The tempo of evolutionary decline in self-compatible plant lineages

William A. Freyman[1] and Sebastian Höhna[2]
[1]*Department of Integrative Biology, University of California, Berkeley, CA, 94720, USA;*
[2]*Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Germany*

Corresponding author: William A. Freyman; E-mail: willfreyman@gmail.com

Abstract

A major goal of evolutionary biology is to identify key evolutionary transitions that correspond with shifts in speciation and extinction rates. Here we test the association of transitions in plant mating system with shifts in diversification rates using a novel stochastic character mapping method that identifies the timing and nature of both transitions in trait evolution and diversification rate shifts over evolutionary trees. Utilizing a state-dependent speciation and extinction (SSE) model and a densely sampled fossil-calibrated phylogeny of the plant family Onagraceae, we confirm long standing theory that self-compatible lineages have higher extinction rates and lower net diversification rates compared to self-incompatible lineages. Furthermore, our results provide the first empirical evidence for the "senescing" diversification rates predicted in highly selfing lineages: our mapped character histories show that the loss of self-incompatibility is followed by a short-term spike in speciation rates, which declines after a time lag of several million years resulting in negative net diversification. Lineages that have long been self-compatible such as *Fuchsia* and *Clarkia* are in a previously unrecognized and ongoing evolutionary decline.

1 Introduction

Evolutionary biologists have long sought to identify key evolutionary transitions that drive the diversification of life (Szathmary and Smith 1995; Sanderson and Donoghue 1996). One such evolutionary transition is the loss of self-incompatibility (SI) in flowering plants (Stebbins 1974; Grant 1981; Barrett 2002). The majority of flowering plants are hermaphrodites, and SI is the genetic system that encourages outcrossing and prevents self-fertilization. Independent transitions to self-compatibility (SC) have occurred repeatedly across the angiosperm phylogeny (Igic et al. 2008) and within the plant family Onagraceae (Raven 1979). Despite the repeated loss of SI, outcrossing is widespread and prevalent in plants, an observation that led Stebbins to hypothesize that SC was an evolutionary dead-end (Stebbins 1957). Stebbins proposed that over evolutionary time SC lineages will have higher extinction rates due to reduced genetic variation and an inability to adapt to changing conditions. However, Stebbins also speculated that SC is maintained by providing a short-term advantage in the form of reproductive assurance. The ability of SC lineages to self reproduce has long been understood to be potentially beneficial in droughts and other conditions where pollinators are rare (Darwin 1876) or after long distance dispersal when a single individual can establish a new population (Baker 1955).

One method that could be used test such hypotheses about the short-term versus long-term consequences of evolutionary transitions is stochastic character mapping on a phylogeny (Nielsen 2002; Huelsenbeck et al. 2003). While most ancestral state reconstruction methods estimate states only at the nodes of a phylogeny, stochastic character mapping explicitly infers the timing and nature of each evolutionary transition along the branches of a phylogeny. However, current approaches to stochastic character mapping have two major limitations: the commonly used rejection sampling approach proposed by Nielsen (2002) is inefficient for characters with large state spaces (Huelsenbeck et al. 2003; Hobolth and Stone 2009), and more importantly current methods only apply to models of character evolution that are finite state substitution processes. While the first limitation has been partially overcome through uniformization techniques (Rodrigue et al. 2008; Irvahn and Minin 2014), a novel approach is needed for models with infinite state spaces, such as

models that specifically test the association of character state transitions with shifts in diversification rates. These models describe the joint evolution of both a character and the phylogeny itself, and define a class of widely used models called state-dependent speciation and extinction models (SSE models; Maddison et al. 2007; FitzJohn 2012; Goldberg and Igić 2012; Freyman and Höhna 2017).

In this work we introduce a method to sample character histories directly from their joint distribution, conditional on the observed tip data and the parameters of the model of character evolution. The method is applicable to standard finite state Markov processes of character evolution and also more complex SSE models that are infinite state Markov processes. The method does not rely on rejection sampling and does not require complex data augmentation (Van Dyk and Meng 2001) schemes to handle unobserved speciation/extinction events. Our implementation directly simulates the number, type, and timing of diversification rate shifts and character state transitions on each branch of the phylogeny. Thus, when applying our method together with a Markov chain Monte Carlo (MCMC; Metropolis et al. 1953) algorithm we can sample efficiently from the posterior distribution of both character state transitions and shifts in diversification rates over the phylogeny.

To illustrate the usefulness of our method to sample stochastic character maps from SSE models, we applied the method to examine the association of diversification rate shifts with mating system transitions over a densely sampled fossil calibrated phylogeny of Onagraceae. Recent studies have reported higher net diversification rates for SI lineages, supporting Stebbins' dead-end hypothesis (Goldberg et al. 2010; Ferrer and Good 2012). However, explicit phylogenetic tests for increased extinction rates in SC lineages are limited to the plant family Solanaceae, where increased rates of speciation in SC lineages were offset by higher extinction rates, leading to lower overall rates of net diversification in SC lineages compared to SI lineages (Goldberg et al. 2010). In the study by Goldberg et al., the association of mating system transitions with shifts in extinction and speciation rates was tested using the Binary State Speciation and Extinction model (BiSSE; Maddison et al. 2007). More recently, BiSSE has been shown to be prone to falsely identifying a positive association when diversification rate shifts are associated with another character *not* included in the model (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015). One approach to reduce the possibility of falsely associating a character with diversification rate heterogeneity is to incorporate a second, unobserved character into the model (i.e., a Hidden State Speciation and Extinction (HiSSE) model; Beaulieu and OMeara 2016). The changes in the unobserved character's state represent background diversification rate changes that are not correlated with the observed character. Our work here is the first to apply a HiSSE-type model to test Stebbins' dead-end hypothesis. We additionally use simulations and Bayes factors (Kass and Raftery 1995) to evaluate the false positive error rate of our model. Most notably, we employ our novel stochastic character mapping method to reconstruct the timing of both diversification rate shifts and transitions in mating system over the Onagraceae phylogeny. We test the hypothesis that SC lineages have higher extinction and speciation rates yet lower net diversification rates compared to SI lineages, and investigate the short-term versus long-term macroevolutionary consequences of the loss of SI.

2 Methods

2.1 Stochastic Character Mapping Method

The primary steps of the novel stochastic character mapping algorithm introduced here are illustrated in Fig. 1. A pseudocode formulation of the algorithm is provided in the Supporting Information (Alg. S1). Additionally, Supporting Information Fig. S1 gives a side by side comparison of the standard stochastic character mapping algorithm as originally described by Nielsen (2002) and the approach introduced in this work. In standard stochastic character mapping the first step is to traverse the tree post-order (tip to root) calculating the conditional likelihood of the character being in each state at each node using Felsenstein's pruning algorithm (Felsenstein 1981). Transition probabilities are computed along each branch using matrix exponentiation. Ancestral states are then sampled at each node during a pre-order (root to tip) traversal. Finally, character histories are repeatedly simulated using rejection sampling for each branch of the tree.

In our new stochastic character mapping algorithm we begin similarly by traversing the tree post-order and calculating conditional likelihoods. However, instead of using matrix exponentiation we calculate the likelihood using a set of ordinary differential equations. We numerically integrate these equations for every arbitrarily small time interval along each branch and store a vector of conditional likelihoods for the character

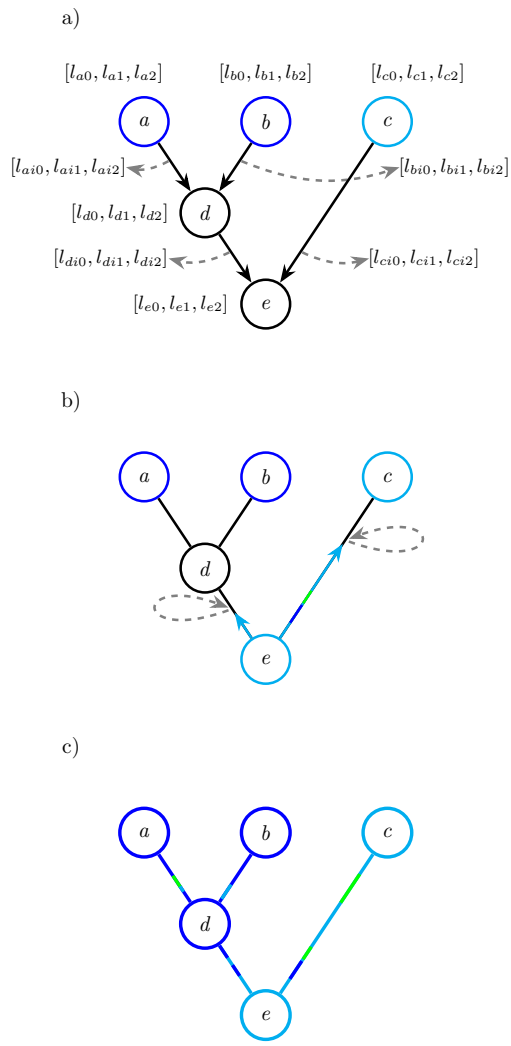

Figure 1: Schematic of the new stochastic character mapping method introduced in this work. The first step in the stochastic character mapping method introduced in this work is *(a)* traversing the tree post-order (tip to root) calculating conditional likelihoods for every arbitrarily small time interval along each branch and at nodes. Next, during a pre-order traversal (root to tip) ancestral states are sampled for each time interval *(b)*, resulting in a full character history *(c)* without the need for a rejection sampling step. See Supporting Information Fig. S1 for a side by side comparison of the standard stochastic character mapping algorithm as originally described by Nielsen (2002) and the approach introduced in this work.

being in each state for every small time interval. The two functions we must numerically integrate are $D_{N,i}(t)$ which is defined as the probability that a lineage in state i at time t evolves into the observed clade N, and $E_i(t)$ which is the probability that a lineage in state i at time t goes extinct before the present, or is not sampled at the present. The equations for these two probabilities are given as Supporting Information Eq. S1 and Eq. S2. Note these equations are identical to the ones describing the Cladogenetic State Speciation and Extinction model (ClaSSE; Goldberg and Igić 2012), which all other discrete SSE models are nested within.

At the tips of the phylogeny (time $t = 0$) the extinction probabilities are $E_i(0) = 1 - \rho$ for all i where ρ is the sampling probability of including that lineage. For lineages with the observed state i, the initial condition is $D_{N,i}(0) = \rho$. The initial condition for all other states j is $D_{N,j}(0) = 0$. When a node L is reached, the probability of it being in state i is calculated by combining the probabilities of its descendant nodes M and N as such:

$$D_{L,i}(t) = \sum_j \sum_k \lambda_{ijk} D_{M,j}(t) D_{N,k}(t), \tag{1}$$

where the rate of a lineage in state i splitting into two lineages in states j and k is λ_{ijk}. Letting $\mathcal{X}$ represent the observed tip data, Ψ an observed phylogeny, and θ_q a particular set of character evolution model parameters, then the likelihood is given by:

$$P(\mathcal{X}, \Psi|\theta_q) = \sum_i \pi_i D_{R,i}(t), \tag{2}$$

where π_i is the root frequency of state i and $D_{R,i}(t)$ is the likelihood of the root node being in state i conditional on having given rise to the observed tree Ψ and the observed tip data $\mathcal{X}$ (Maddison et al. 2007; FitzJohn 2012).

