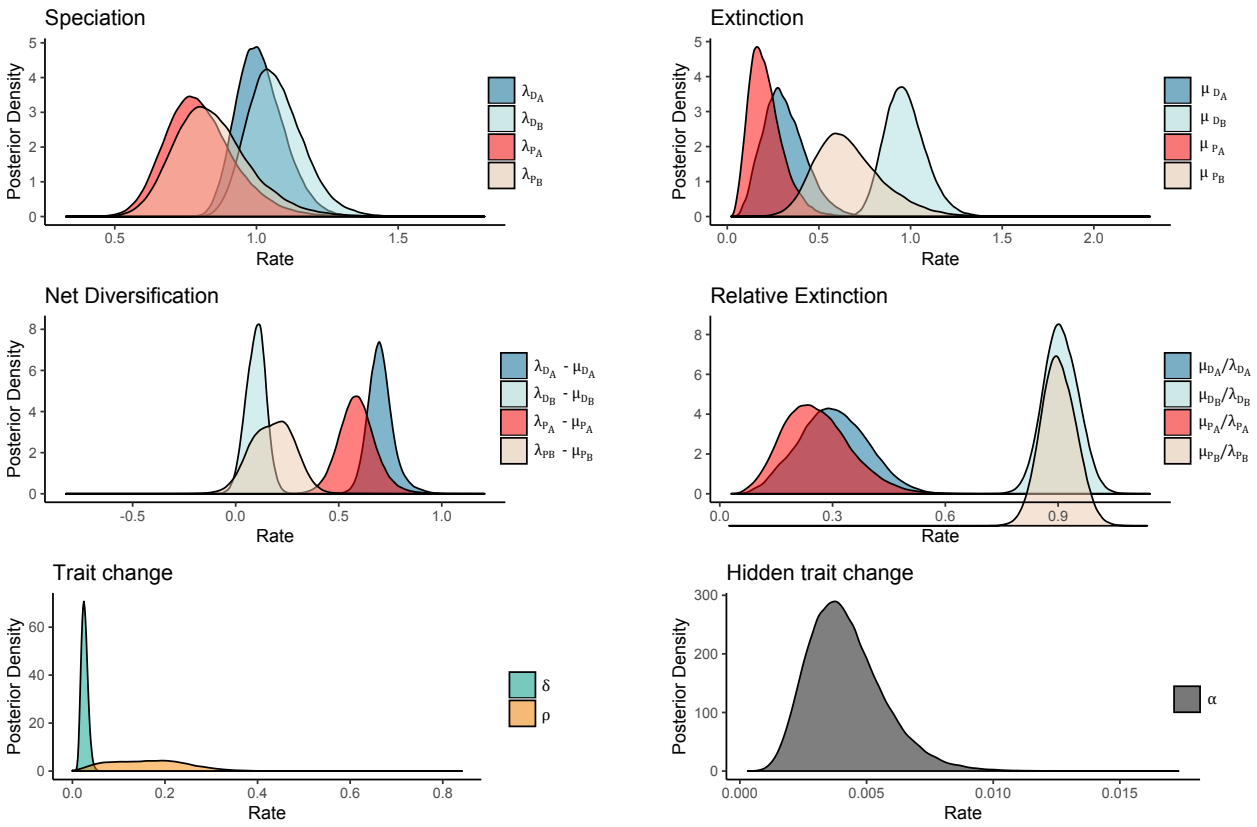

Figure S3: Posterior distribution for each of the parameters in the D/P-A/B polyploidy model