We then sample a complete character history during a pre-order tree traversal in which the root state is first drawn from the marginal likelihoods at the root, and then states are drawn for each small time interval moving towards the tip of the tree conditioned on the state of the previous small time interval. We must again numerically integrate over a set differential equations during this root-to-tip tree traversal. This integration, however, is performed in forward-time, thus a different and new set of differential equations must be used. Letting the rate of anagenetic change from state i to j to be Q_{ij} and the rate of extinction in state i to be μ_i:

$$\begin{aligned} E_i(t-\Delta t) \approx E_i(t) - \Bigg[\mu_i - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i \right) E_i(t) \\ + \sum_{j \neq i} Q_{ij} E_j(t) + \sum_j \sum_k \lambda_{ijk} E_j(t) E_k(t) \Bigg] \Delta t, \end{aligned} \tag{3}$$

$$\begin{aligned} D_{N,i}(t-\Delta t) \approx D_{N,i}(t) + \Bigg[- \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i \right) D_{N,i}(t) \\ + \sum_{j \neq i} Q_{ji} D_{N,j}(t) + D_{N,j}(t) E_k(t) \left(\sum_j \sum_k \lambda_{jik} + \sum_j \sum_k \lambda_{jki} \right) \Bigg] \Delta t. \end{aligned} \tag{4}$$

In the Supporting Information we derive these forward-time differential equations. We demonstrate how the forward-time equations correctly handle non-reversible models of character evolution and validate the forward-time computation of $D_{N,i}(t)$ and $E_i(t)$. With this approach we can directly sample character histories from an SSE process in forward-time, resulting in a complete stochastic character map sample without the need for rejection sampling or uniformization, see Figure 1.

2.1.1 Implementation

The stochastic character mapping method described here is implemented in C++ in the software RevBayes (Höhna et al. 2014, 2016). The RevGadgets R package (available at `https://github.com/revbayes/RevGadgets`) can be used to generate plots from RevBayes output. Scripts to run all RevBayes analyses presented here can be found in the repository at `https://github.com/wf8/onagraceae`.

2.2 Onagraceae Phylogenetic Analyses

DNA sequences for Onagraceae and Lythraceae were mined from GenBank using SUMAC (Freyman 2015). Lythraceae was selected as an outgroup since previous molecular phylogenetic analyses place it sister to Onagraceae (Sytsma et al. 2004). Information about the alignments and GenBank accessions used can be found in the Supporting Information. Phylogeny and divergence times were inferred using RevBayes (Höhna et al. 2016). Details regarding the fossil and secondary calibrations, the model of molecular evolution, and MCMC analyses are given in the Supporting Information.

2.3 Analyses of Mating System Evolution

The mating system of Onagraceae species were scored as either self-compatible or self-incompatible following Wagner et al. (Wagner et al. 2007).

2.3.1 HiSSE Model

To test whether diversification rate heterogeneity is associated with shifts in mating system or changes in other unmeasured traits, we used a model with 4 states that describes the joint evolution of mating system as well as an unobserved character with hidden states a and b (Fig. 2). Since the RNase-based gametophytic system of self-incompatibility found in Onagraceae is ancestral for all eudicots (Steinbachs and Holsinger 2002), we used an irreversible model that only allowed transitions from self-incompatible to self-compatible. For each of the 4 states we estimated speciation (λ) and extinction (μ) rates. While estimating diversification rates, we accounted for uncertainty in phylogeny and divergence times by sampling 200 trees from the posterior distribution of trees. For details on priors used and the MCMC analyses see the Supporting Information.

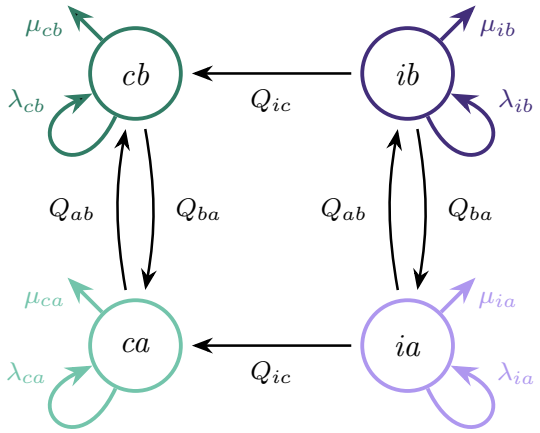

Figure 2: SSE model depicting states and rate parameters used to infer mating system evolution. The states are labeled ca, cb, ia, and ib, representing self-compatible hidden state a, self-compatible hidden state b, self-incompatible hidden state a, and self-incompatible hidden state b, respectively. Independent extinction μ and speciation λ rates were estimated for each of the 4 states, as well as the rate of transitioning from self-incompatible to self-compatible Q_{ic} and the rates of transitioning between the hidden states Q_{ab} and Q_{ba}.

2.3.2 Model Comparisons and Error Rates

To test whether diversification rate heterogeneity was *not* associated with shifts in mating system, we calculated a Bayes factor (Kass and Raftery 1995) to compare the mating system dependent diversification model described above with a mating system independent diversification model. The independent model had 4 states and the same parameters as the dependent model, except that the speciation and extinction rates were fixed so they only varied between the hidden states a and b. Hence, λ_{ca} was fixed to equal λ_{ia}, λ_{cb} was fixed to λ_{ib}, μ_{ca} was fixed to μ_{ia}, and μ_{cb} was fixed to μ_{ib}.

To evaluate the false positive error rate we performed a series of simulations that tested the power of our models to reject false associations between shifts in mating system and diversification rate shifts. Trees were simulated under a BiSSE model, and then diversification *independent* binary characters representing mating system were simulated over the trees. For each simulation replicate Bayes factors were calculated to compare the fit of the mating system dependent diversification model and mating system independent diversification model. Details on the simulations are provided in the Supporting Information.

All Bayes factors were calculated using the stepping stone method (Xie et al. 2010; Höhna et al. 2017) as implemented in RevBayes. Marginal likelihood estimates were run for 50 path steps and 19000 generations within each step. The Bayes factor was then calculated as twice the difference in the natural log marginal likelihoods (Kass and Raftery 1995).

3 Results

3.1 Onagraceae Phylogeny

In our estimated phylogeny, all currently recognized Onagraceae genera (Wagner et al. 2007) were strongly supported to be monophyletic with posterior probabilities > 0.98. The crown age of Onagraceae was estimated to be 98.8 Ma (94.0 Ma – 107.3 Ma 95% HPD; Fig. 3), and a summary of the divergence times of major clades within Onagraceae can be found in Supporting Information Table 3.

3.2 Stochastic Character Maps

Under the state-dependent diversification model, repeated independent losses of SI across the Onagraceae phylogeny were found to be associated with shifts in diversification rates (Fig. 3). Additionally, transitions between the unobserved character states a and b were also associated with diversification rate heterogeneity. Uncertainty in the timing of diversification rate shifts and character state transitions was generally low, but increased along long branches where there was relatively little information regarding the exact timing of transitions (Fig. 4). Following the loss of self-incompatibility, there was an evolutionary time lag (mean 1.97 My) until net diversification (speciation minus extinction) turned negative (Fig. 5). In many cases the loss of self-incompatibility occurred in an ancestral lineage followed by multiple shifts to negative net diversification in descendant lineages. To account for these non-independent time lags we divided the time during ancestral lineages evenly among the individual shifts to negative net diversification in descendant lineages.

3.3 Diversification Rate Estimates

Within either hidden state (a or b) SC lineages had generally higher speciation and extinction rates compared to SI lineages (Fig. 3). SC lineages in state a had a speciation rate of 0.12 (0.02 – 0.23 95% HPD) compared to 0.16 (0.09 – 0.24 95% HPD) in SI lineages in state a. For SC lineages in state b the speciation rate was 1.66 (0.98 – 2.41 95% HPD) compared to 0.65 (0.45 – 0.85 95% HPD) in SI lineages in state b. Similarly, SC lineages in state a had an extinction rate of 0.35 (0.25 – 0.48 95% HPD) compared to 0.04 (0.00 – 0.09 95% HPD) in SI lineages in state a. For SC lineages in state b the extinction rate was 1.36 (0.65 – 2.19 95% HPD) compared to 0.10 (0.00 – 0.29 95% HPD) in SI lineages in state b.

Despite higher speciation and extinction rates, SC lineages had lower net diversification compared to SI lineages. Net diversification was found to be negative for most but not all extant SC lineages. The net diversification rate for SC lineages in state a was -0.23 (-0.32 – -0.14 95% HPD), compared to 0.13 (0.05 – 0.19 95% HPD) in SI lineages in state a. For SC lineages in state b the net diversification rate was 0.30 (0.15 – 0.46 95% HPD), compared to 0.55 (0.39 – 0.71 95% HPD) in SI lineages in state b.

3.4 Model Comparisons and Error Rates

The state-dependent diversification model of mating system evolution (Fig. 2) was "decisively" supported over the state-independent diversification model with a Bayes factor ($2ln$BF) of 19.9 (Jeffreys 1961). Bayes factors calculated using simulated datasets showed that the false positive error rate was low (Figure 6). The false positive rate for "strong" support ($2ln\text{BF} > 6$; Kass and Raftery 1995) was 0.05, and the false positive rate for "very strong" support ($2ln\text{BF} > 10$; Kass and Raftery 1995) was 0.0.

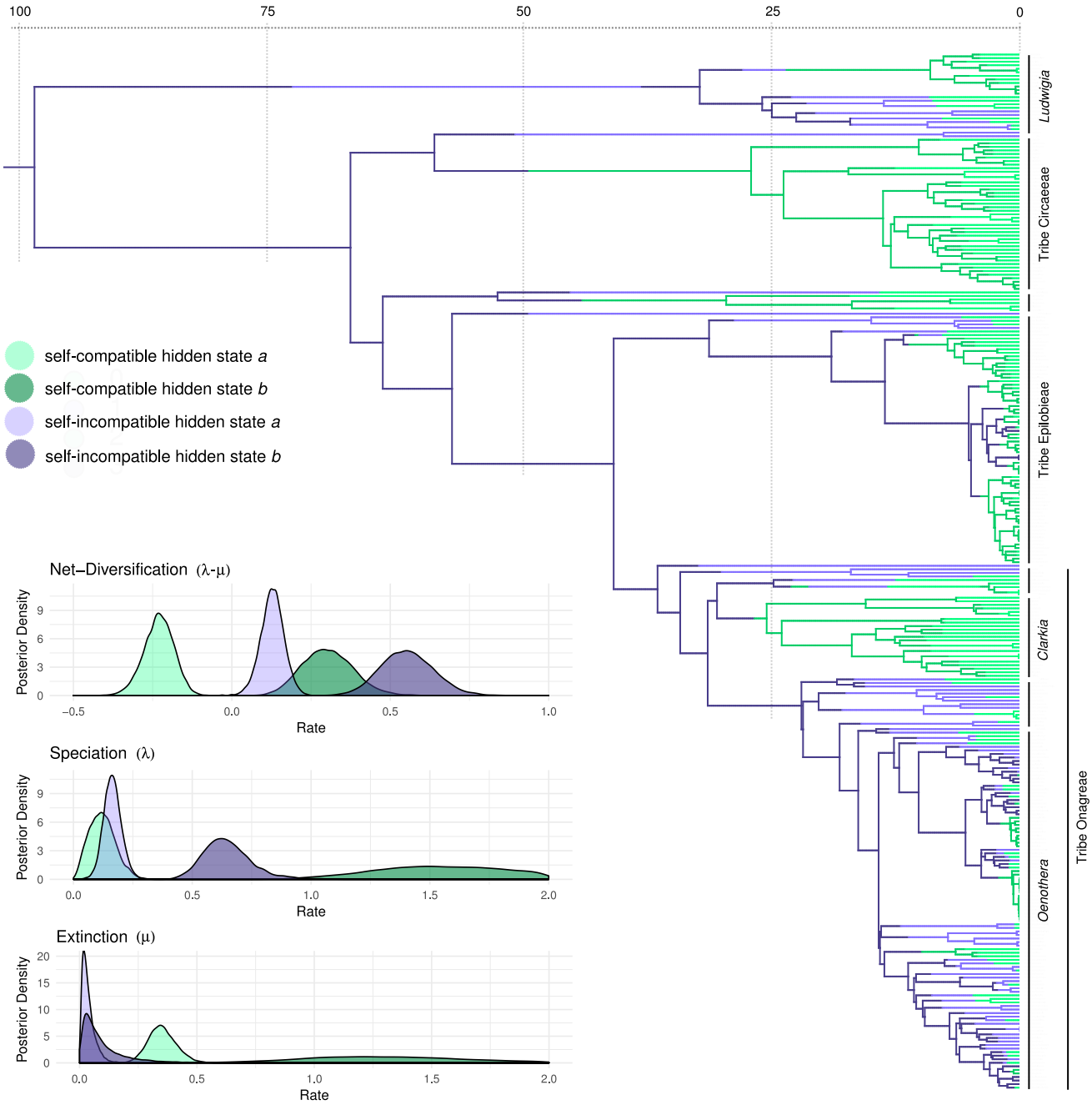

Figure 3: Maximum a posteriori reconstruction of mating system evolution and shifts in diversification rates in Onagraceae. Divergence times in millions of years are indicated by the axis at the top. Note that in this marginal summary reconstruction some transitions are displayed such as the loss and regain of self-incompatibility that were impossible in any single sampled character history. This indicates high uncertainty in the exact timing of transitions (see Fig. 4). The inset panels show posterior densities of net diversification ($\lambda - \mu$), speciation (λ), and extinction (μ) rates in millions of years. Changes in mating system and an unobserved character (hidden states a and b) are both associated with diversification rate heterogeneity. Within either hidden state (a or b) self-compatible lineages have higher extinction and speciation rates yet lower net diversification rates compared to self-incompatible lineages.

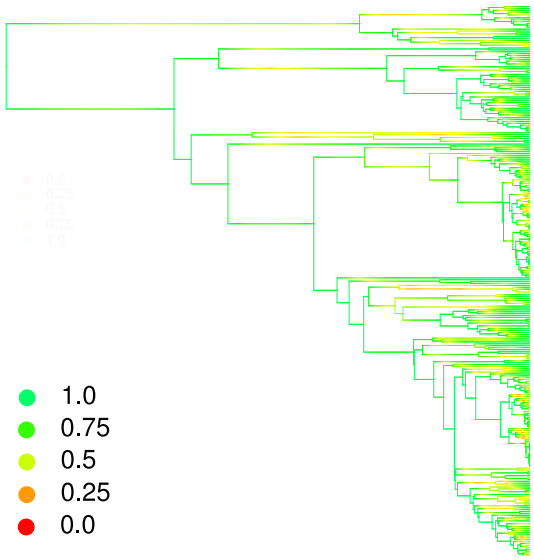

Figure 4: Posterior probabilities of the maximum a posteriori reconstruction of mating system evolution and shifts in diversification rates in Onagraceae. Marginal posterior probabilities of the character states shown in Fig. 3. Uncertainty was highest along long branches where there was relatively little information regarding the timing of transitions.

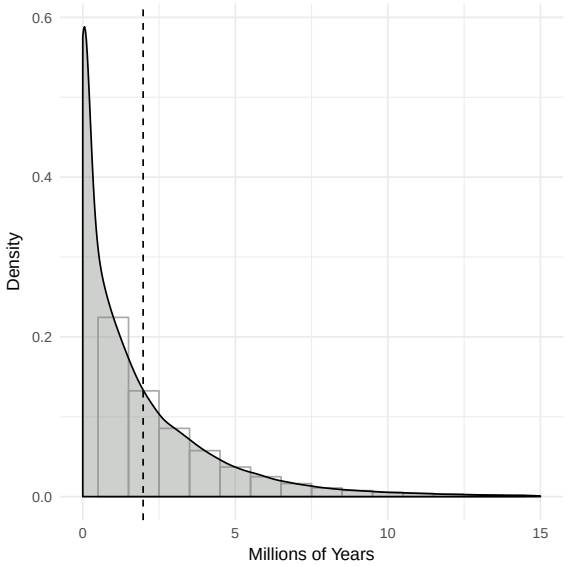

Figure 5: The time lag from the loss of self-incompatibility until the onset of evolutionary decline. The time in millions of years after the loss of self-incompatibility until the net diversification rate became negative measured over 10000 stochastic character map samples. The mean time lag until evolutionary decline was 1.97 million years (indicated by a dashed line).

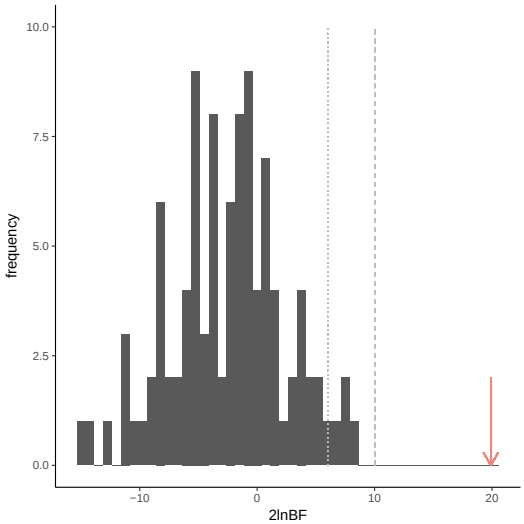

Figure 6: Bayes factors ($2ln$BF) comparing the fit of the state-dependent diversification model of mating system evolution with the state-independent diversification model. The red arrow indicates the "decisive" support found for the empirical Onagraceae data ($2ln\text{BF} = 19.9$; Jeffreys 1961). The dark grey bars represent Bayes factors calculated for 100 datasets simulated under a state-independent diversification model. The dotted light grey line indicates "strong" support ($2ln\text{BF} > 6$; Kass and Raftery 1995), and the dashed light grey line indicates "very strong" support ($2ln\text{BF} > 10$; Kass and Raftery 1995).

4 Discussion

The stochastic character map results reveal that the loss of SI has different short term and long term macroevolutionary consequences. Lineages with relatively recent losses of SI like *Epilobium* are undergoing a burst in both speciation and extinction rates with a positive net diversification rate. However, lineages that have long been SC such as *Fuchsia* (Tribe Circaeeae) and *Clarkia* are in a previously unrecognized evolutionary decline. These lineages went through an increase in both speciation and extinction rates a long time ago —after the loss of SI— but now only the extinction rates remain elevated and the speciation rates have declined, resulting in negative net diversification. The stochastic character maps quantify the speed of this evolutionary decline in SC lineages; while the mean time until evolutionary decline was 1.97 My, there was a large amount of variation in time estimates (Fig. 5). This variation could be due to differences in realized selfing/outcrossing rates of different lineages. Lineages with higher selfing rates likely build up load due to weakly deleterious mutations more quickly, leading to a more rapid mutational meltdown and eventual evolutionary decline. Furthermore, even if mutational load is low, the loss of genetic variation in highly selfing lineages will reduce the probability that such lineages can respond adequately to natural selection, such as imposed by a changing or new environment, thus increasing potential for extinction.

These results confirm long-standing theory about the macroevolutionary consequences of SC (Darwin 1876; Stebbins 1957). These consequences include the increased probability of going extinct due to the accumulation of harmful mutations (Lynch et al. 1995a,b) and an increased rate of speciation which may be driven by higher among-population differentiation and reproductive assurance that facilitates colonization of new habitats (Baker 1955; Hartfield 2016). The advantages of reproductive assurance may explain why transitions to SC occur repeatedly (Igic et al. 2008; Lande and Schemske 1985). However, our results reveal that this advantage is short-lived; the burst of increased speciation following the loss of SI eventually declines, possibly due to failing to adapt to changing conditions and the accumulation of deleterious mutations. The overall macroevolutionary pattern is one in which SC lineages undergo rapid bursts of increased speciation that eventually decline, doomed by intensified extinction and thus supporting Stebbins' hypothesis of SC as an evolutionary dead-end (Stebbins 1957). These results provide the first empirical evidence for the "senescing" diversification rates predicted in highly selfing lineages by Ho and Agrawal (2017), who proposed that primarily selfing lineages may at first diversify at higher rates than outcrossing lineages but over time slow down due to elevated extinction rates.

Our findings corroborate previous analyses performed in the plant family Solanceae (Goldberg et al. 2010), where SC lineages were also found to have higher speciation and extinction rates yet lower net diversification. Our results, however, are the first to show that this pattern is supported even when other unmeasured factors affect diversification rate heterogeneity. Intuitively it is clear that no single factor drives all diversfication rate heterogeneity in diverse and complex clades such as Onagraceae. Indeed, in some lineages of *Oenothera* the loss of sexual recombination and segregation due to extensive chromosome translocations (a condition called Permanent Translocation Heterozygosity) is associated with increased diversification rates (Johnson et al. 2011). Furthermore, other factors such as polyploidy and shifts in habitat, growth form, or life cycle may impact diversification rates (Mayrose et al. 2011; Donoghue 2005; Eriksson and Bremer 1992).

Stochastic character mapping of state-dependent diversification can be a powerful tool for examining the timing and nature of both shifts in diversification rates and character state transitions on a phylogeny. Character mapping reveals which stages of the unobserved character a lineage goes through; e.g. after the loss of self-incompatibility transitions are predominantly from hidden state b to a, representing shifts from positive net diversification to negative net diversification. Furthermore, character mapping infers the state of the lineages in the present and so reveals which tips of the phylogeny are currently undergoing positive or negative net diversification. If used with an SSE model in which all states are hidden (no observed states) our method will "paint" the location of shifts in diversification rate regimes over the tree. Distributions of character map samples could be used for posterior predictive assessments of model fit (Nielsen 2002; Bollback 2006; Höhna et al. 2017) and for testing whether multiple characters coevolve (Huelsenbeck et al. 2003; Bollback 2006). Our hope is that these approaches enable researchers to examine the macroevolutionary impacts of the diverse processes shaping the tree of life with increasing quantitative rigor.

5 Acknowledgements

Thank you to Bruce Baldwin, John Huelsenbeck, Emma Goldberg, Michael Landis, Seema Sheth, and Carl Rothfels and his lab group for discussions that have improved our work. W.A.F. was supported by grants from the National Science Foundation (GRFP DGE 1106400 and DDIG DEB 1601402). Computations were performed on the Savio computational cluster provided by the Berkeley Research Computing program at the University of California, Berkeley.

References

Baker, H. G. 1955. Self-compatibility and establishment after'long-distance'dispersal. Evolution 9:347–349.

Barrett, S. C. 2002. The evolution of plant sexual diversity. Nature Reviews Genetics 3:274–284.

Beaulieu, J. M. and B. C. OMeara. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Systematic Biology 65:583–601.

Bollback, J. P. 2006. Simmap: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.

Darwin, C. 1876. The effects of cross and self fertilization in the vegetable kingdom.

Donoghue, M. J. 2005. Key innovations, convergence, and success: macroevolutionary lessons from plant phylogeny. Paleobiology 31:77–93.