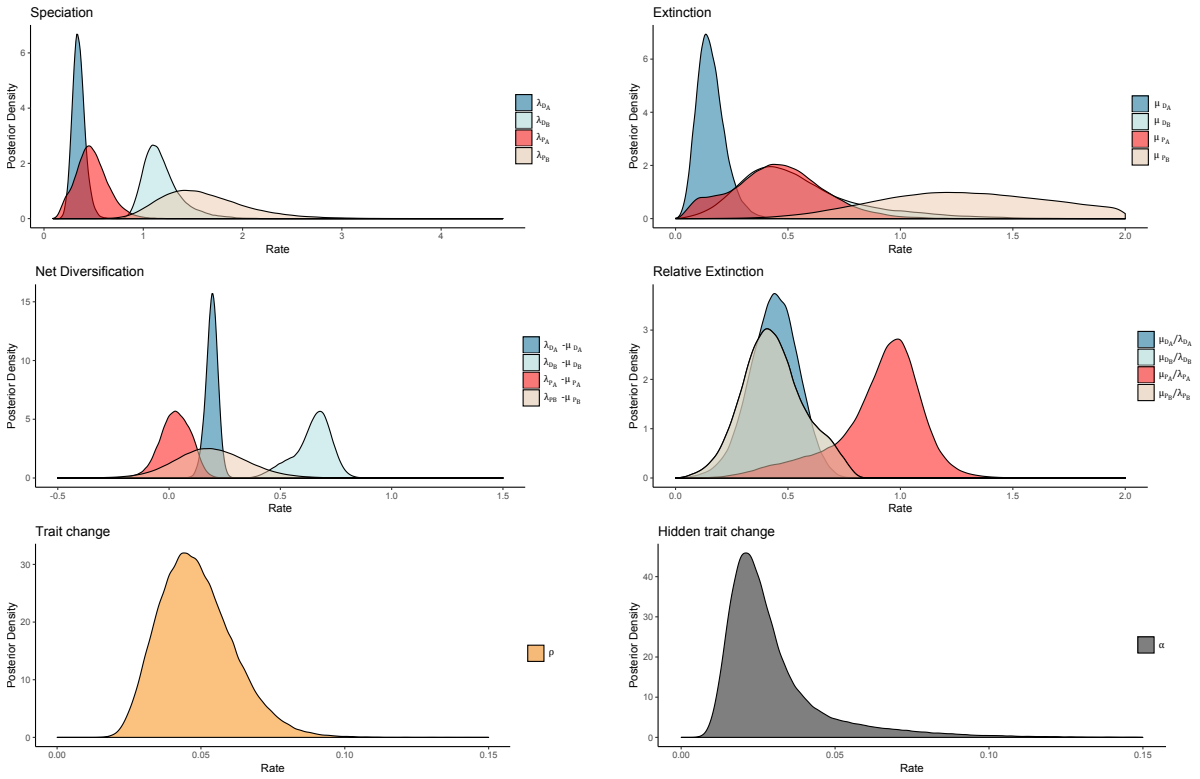

Figure S4: Posterior distribution for each of the parameters in the D/P no δ-A/B polyploidy model