Eriksson, O. and B. Bremer. 1992. Pollination systems, dispersal modes, life forms, and diversification rates in angiosperm families. Evolution 46:258–266.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

Ferrer, M. M. and S. V. Good. 2012. Self-sterility in flowering plants: preventing self-fertilization increases family diversification rates. Annals of Botany Page mcs124.

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

Freyman, W. A. 2015. SUMAC: Constructing phylogenetic supermatrices and assessing partially decisive taxon coverage. Evolutionary Bioinformatics 11:263.

Freyman, W. A. and S. Höhna. 2017. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Systematic Biology Page in press.

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.

Grant, V. 1981. Plant speciation. New York: Columbia University Press xii, 563p.-illus., maps, chrom. nos.. En 2nd edition. Maps, Chromosome numbers. General (KR, 198300748).

Hartfield, M. 2016. Evolutionary genetic consequences of facultative sex and outcrossing. Journal of Evolutionary Biology 29:5–22.

Ho, E. K. and A. F. Agrawal. 2017. Aging asexual lineages and the evolutionary maintenance of sex. Evolution .

Hobolth, A. and E. A. Stone. 2009. Efficient simulation from finite-state, continuous-time Markov chains with incomplete observations. Annals of Applied Statistics 3:1204–1231.

Höhna, S., L. M. Coghill, G. G. Mount, R. C. Thomson, and J. M. Brown. 2017. P^3: Phylogenetic Posterior Prediction in RevBayes. Molecular Biology and Evolution accepted.

Höhna, S., T. A. Heath, B. Boussau, M. J. Landis, F. Ronquist, and J. P. Huelsenbeck. 2014. Probabilistic Graphical Model Representation in Phylogenetics. Systematic biology 63:753–771.

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. Systematic Biology 65:726–736.

Höhna, S., M. L. Landis, and J. P. Huelsenbeck. 2017. Parallel power posterior analyses for fast computation of marginal likelihoods in phylogenetics. bioRxiv .

Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic mapping of morphological characters. Systematic Biology 52:131–158.

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

Irvahn, J. and V. N. Minin. 2014. Phylogenetic stochastic mapping without matrix exponentiation. Journal of Computational Biology 21:676–690.

Jeffreys, H. 1961. Theory of probability. 3 ed. Oxford University Press.

Johnson, M. T., R. G. FitzJohn, S. D. Smith, M. D. Rausher, and S. P. Otto. 2011. Loss of sexual recombination and segregation is associated with increased diversification in evening primroses. Evolution 65:3230–3240.

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

Lande, R. and D. W. Schemske. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. Evolution 39:24–40.

Lynch, M., J. Conery, and R. Burger. 1995a. Mutation accumulation and the extinction of small populations. The American Naturalist 146:489–518.

Lynch, M., J. Conery, and R. Burger. 1995b. Mutational meltdowns in sexual populations. Evolution Pages 1067–1080.

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.

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 333:1257–1257.

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.

Nielsen, R. 2002. Mapping mutations on phylogenies. Systematic Biology 51:729–739.

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

Raven, P. H. 1979. A survey of reproductive biology in Onagraceae. New Zealand Journal of Botany 17:575–593.

Rodrigue, N., H. Philippe, and N. Lartillot. 2008. Uniformization for sampling realizations of Markov processes: applications to Bayesian implementations of codon substitution models. Bioinformatics 24:56–62.

Sanderson, M. J. and M. J. Donoghue. 1996. Reconstructing shifts in diversification rates on phylogenetic trees. Trends in Ecology & Evolution 11:15–20.

Stebbins, G. L. 1957. Self fertilization and population variability in the higher plants. The American Naturalist 91:337–354.

Stebbins, G. L. 1974. Flowering plants: evolution above the species level.

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

Sytsma, K., A. Litt, and M. Zjhra. 2004. Clades, clocks, and continents: Historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the southern Hemisphere source. International Journal of Plant Sciences 165:S85–S105.

Szathmary, E. and J. M. Smith. 1995. The major evolutionary transitions. Nature 374:227.

Van Dyk, D. A. and X.-L. Meng. 2001. The art of data augmentation. Journal of Computational and Graphical Statistics 10:1–50.

Wagner, W. L., P. C. Hoch, and P. H. Raven. 2007. Revised classification of the Onagraceae. Systematic Botany Monographs 83.

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.

Supporting Information: The tempo of evolutionary decline in self-compatible plant lineages

William A. Freyman[1] and Sebastian Höhna[2]
[1]*Department of Integrative Biology, University of California, Berkeley, CA, 94720, USA;*
[2]*Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Germany*

Corresponding author: William A. Freyman; E-mail: willfreyman@gmail.com

S1 Stochastic Character Mapping Method

S1.1 Comparison of Algorithms

Our novel stochastic character mapping method is formulated as Algorithm (S1). A side by side illustration comparing the primary steps of the standard stochastic character mapping algorithm originally described by Nielsen (2002) and Algorithm (S1) is provided in Figure S1. There are three primary differences in the two algorithms. First, in the original algorithm likelihoods are calculated using matrix exponentiation whereas in the new algorithm likelihoods are calculated using numerical integration of ordinary differential equations. Second, during the post-order tree traversal the original algorithm stores conditional likelihoods of the process being in each state only at the nodes, whereas in the new algorithm they are stored for every very small time interval between which we apply a numerical integration algorithm. Third, during a pre-order tree traversal the original algorithm estimates ancestral states at all nodes and then uses rejection sampling to simulate branch histories, whereas the new method simulates branch histories by sampling directly from the process in forward time using numerical integration. The central challenge in developing our new method was deriving and validating the forward time differential equations necessary for this last pre-order tree traversal step.

S1.2 Derivation of our differential equations

In the following section we will derive the differential equations for our algorithm to compute the probability of the observed lineages and the extinction probabilities both backwards and forwards in time. We additionally show how the forward-time equations must be modified to handle non-reversible models of character evolution when sampling ancestral states or stochastic character maps.

S1.2.1 Differential equations backwards in time

The original derivation of the differential equations for the state-dependent speciation and extinction (SSE) process look backward in time (Maddison et al. 2007). Here we present a generalization of the SSE process to allow for cladogenetic events where daughter lineages may inherit different states (Goldberg and Igić 2012; Ng and Smith 2014). We repeat this known derivation of the backwards process to show the similarities to our forward in time derivation. We present an overview of the possible scenarios of what can happen in a small time interval Δt in Figure S2. We need to consider all these scenarios in our differential equations.

First, let us start with the computation of the extinction probability. That is, we want to compute the probability of a lineage going extinct at time $t+\Delta t$, denoted by $E(t+\Delta t)$, before the present time $t = 0$. We assume that we know the extinction probability of a lineage at time t, denoted by $E(t)$, which is provided by our initial condition that $E(t = 0) = 0$ because the probability of a lineage alive at the present cannot go extinct before the present, or $E(t = 0) = 1 - \rho$ in the case of incomplete taxon sampling. We have five different cases (top row in Figure S2): (1) the lineage goes extinct within the interval Δt; (2) nothing happens in the interval Δt but the lineage eventually goes extinct before the present; (3) a state-change to state j occurs and the lineage now in state j goes extinct before the present; (4) the lineage speciates, giving birth to a left daughter lineage in state j and a right daughter lineage in state k and both lineages eventually

Algorithm S1 Stochastic character mapping algorithm. $D_{Ni}(t)$ is the probability that a lineage in state i at time t evolves into the observed clade N. $E_i(t)$ is the probability that a lineage in state i at time t goes extinct or is not sampled before the present.

Inputs:
- $\mathcal{X}$: the vector of observed tip states.
- t_r: the starting time of the process.
- π: the vector of root state frequencies.
- λ: the vector of speciation rates.
- μ: the vector of extinction rates.
- ρ: the probability of sampling a lineage in the present.
- Q: The matrix of transition rates between states.

Initialize:
- $t \leftarrow 0$ // start at the present
- $E_i(t=0) \leftarrow 1-\rho$ // extinction probability at present time
- **if** $i = \mathcal{X}_{\text{observed}}$ **then**
  - $D_{N,i}(t=0) \leftarrow \rho$ // probability of observed character
- **else**
  - $D_{N,i}(t=0) \leftarrow 0$

while $t \leq t_r$ **do** // post-order tree traversal
- **if** node L is reached **then**
  - $D_{L,i}(t) \leftarrow \sum_j \sum_k \lambda_{ijk} D_{M,j}(t) D_{N,k}(t)$ // combine descendant probabilities
- **else**
  - $L_{N,i}(t) \leftarrow D_{N,i}(t)$ // store the conditional likelihoods for this time interval
  - $E_i(t+\Delta t) \leftarrow E_i(t) +$ // compute conditional likelihoods for next time interval

$$\left[\mu_i - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) E_i(t) + \sum_{j\neq i} Q_{ij} E_j(t) + \sum_j \sum_k \lambda_{ijk} E_j(t) E_k(t)\right] \Delta t$$

// backward-time Equation (S1)

  - $D_{N,i}(t+\Delta t) \leftarrow D_{N,i}(t) +$

$$\left[-\left(\sum_j \sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) D_{N,i}(t) + \sum_{j\neq i} Q_{ij} D_{N,j}(t) + \sum_j \sum_k \lambda_{ijk} \left(D_{N,k}(t) E_j(t) + D_{N,j}(t) E_k(t)\right)\right] \Delta t$$

// backward-time Equation (S2)

  - $t \leftarrow t + \Delta t$ // increment the current t
- **end if**

end while

while $t \geq 0$ **do** // pre-order tree traversal
- **if** $t = t_r$ **then**
  - $s_t \sim \text{Multinomial}\big(n=1, D_N(t_r) \times \pi\big)$ // draw character state at the root
- **else**
  - $s_t \sim \text{Multinomial}\big(n=1, D_N(t) \times L_N(t)\big)$ // draw character state for time t
- **end if**
- $D_{N,s_t} \leftarrow 1$ // condition on the sampled character state
- $D_{N,i\neq s_t} \leftarrow 0$
- $E_i(t-\Delta t) \leftarrow E_i(t) -$

$$\left[\mu_i - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) E_i(t) + \sum_{j\neq i} Q_{ij} E_j(t) + \sum_j \sum_k \lambda_{ijk} E_j(t) E_k(t)\right] \Delta t$$

// forward-time Equation (S4)

- $D_{N,i}(t-\Delta t) \leftarrow D_{N,i}(t) +$

$$\left[-\left(\sum_j \sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) D_{N,i} + \sum_{j\neq i} Q_{ji} D_{N,j}(t) + D_{N,j}(t) E_k(t) \left(\sum_j \sum_k \lambda_{jik} + \sum_j \sum_k \lambda_{jki}\right)\right.$$

// forward-time Equation (S6)