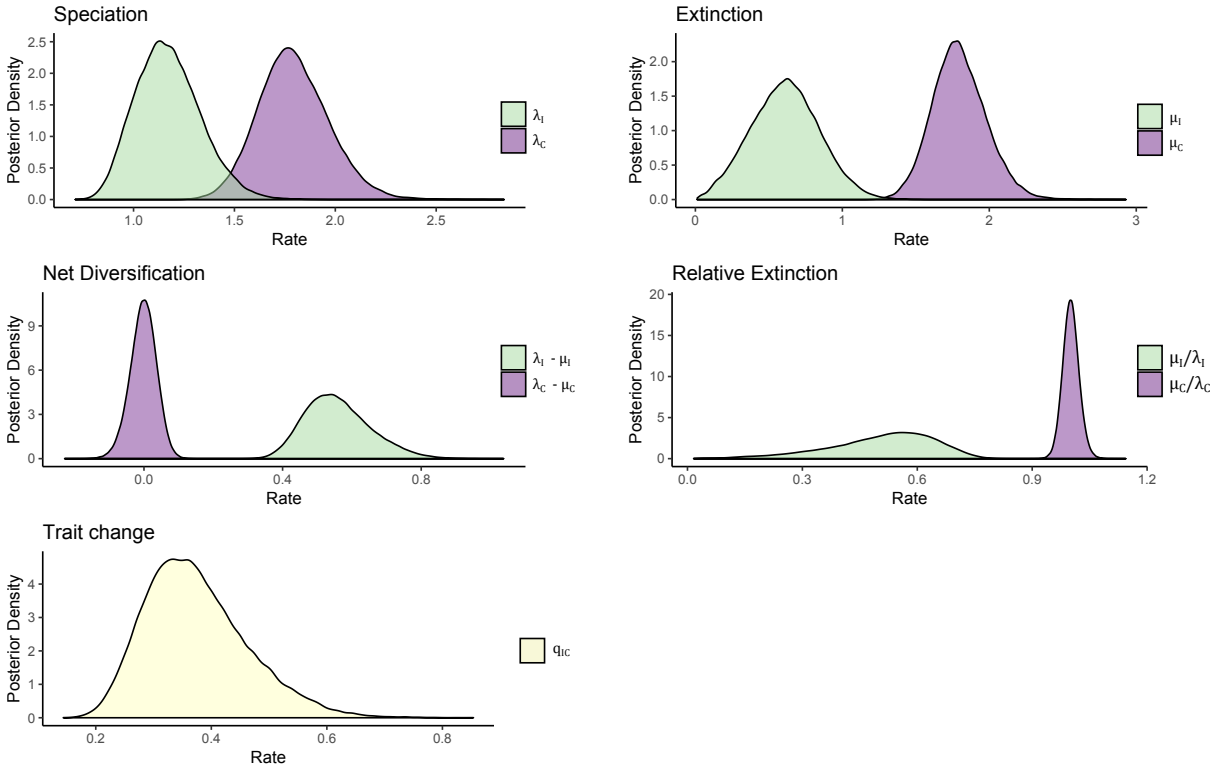

Figure S5: Posterior distribution for each of the parameters in the I/C breeding system model

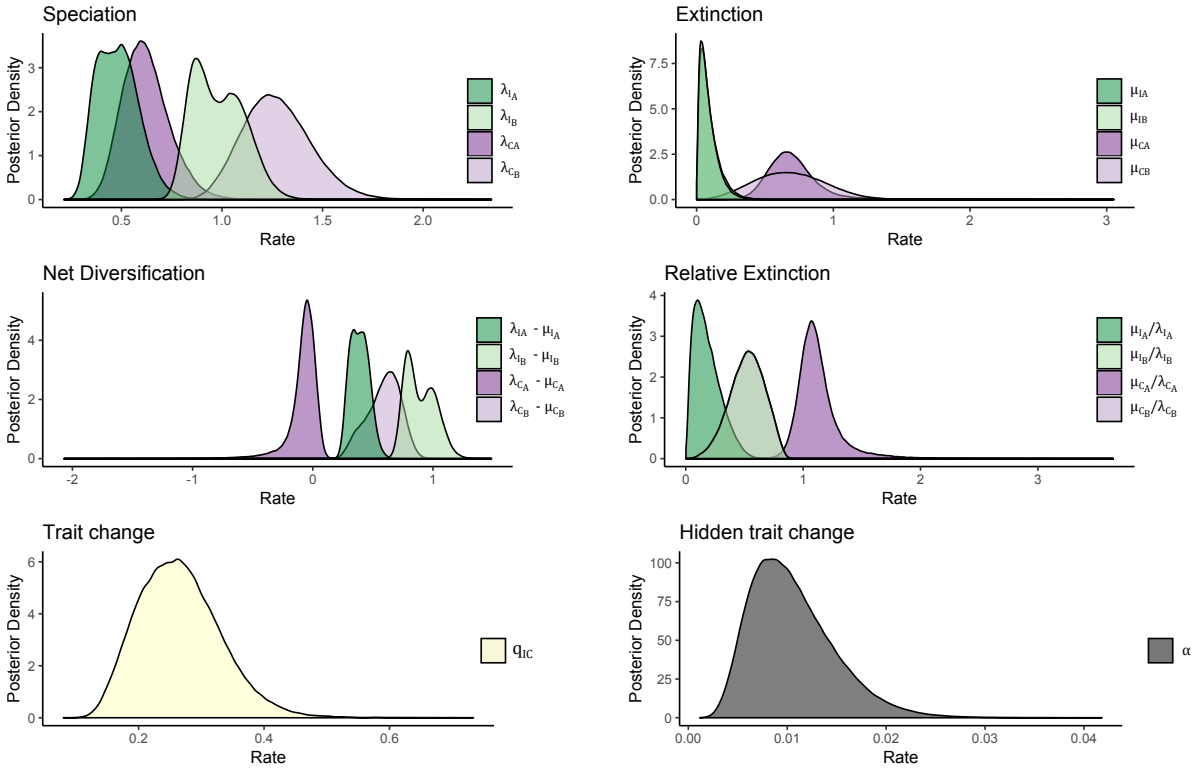

Figure S6: Posterior distribution for each of the parameters in the I/C-A/B breeding system model

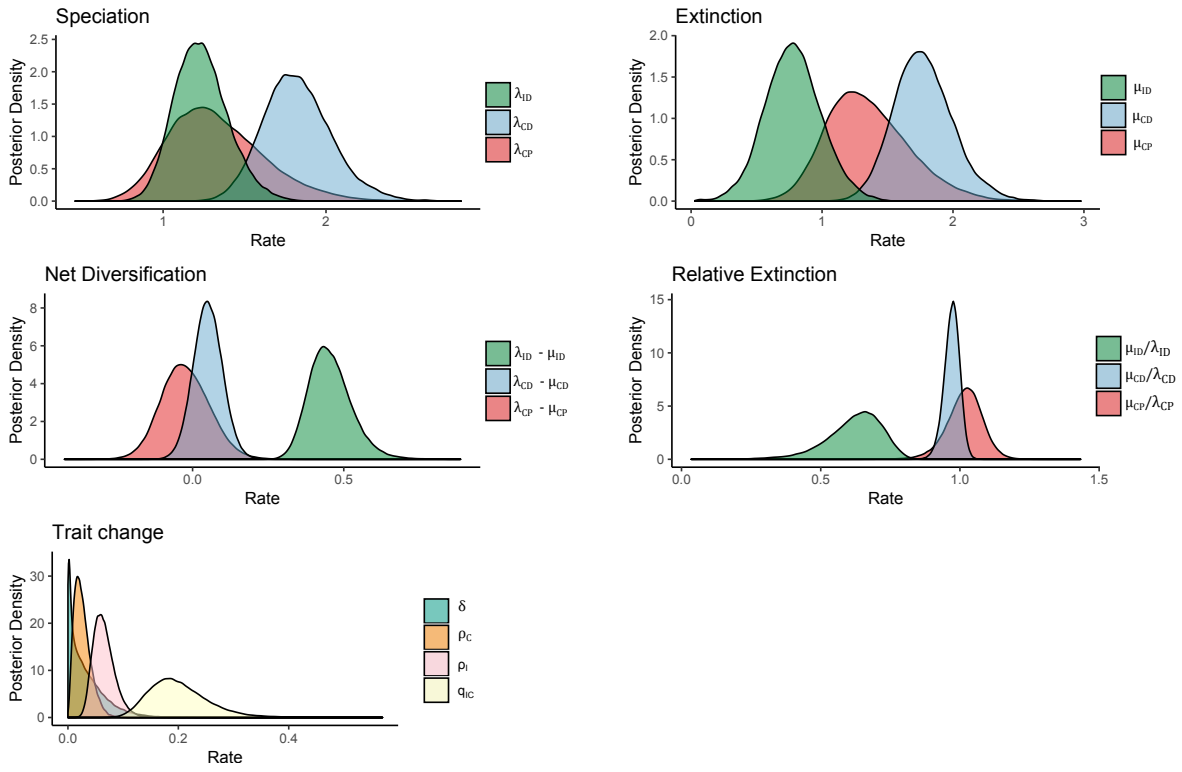

Figure S7: Posterior distribution for each of the parameters in the ID/CD/CP polyploidy and breeding system model