- $t \leftarrow t - \Delta t$ // decrement the current t

end while

return vector of all sampled character states s

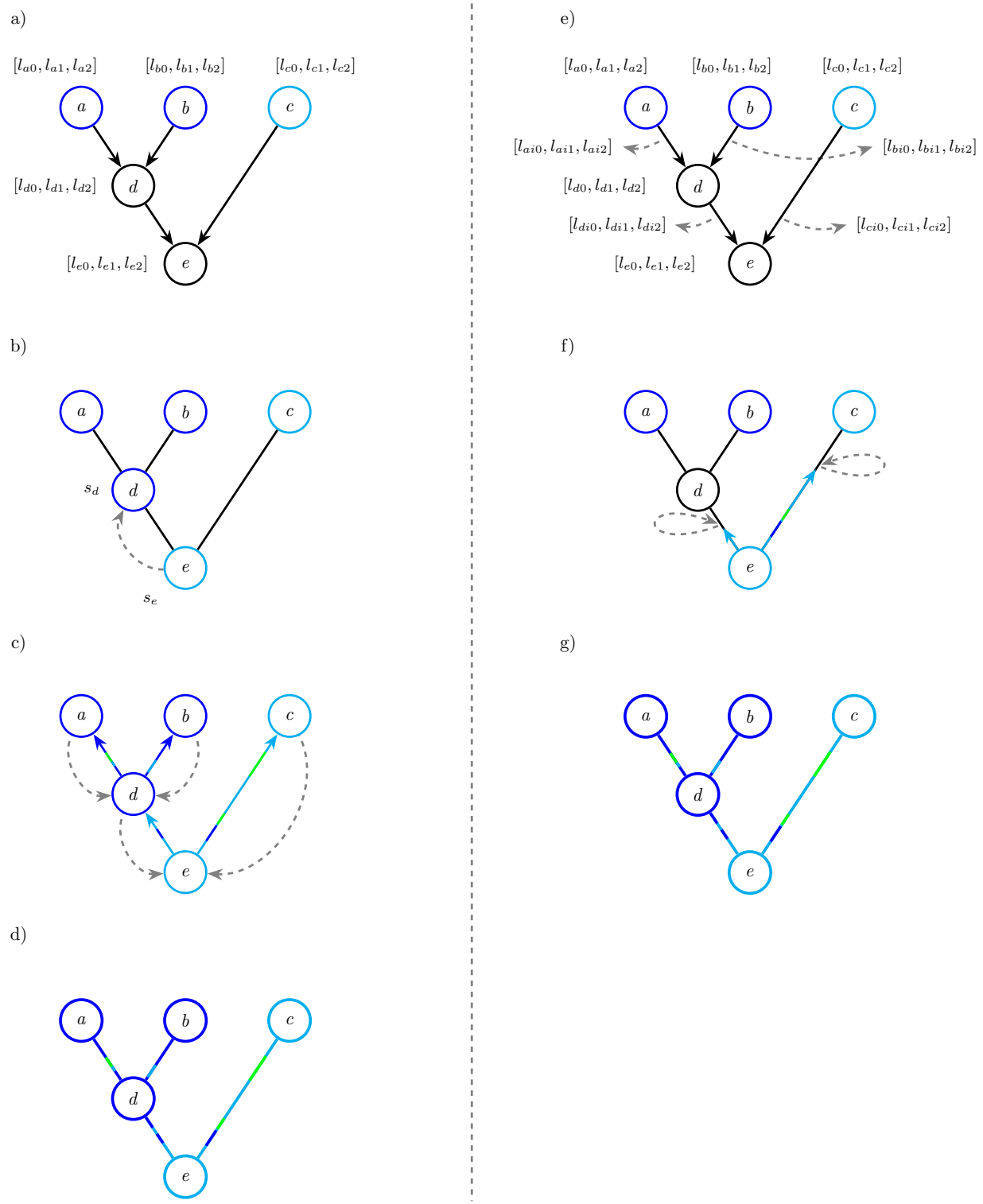

Figure S1: Comparison of stochastic character mapping methods. On the left *(a, b, c, d)* is an illustration of the standard stochastic character mapping algorithm as originally described by Nielsen (2002). On the right *(e, f, g)* is the approach introduced in this work. The first step in standard stochastic character mapping is *(a)* traversing the tree post-order (tip to root) calculating conditional likelihoods for each node. Next, ancestral states are sampled at each node during a pre-order (root to tip) traversal *(b)*. Branch by branch, character histories are then repeatedly simulated using rejection sampling *(c)*, resulting in a full character history *(d)*. The first step in the stochastic character mapping method introduced in this work is *(e)* traversing the tree post-order calculating conditional likelihoods for every arbitrarily small time interval along each branch and at nodes. Next, during a pre-order traversal ancestral states are sampled for each time interval *(f)*, resulting in a full character history *(g)* without the need for a rejection sampling step. See the main text for more details.

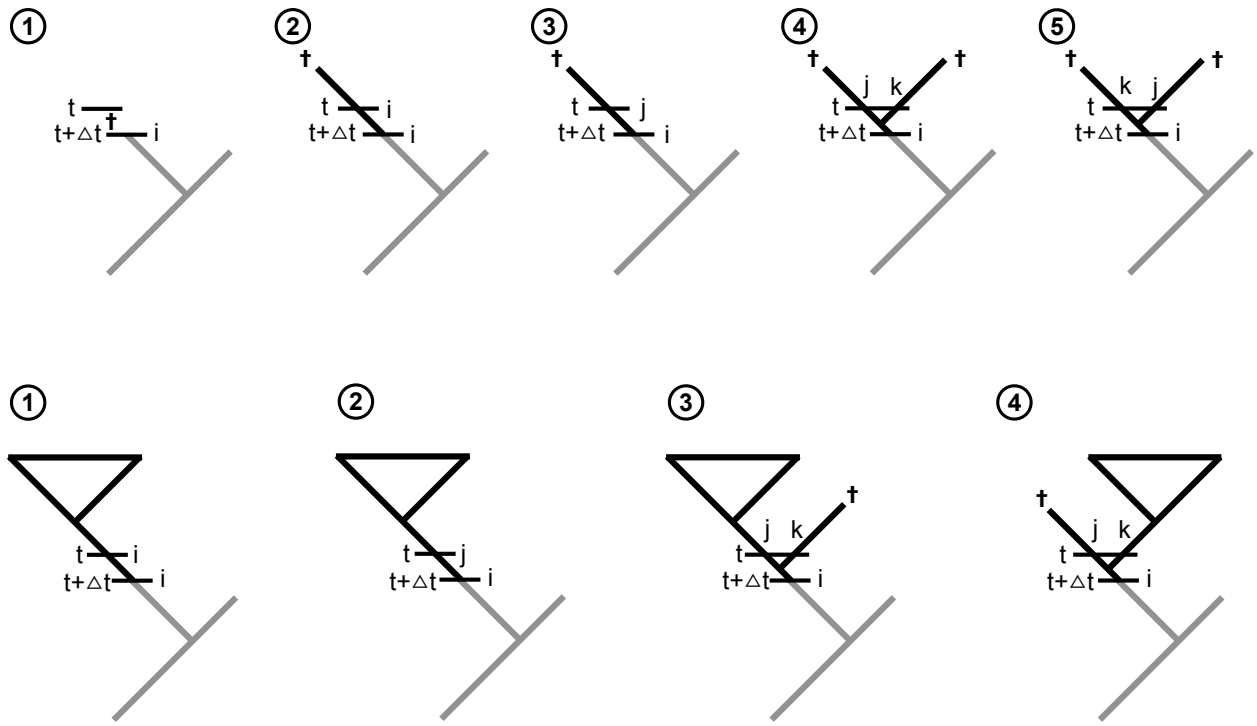

Figure S2: Alternative scenarios of events in a small time interval Δt looking backwards in time. The top row shows the different scenarios for a lineage that goes extinct before the present. Case 1: The lineage goes extinct in the time interval Δt. Case 2: There is no event in the time interval Δt and the lineage goes extinct before the present. Case 3: The lineage undergoes a state-shift event to state j in the time interval Δt and the lineage goes extinct before the present. Case 4: The lineage speciates and leaves a left daughter lineage in state j and a right daughter lineage in state k and both daughter lineages go extinct before the present. Case 5: The lineage speciates and leaves a left daughter lineage in state k and a right daughter lineage in state j and both daughter lineages go extinct before the present. The bottom row shows the different scenarios for an observed lineage. Case 1: There is no event in the time interval Δt. Case 2: The lineage undergoes a state-shift event to state j in the time interval Δt. Case 3: The lineage speciates and leaves a left daughter lineage in state j and a right daughter lineage in state k and only the left daughter lineage survives. Case 4: The lineage speciates and leaves a left daughter lineage in state j and a right daughter lineage in state k and only the right daughter lineage survives.

go extinct before the present, or; (5) the lineage speciates, giving birth to a left daughter lineage in state k and a right daughter lineage in state j and both lineages eventually go extinct before the present. With this description of all possible scenarios we can derive the differential equation.

$$
\begin{aligned}
E_i(t+\Delta t) = E_i(t) + & \qquad \text{(S1)} \\
\Bigg[& \mu_i && \text{Case (1)} \\
& - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i \right) E_i(t) && \text{Case (2)} \\
& + \sum_{j \neq i} Q_{ij} E_j(t) && \text{Case (3)} \\
& + \sum_j \sum_k \lambda_{ijk} E_j(t) E_k(t) \Bigg] \Delta t && \text{Case (4) and (5)}
\end{aligned}
$$

Similarly, we can consider all possible scenarios for an observed lineage. We have four different cases (bottom row in Figure S2): (1) nothing happens in the interval Δt; (2) a state-change to state j occurs; (3) the lineage speciates, giving birth to a left daughter lineage in state j and a right daughter lineage in state k and only the left daughter lineage survives until the present, or; (4) the lineage speciates, giving birth to a left daughter

lineage in state j and a right daughter lineage in state k and only the right daughter lineage survives until the present. Again, these scenarios are sufficient to derive the differential equation for the probability of an observed lineage, denoted $D(t)$.

$$
\begin{aligned}
D_{N,i}(t+\Delta t) = D_{N,i}(t) + & \qquad \text{(S2)} \\
\Bigg[& -\left(\sum_j \sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) D_{N,i}(t) && \text{Case (1)} \\
& + \sum_{j\neq i} Q_{ij} D_{N,j}(t) && \text{Case (2)} \\
& + \sum_j \sum_k \lambda_{ijk} \Big(D_{N,k}(t) E_j(t) + D_{N,j}(t) E_k(t) \Big) \Bigg] \Delta t && \text{Case (3) and (4)}
\end{aligned}
$$

S1.2.2 Differential equations forward in time

Next, we want to compute the probability of extinction and the probability of an observed lineage forward in time. For the probability of extinction this is, in principle, almost identical to the backward in time equations. However, now we assume that we know $E(t)$ and want to compute $E(t-\Delta t)$. We already computed $E(t_{root})$ and $D(t_{root})$ in our post-order tree traversal (from the tips to root). We use $E(t_{root})$ as the initial conditions to approximate $E(t-\Delta t)$. Again, we have the same five different cases (top row in Figure S2): (1) the lineage goes extinct within the interval Δt; (2) nothing happens in the interval Δt but the lineage eventually goes extinct before the present; (3) a state-change to state j occurs and the lineage now in state j goes extinct before the present; (4) the lineage speciates, giving birth to a left daughter lineage in state j and a right daughter lineage in state k and both lineages eventually go extinct before the present, or; (5) the lineage speciates, giving birth to a left daughter lineage in state k and a right daughter lineage in state j and both lineages eventually go extinct before the present. However, these are the events that can happen in the future and we included the probabilities of these events already in $E(t)$. Thus, we need to subtract instead of adding all possible scenarios that lead to the extinction of the lineage in the time interval Δt from $E(t)$ to obtain $E(t-\Delta t)$. This gives us the differential equation for the extinction probability as

$$
\begin{aligned}
E_i(t-\Delta t) = E_i(t) - & \qquad \text{(S3)} \\
\Bigg[& \mu_i && \text{Case (1)} \\
& - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) E_i(t) && \text{Case (2)} \\
& + \sum_{j\neq i} Q_{ij} E_j(t-\Delta t) && \text{Case (3)} \\
& + \sum_j \sum_k \lambda_{ijk} E_j(t-\Delta t) E_k(t-\Delta t) \Bigg] \Delta t && \text{Case (4) and (5)}
\end{aligned}
$$