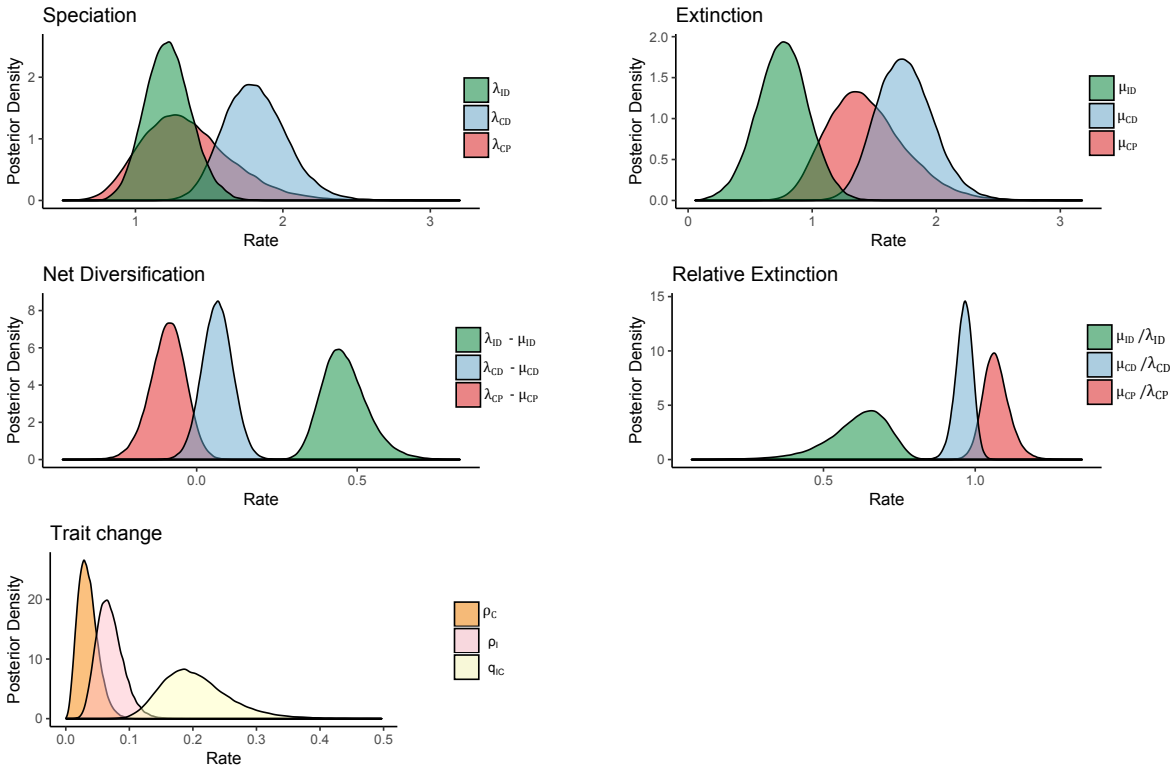

Figure S8: Posterior distribution for each of the parameters in the ID/CD/CP no δ polyploidy and breeding system model

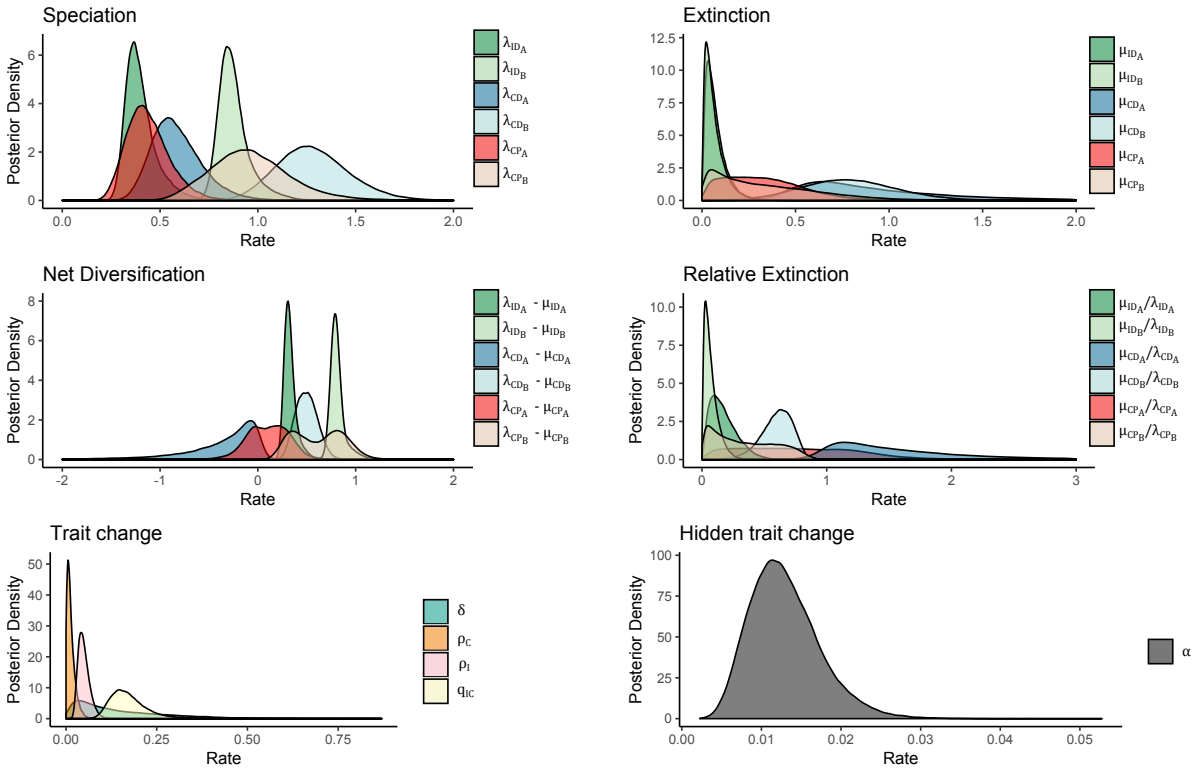

Figure S9: Posterior distribution for each of the parameters in the ID/CD/CP-A/B polyploidy and breeding system model