Unfortunately, we cannot solve Equation (S3) directly because we do not know $E_j(t-\Delta t)$ and $E_k(t-\Delta t)$. Instead, we will approximate Equation (S3) by using $E_j(t)$ instead of $E_j(t-\Delta t)$, and $E_k(t)$ instead of $E_k(t-\Delta t)$, respectively. Our approximation yields the new differential equation of the extinction probability

by

$$
\begin{aligned}
E_i(t-\Delta t) \approx E_i(t) - & \\
\Bigg[& \mu_i && \text{Case (1)} \\
& - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i \right) E_i(t) && \text{Case (2)} \\
& + \sum_{j \neq i} Q_{ij} E_j(t) && \text{Case (3)} \\
& + \sum_j \sum_k \lambda_{ijk} E_j(t) E_k(t) \Bigg] \Delta t && \text{Case (4) and (5)}
\end{aligned}
\tag{S4}
$$

The derivation of the probability of an observed lineage in forward time is slightly different. When sampling a character history from the process we must compute $D(t-\Delta t)$ conditioned upon the character state sampled at time t. This does not effect the probability of a lineage going extinct before the present, so we can use $E(t_{root})$ as the initial conditions to approximate $E(t-\Delta t)$. The initial conditions for the probability of an observed lineage, on the other hand, must account for the sampled character state. For example, if we sample the state a at time t our initial conditions to compute $D(t-\Delta t)$ must be $D_a(t) = 1.0$ and $D_b(t) = 0.0$ for all other character states b. Additionally, we must consider the process in forward time with all possible scenarios instead of backwards in time and subtracting the possible scenarios. We have four different cases that are similar to the cases for the backward in time computation (bottom row in Figure S2), however here the character state transitions are reversed since we are looking forward in time: (1) nothing happens in the interval Δt; (2) with probability $D_{N,j}(t)$ the lineage was in state j and then a state-change to state i occurs; (3) with probability $D_{N,j}(t)$ the lineage was in state j and then speciates, giving birth to a left daughter lineage in state i and a right daughter lineage in state k and only the left daughter lineage survives until the present (the probability of extinction of the right daughter lineage is given by $E_k(t-\Delta t)$), or; (4) with probability $D_{N,j}(t)$ the lineage was in state j and then speciates, giving birth to a left daughter lineage in state k and a right daughter lineage in state i and only the right daughter lineage survives until the present (the probability of extinction of the left daughter lineage is given by $E_k(t-\Delta t)$). From these four scenarios we derive the differential equation.

$$
\begin{aligned}
D_{N,i}(t-\Delta t) = D_{N,i}(t) + & \\
\Bigg[& - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i \right) D_{N,i}(t) && \text{Case (1)} \\
& + \sum_{j \neq i} Q_{ji} D_{N,j}(t) && \text{Case (2)} \\
& + D_{N,j}(t) E_k(t-\Delta t) \left(\sum_j \sum_k \lambda_{jik} + \sum_j \sum_k \lambda_{jki} \right) \Bigg] \Delta t && \text{Case (3) and (4)}
\end{aligned}
\tag{S5}
$$

As before, we cannot solve Equation (S5) directly because we do not know $E_k(t-\Delta t)$. Thus, we use the same approximation as before and substitute $E_k(t)$ for $E_k(t-\Delta t)$. This substitution gives our approximated

differential equation.

$$
\begin{aligned}
D_{N,i}(t-\Delta t) \approx D_{N,i}(t) + & \\
\Bigg[& -\left(\sum_j \sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) D_{N,i}(t) && \text{Case (1)} \\
& + \sum_{j\neq i} Q_{ji} D_{N,j}(t) && \text{Case (2)} \\
& + D_{N,j}(t) E_k(t) \left(\sum_j \sum_k \lambda_{jik} + \sum_j \sum_k \lambda_{jki}\right)\Bigg] \Delta t && \text{Case (3) and (4)}
\end{aligned}
\tag{S6}
$$

To sample character histories from an SSE process in forward-time during Algorithm (S1) we calculate $E(t-\Delta t)$ using the approximation given by Equation (S4) and $D(t-\Delta t)$ using Equation (S6).

S1.3 Correctness of the forward time equations

For the purpose of demonstrating our forward time equations, we will use a non-symmetrical BiSSE model with states 0 and 1 which have the speciation rates $\lambda_0 = 1$ and $\lambda_1 = 2$, the extinction rates $\mu_0 = 0.5$ and $\mu_1 = 1.5$, and the transition rates $Q_{01} = 0.2$ and $Q_{10} = 2.0$. For simplicity we assume that there are no state changes at speciation events. We will first show that the approximations given by Equation (S4) actually converge to the true probability of extinction if the time interval Δt is very small (goes to zero). Note that we cannot show the same behavior for the forward in time probabilities of the observed lineage, $D(t)$, because when conditioning on a sampled character state the forward in time probabilities will be different than the backward in time probabilities. For these probabilities we provide a different type of validation in Section S1.3.1.

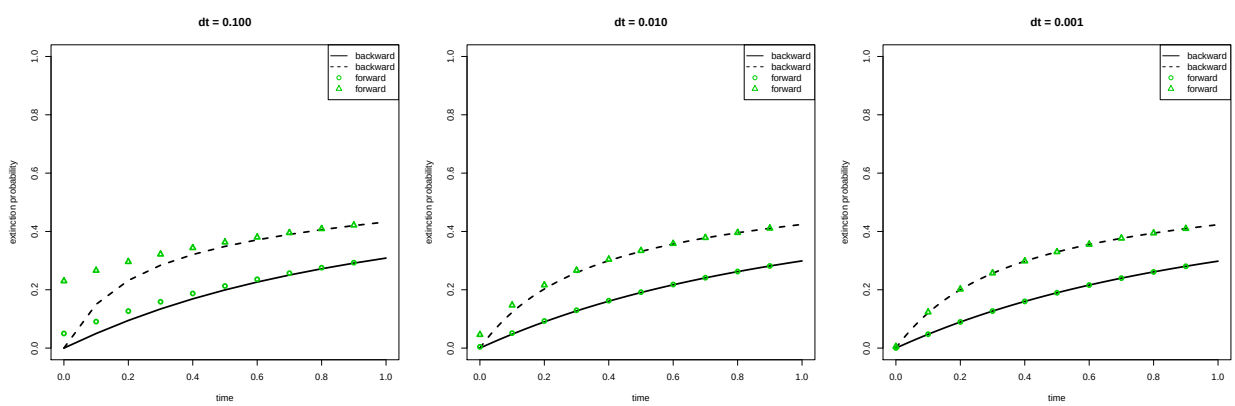

Figure S3: The probability of extinction computed backward and forward in time. Here we compute the extinction probabilities $E_0(t)$ and $E_1(t)$ for a BiSSE model backward and forward in time. Details about the parameters of the BiSSE model are given in the text. We varied the step-size Δt for the numerical integration between 0.1, 0.01, and 0.001 to show that both computations give the same probabilities once Δt is small enough.

We start by computing the probability of extinction and the probability of an observed lineage backward in time for a total time interval of 1.0. We initialize the computation with $E_i(t=0) = 0$ and then compute $E_0(t)$ and $E_1(t)$ backward in time. Then, we use the computed values of $E_i(t=1)$ as the initial values for our forward in time computation. If our approximation is correct, then we should get identical values for the extinction probabilities $E_i(t)$ for any value of t.

Figure S3 shows our computation using three different values for Δt: 0.1, 0.01 and 0.001. We observe that our approximation of the forward in time computation of the probabilities converges to the backward in time computation when $\Delta t \leq 0.001$, which confirms our expectation. An explanation for the convergence is that $E_0(t)$ will be approximately equal to $E_0(t-\Delta t)$, (and $E_1(t)$ to $E_1(t-\Delta t)$) the smaller Δt becomes.

In our actual implementation in `RevBayes` we use an initial step-size of $\Delta t = 10^{-7}$ but apply an adaptive numerical integration routine to minimize the error in the integrated function.

S1.3.1 Validation of the forward time equations against `diversitree`

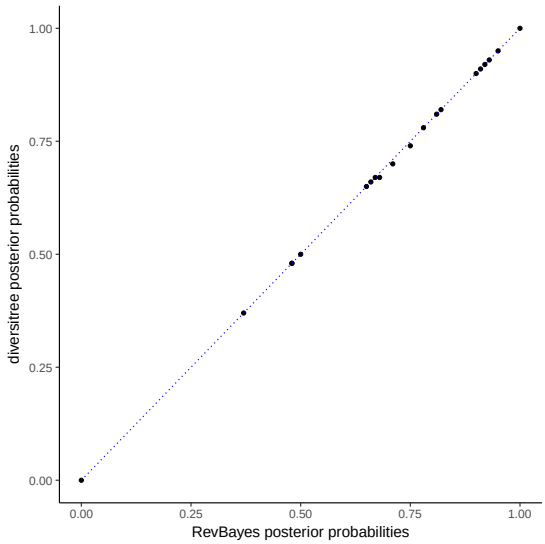

Figure S4: Comparing marginal posterior ancestral state estimates from `diversitree` to those calculated in `RevBayes`. Each point represents the posterior probability of a given node having the ancestral state 0. On the y-axis are the posterior probabilities as analytically calculated by `diversitree`. On the x-axis are the posterior probabilities as calculated by `RevBayes` using Algorithm (S1). Our approximation given in Equation (S6) yields the same posterior probabilities of the ancestral states as `diversitree`. Scripts to repeat this test with various parameter settings are provided in https://github.com/wf8/anc_state_validation.

Finally, we validate our method of sampling character histories from an SSE process in forward-time by testing it against the analytical marginal ancestral state estimation implemented in the `R` package `diversitree` (FitzJohn 2012). Our method as implemented in `RevBayes` works for sampling both ancestral states and stochastic character maps, however `diversitree` can not sample stochastic character maps. Thus we limit our comparison to ancestral states estimated at the nodes of a phylogeny. Though our method works for all SSE models nested within ClaSSE, ancestral state estimation for ClaSSE is not implemented in `diversitree`, so we further limit our comparison to ancestral state estimates for a BiSSE model. Note that as implemented in `RevBayes` the BiSSE, ClaSSE, MuSSE (FitzJohn 2012), HiSSE (Beaulieu and OMeara 2016), ChromoSSE (Freyman and Höhna 2017), and GeoSSE (Goldberg et al. 2011) models use the same C++ classes and algorithms for parameter and ancestral state estimation, so validating under BiSSE should provide confidence in estimates made by `RevBayes` for all these SSE models.

Our method samples character histories from SSE models from their joint distribution conditioned on the tip states and the model parameters during MCMC. In contrast, `diversitree` computes marginal ancestral states analytically. Thus to directly compare results from these two approaches we calculated the marginal posterior probability of each node being in each state from a set of 10000 samples drawn by our Monte Carlo method. Figure S4 compares these estimates under a non-reversible BiSSE model where the tree and tip data were simulated in `diversitree` with the following parameters: $\lambda_0 = 0.2, \lambda_1 = 0.4, \mu_0 = 0.01, \mu_1 = 0.1$, and $q_{01} = 0.1, q_{10} = 0.4$. Figure S4 shows that using the approximation of $E(t - \Delta t)$ given by Equation (S4) and the approximation to compute $D(t - \Delta t)$ in Equation (S6) during Algorithm (S1) results in marginal posterior estimates for the ancestral states that are nearly identical (up to some expected numerical and sampling errors) to those calculated analytically by `diversitree`. Scripts to perform this test with various parameter settings are provided in https://github.com/wf8/anc_state_validation.

S1.4 MCMC Sampling and Computational Efficiency

Our method approximates the posterior distribution of the timing and nature of all character transitions and diversification rate shifts by sampling a large number of stochastically mapped character histories using MCMC. Uncertainty in the phylogeny and other parameters is incorporated by integrating over all possible phylogenetic trees and other parameters jointly. From these sampled character histories the maximum a posteriori character history can be summarized in a number of ways. The approach presented here is to calculate the marginal probabilities of character states for every small time interval along each branch, however one could also calculate the joint posterior probability of an entire character history.

During Algorithm (S1) the rate-limiting step is writing conditional likelihood vectors for every small time interval along every branch on the tree, particularly when the state space of the model is large. The time required is of order $O(n \times m \times r)$, where n is the number of taxa in the tree, m is the number of character states, and r is the number of time intervals. This is reduced by only storing conditional likelihood vectors for all time intervals during the MCMC iterations that are sampled. During unsampled (*i.e.,* thinned) MCMC iterations the likelihood is calculated in the standard way storing conditional likelihood vectors only at the nodes, thus the use of the stochastic mapping algorithm has little impact on the overall computation time.

S2 Onagraceae Phylogenetic Analyses

S2.1 Methods

S2.1.1 Supermatrix Assembly

DNA sequences for Onagraceae and Lythraceae were mined from GenBank using SUMAC (Freyman 2015). Lythraceae was selected as an outgroup since previous molecular phylogenetic analyses place it sister to Onagraceae (Sytsma et al. 2004). SUMAC assembled an 8 gene supermatrix (7 chloroplast loci plus the nuclear ribosomal internal transcribed spacer region) representing a total of 340 taxa. Table S1 summarizes the genes used, their length, and the percent of missing data. Sequences were aligned using MAFFT v7.123b (Katoh and Standley 2013). The default settings in MAFFT were used except that proper sequence polarity was ensured by using the direction adjustment option. Alignments were then concatenated resulting in chimeric operational taxonomic units (OTUs) that do not necessarily represent a single individual.

Table S1: DNA regions mined from GenBank. A total of 340 taxa were included.

DNA Region	# Taxa	Aligned Length	# Variable Sites	Missing data (%)	Taxon Coverage Density
ITS	250	904	481	26.5	0.735
matK	42	895	276	87.6	0.124
ndhF	39	1085	429	88.5	0.115
pgiC	66	7664	3828	80.6	0.194
rbcL	108	1427	388	68.2	0.318
rpl16	54	1139	343	84.1	0.159
rps16	78	1056	335	77.1	0.229
trnL-trnF	261	1434	644	23.2	0.768

Table S2: Fossil and secondary calibrations used as priors in the Bayesian divergence time analysis. Units are in millions of years.

Group	Calibration Type	Placement	Prior Distribution	Mean	SD	Offset	Reference
Circaea	fossil	stem	lognormal	10	2	12	(Grímsson et al. 2012)
Tribe Epilobieae	fossil	stem	lognormal	10	2	12	(Grímsson et al. 2012)
Fuchsia section *Skinnera*	fossil	stem	lognormal	10	2	23	(Lee et al. 2013)
Lythraceae	fossil	crown	lognormal	20	2	81.5	(Graham 2013)
Ludwigia	fossil	stem	lognormal	10	2	57.6	(Zhi-Chen et al. 2004)
Onagraceae + Lythraceae	secondary	crown	normal	93	5	0	(Sytsma et al. 2004)

S2.1.2 Phylogenetic Analyses

Divergence times and phylogeny were jointly estimated using `RevBayes` (Höhna et al. 2014, 2016). Estimates were time calibrated using six node calibrations: four stem fossil calibrations, one crown fossil calibration, and a secondary calibration for the root split between Onagraceae and Lythraceae (Table S2). An uncorrelated lognormal relaxed clock model was used, and each of the eight gene partitions were assigned independent GTR substitution models (Tavaré 1986; Rodriguez et al. 1990). Rate variation across sites was modeled under a gamma distribution approximated by four discrete rate categories (Yang 1994). The constant rate birth-death-sampling tree prior (Nee et al. 1994; Yang and Rannala 1997) was used with the probability of sampling species at the present (ρ) set to 0.27. ρ was calculated by dividing the number of extant species sampled in the supermatrix (340) by the sum of the number of species recognized in Onagraceae (~650) and in Lythraceae (~620).

Four independent MCMC analyses were performed. Each MCMC ran for 15000 generations, where each generation consisted of 837 randomly scheduled Metropolis-Hastings moves. This resulted in four chains that each performed a total of 12,555,000 MCMC steps. Samples of the posterior distribution were drawn every 10 generations, and the first 50% of samples from each chain were discarded as burnin resulting in 750 trees sampled from each of the 4 independent chains. Convergence was assessed by ensuring the effective sample size of each parameter was over 200 for each independent chain. The maximum a posteriori (MAP) tree was then calculated from the combined 3000 tree samples of all 4 chains.

S2.2 Results

All Onagraceae genera described in Wagner et al. (2007) were recovered as monophyletic clades in the MAP summary tree with posterior probabilities > 0.95 (Figure S6). Onagraceae was found to diverge from Lythraceae at 111.3 My (95% HPD interval 106.0 - 116.6 My). Divergence time estimates of other major clades and 95% HPD intervals can be seen in Table S3.

Table S3: Divergence time estimates of major clades.

Clade	Age Type	Mean Age (Ma)	95% HPD Min	95% HPD Max
Onagraceae + Lythraceae	crown	111.3	106.0	116.6
Onagraceae	crown	98.8	94.0	107.3
Ludwigia	crown	32.1	31.3	50.6
Tribe Circaeeae	stem	58.7	42.3	65.0
Tribe Circaeeae	crown	27.0	24.6	29.8
Fuchshia	crown	23.7	23.0	24.1
Lopezia	crown	29.5	25.3	36.7
Tribe Epilobieae	stem	40.8	39.3	47.0
Tribe Epilobieae	crown	31.2	31.2	40.6
Chamerion	crown	14.9	12.6	25.2
Epilobium	crown	18.9	15.7	22.9
Tribe Onagreae	stem	40.8	39.3	47.0
Tribe Onagreae	crown	36.4	31.6	40.5
Taraxia	crown	17.0	9.9	19.5
Gayophytum	crown	3.1	1.8	6.1
Clarkia	crown	25.4	24.6	31.2
Eremothera	crown	10.4	8.1	15.1
Camissonia	crown	9.7	6.4	14.6
Eulobus	crown	4.7	2.6	8.1
Chylismia	crown	14.4	10.8	19.0
Oenothera	crown	14.2	13.0	17.6

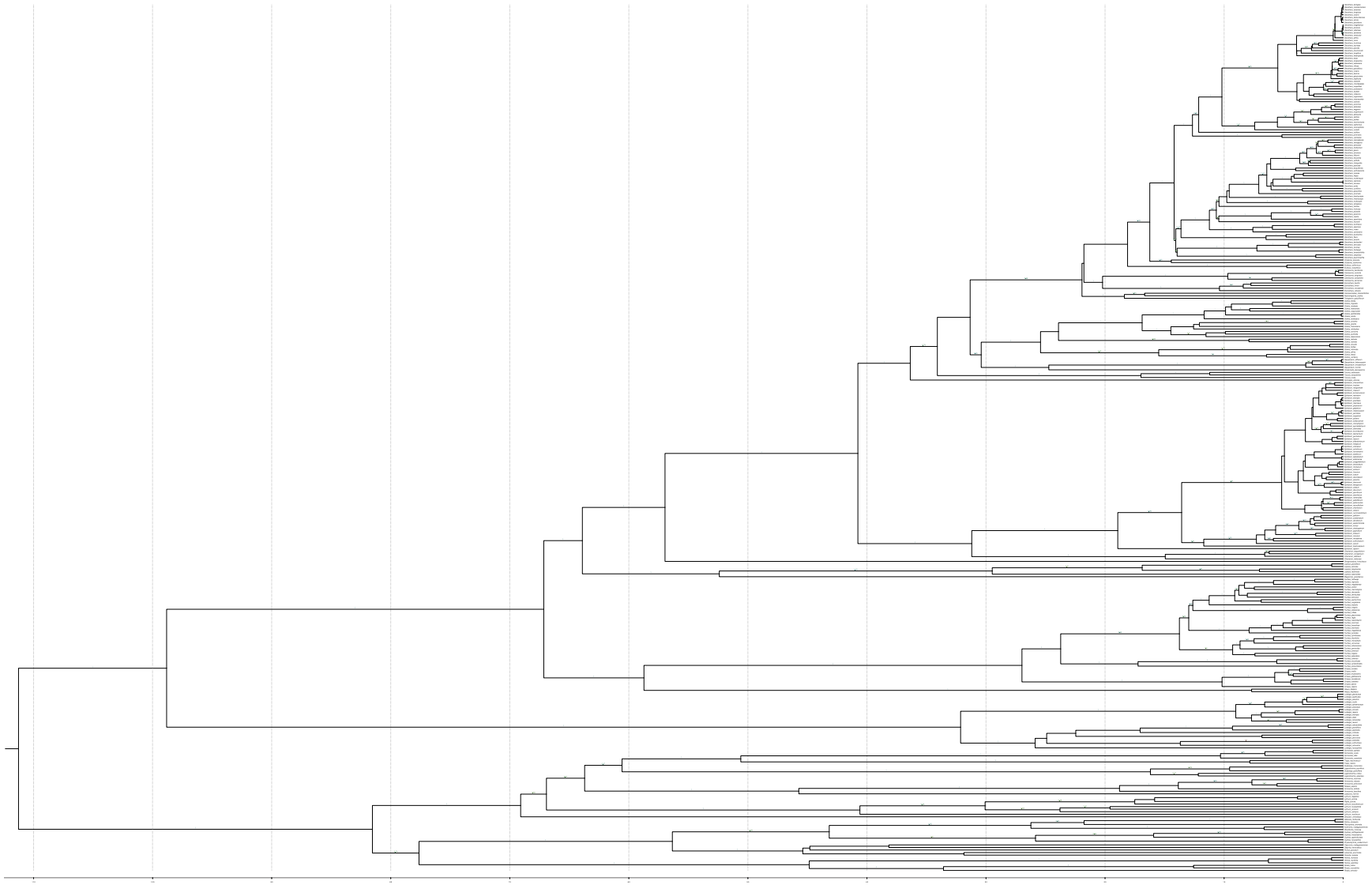

Figure S5: Maximum a posteriori estimate of phylogeny and divergence times of Onagraceae. Divergence times in millions of years are indicated by the axis at the bottom. Posterior probabilities > 0.50 are labeled on the branches.

S3 Mating System Evolution Analyses

S3.1 Model Priors

Model parameter priors are listed in Table S4. The rate of loss of self-incompatibility (q_{ic}), and the rates of switching between hidden states a and b (q_{ab} and q_{ba}) were each given an exponential distribution with a mean of n/Ψ_l, where Ψ_l is the length of the tree Ψ and n is the expected number of transitions. n was given an exponential hyperprior with a mean of 20.