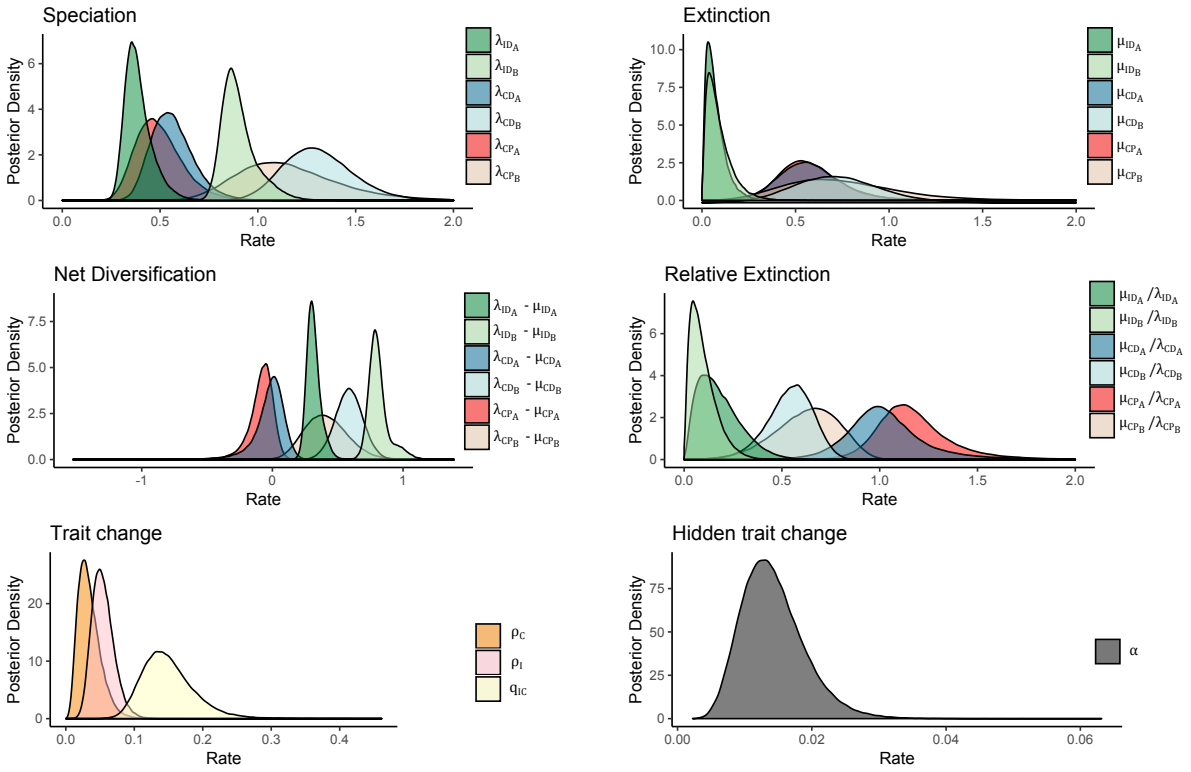

Figure S10: Posterior distribution for each of the parameters in the ID/CD/CP no δ- A/B polyploidy and breeding system model

	D/P	D/P+A/B	I/C	I/C+A/B	ID/CD/CP	ID/CD/CP+A/B
r_D	0.382	0.698, 0.100	—	—	—	—
r_P	0.109	0.587, 0.182	—	—	—	—
r_I	—	—	0.550	0.386, 0.877	—	—
r_C	—	—	-0.001	-0.059, 0.606	—	—
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}	—	—	0.364		0.194	0.164

	D/P	D/P+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	—	—	—	—
r_C	—	—	—	—
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}	—	—	0.198	0.145

Table S1: Median rate estimates for all 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. The upper section is for models with diploidization, and the lower section is for models without. The supplemental figures show the corresponding distributions of parameter estimates.