The speciation and extinction rates were drawn from exponential priors with a mean equal to an estimate of the net diversification rate $\hat{d}$. Under a constant rate birth-death process not conditioning on survival of the process, the expected number of lineages at time t is given by:

$$E(N_t) = N_0 e^{td}, \tag{S7}$$

where N_0 is the number of lineages at time 0 and d is the net diversification rate $\lambda - \mu$ (Nee et al. 1994; Höhna 2015). Therefore, we estimate $\hat{d}$ as:

$$\hat{d} = (\ln N_t - \ln N_0)/t, \tag{S8}$$

where N_t is the number of lineages in the clade that survived to the present, t is the age of the root, and $N_0 = 2$. The root state probabilities π were set to start the process equally in either self-incompatible hidden state a or self-incompatible hidden state b.

Table S4: Model parameter names and prior distributions. See the main text for complete description of model parameters and prior distributions. Ψ_l represents the length of tree Ψ and $\hat{d}$ is the expected diversification rate under a constant rate birth-death process.

Parameter	X	$f(X)$
Speciation self-incompatible a	λ_{ia}	Exponential($\lambda = 1/\hat{d}$)
Speciation self-incompatible b	λ_{ib}	Exponential($\lambda = 1/\hat{d}$)
Speciation self-compatible a	λ_{ca}	Exponential($\lambda = 1/\hat{d}$)
Speciation self-compatible b	λ_{cb}	Exponential($\lambda = 1/\hat{d}$)
Extinction self-incompatible a	μ_{ia}	Exponential($\lambda = 1/\hat{d}$)
Extinction self-incompatible b	μ_{ib}	Exponential($\lambda = 1/\hat{d}$)
Extinction self-compatible a	μ_{ca}	Exponential($\lambda = 1/\hat{d}$)
Extinction self-compatible b	μ_{cb}	Exponential($\lambda = 1/\hat{d}$)
Rate of loss of self-incompatibility	q_{ic}	Exponential($\lambda = \Psi_l/n$)
Rate of $a \rightarrow b$	q_{ab}	Exponential($\lambda = \Psi_l/n$)
Rate of $b \rightarrow a$	q_{ba}	Exponential($\lambda = \Psi_l/n$)
Expected number of transitions	n	Exponential($\lambda = 1/20$)

S3.2 MCMC Analyses

To account for uncertainty in phylogeny and divergence times 200 independent MCMC analyses were performed, each sampling a tree from the posterior distribution of trees generated during the phylogenetic analyses. All outgroup (Lythraceae) lineages were pruned off. Each MCMC run drew 10000 samples from the posterior distribution, with 190 randomly scheduled Metropolis-Hastings moves per sample. The first 10% of samples from each run were discarded as burnin. For each run, all parameters had effective sample sizes greater than 200, and the mean effective sample size of the posterior across all 200 tree samples was 1161.6. Estimates of the diversification rates were made by combining samples from all 200 independent runs.

S4 Simulations

S4.1 Simulated Datasets

100 datasets were simulated under a model where the observed binary character was diversification rate independent yet an unobserved binary character drove background diversification rate heterogeneity. First trees were simulated under BiSSE (Maddison et al. 2007) as implemented in the R package Diversitree (FitzJohn 2012). The binary character represented hidden states a and b with diversification rates $\lambda_a = 1.0, \lambda_b = 2.0, \mu_a = 0.4$, and $\mu_b = 0.1$. The rate of change between hidden states a and b was set to $q_{ab} = q_{ba} = 0.1$. This resulted in trees that were qualitatively similar in shape to the empirically estimated Onagraceae tree, with a mix of early diverging depauperate clades and more rapidly radiating recent clades (Figure S6). To simulate incomplete sampling, 55% of the extant tips were randomly pruned off the tree. After pruning, tree samples were discarded unless they had between 100 and 200 sampled lineages that survived to the present. This restriction ensured that the simulated datasets were not too small for reliable inference and yet not so large to be computationally infeasible. Furthermore, we discarded datasets that had fewer than 20% of the tips in either hidden state to ensure that the trees were generated under a sufficiently heterogenous process.

Once the trees were simulated, diversification independent binary characters were simulated over the trees. These characters represented the observed character (mating system) and so were simulated under an irreversible model where the allowed transition occurred with the rate $10/\Psi_s$, where Ψ_s is the length of the simulated tree. This represents an expected 10 irreversible transitions over the length of the tree, and resulted in simulated datasets with a proportion of either state similar to the proportion of self-compatible/self-incompatible in the empirical Onagraceae dataset. These diversification independent characters were then used to calculate Bayes factors that compared the fit of the diversification dependent model to the diversification independent model of mating system. For details on how Bayes factors were calculated see the main text. The false positive error rate was calculated as the percent of simulation replicates in which the Bayes factor supported the false dependent model over the true independent model.

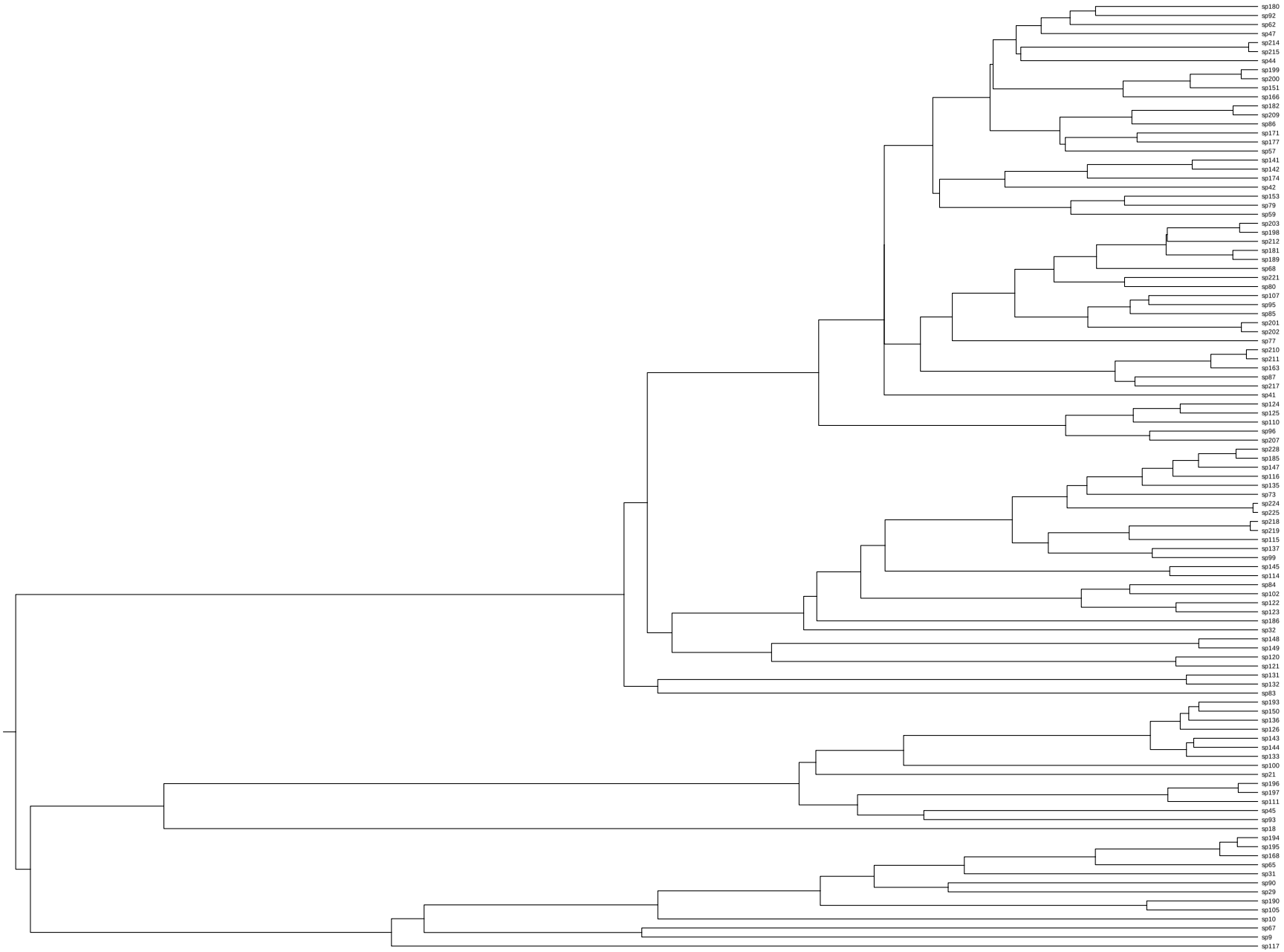

Figure S6: One of the trees simulated under BiSSE used to calculate the false positive error rate. Trees were simulated under a heterogenous diversification process to result in a mix of early diverging depauperate clades and more rapidly radiating recent clades. The tree is shown after 55% of the extant lineages were randomly pruned to replicate incomplete sampling in a reconstructed phylogeny.

References

Beaulieu, J. M. and B. C. OMeara. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Systematic Biology 65:583–601.

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

Freyman, W. A. 2015. SUMAC: Constructing phylogenetic supermatrices and assessing partially decisive taxon coverage. Evolutionary Bioinformatics 11:263.

Freyman, W. A. and S. Höhna. 2017. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Systematic Biology Page in press.

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

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Systematic Biology 60:451–465.

Graham, S. A. 2013. Fossil records in the Lythraceae. The Botanical Review 79:48–145.

Grímsson, F., R. Zetter, and Q. Leng. 2012. Diverse fossil Onagraceae pollen from a Miocene palynoflora of north-east China: early steps in resolving the phytogeographic history of the family. Plant Systematics and Evolution 298:671–687.

Höhna, S. 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. Journal of Theoretical Biology 380:321–331.

Höhna, S., T. A. Heath, B. Boussau, M. J. Landis, F. Ronquist, and J. P. Huelsenbeck. 2014. Probabilistic graphical model representation in phylogenetics. Systematic Biology 63:753–771.

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. Systematic Biology 65:726–736.

Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30:772–780.

Lee, D. E., J. G. Conran, J. M. Bannister, U. Kaulfuss, and D. C. Mildenhall. 2013. A fossil *Fuchsia* (Onagraceae) flower and an anther mass with in situ pollen from the early Miocene of New Zealand. American Journal of Botany 100:2052–2065.

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.

Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society B: Biological Sciences 344:305–311.

Ng, J. and S. D. Smith. 2014. How traits shape trees: new approaches for detecting character state-dependent lineage diversification. Journal of Evolutionary Biology 27:2035–2045.

Nielsen, R. 2002. Mapping mutations on phylogenies. Systematic Biology 51:729–739.

Rodriguez, F., J. Oliver, A. Marin, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142:485–501.

Sytsma, K., A. Litt, and M. Zjhra. 2004. Clades, clocks, and continents: Historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the southern Hemisphere source. International Journal of Plant Sciences 165:S85–S105.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Some Mathematical Questions in Biology—DNA Sequence Analysis, Miura RM (Ed.), American Mathematical Society, Providence (RI) 17:57–86.

Wagner, W. L., P. C. Hoch, and P. H. Raven. 2007. Revised classification of the Onagraceae. Systematic Botany Monographs 83.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of Molecular Evolution 39:306–314.

Yang, Z. and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Molecular Biology and Evolution 14:717–724.

Zhi-Chen, S., W. Wei-Ming, and H. Fei. 2004. Fossil pollen records of extant angiosperms in China. The Botanical Review 70:425–458.