Syst. Biol. 68(3):505–519, 2019

DOI:10.1093/sysbio/syy078
Advance Access publication November 26, 2018

Stochastic Character Mapping of State-Dependent Diversification Reveals the Tempo of Evolutionary Decline in Self-Compatible Onagraceae Lineages

WILLIAM A. FREYMAN[1,*] AND SEBASTIAN HÖHNA[2]

[1]*Department of Integrative Biology, University of California, Berkeley, 3040 Valley Life Sciences Building #3140, CA 94720, USA; and* [2]*Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany*
**Correspondence to be sent to: Department of Integrative Biology, University of California, Berkeley, CA 94720, USA; E-mail: willfreyman@gmail.com.*

Received 16 January 2018; reviews returned 05 November 2018; accepted 13 November 2018

Abstract.—A major goal of evolutionary biology is to identify key evolutionary transitions that correspond with shifts in speciation and extinction rates. Stochastic character mapping has become the primary method used to infer the timing, nature, and number of character state transitions along the branches of a phylogeny. The method is widely employed for standard substitution models of character evolution. However, current approaches cannot be used for models that specifically test the association of character state transitions with shifts in diversification rates such as state-dependent speciation and extinction (SSE) models. Herein, we introduce a new stochastic character mapping algorithm that overcomes these limitations, and apply it to study mating system evolution over a time-calibrated phylogeny of the plant family Onagraceae. Utilizing a hidden state SSE model we tested the association of the loss of self-incompatibility (SI) with shifts in diversification rates. We found that self-compatible lineages have higher extinction rates and lower net-diversification rates compared with self-incompatible lineages. Furthermore, these results provide empirical evidence for the "senescing" diversification rates predicted in highly selfing lineages: our mapped character histories show that the loss of SI 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. Our results demonstrate that stochastic character mapping of SSE models is a powerful tool for examining the timing and nature of both character state transitions and shifts in diversification rates over the phylogeny. [Bayesian inference; Onagraceae; state-dependent speciation and extinction; self-incompatibility; stochastic character mapping.]

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 method frequently used to test hypotheses about 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; Landis et al. 2018), 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 et al. 2009; FitzJohn 2010, 2012; Goldberg and Igić 2012; Magnuson-Ford and Otto 2012; Freyman and Höhna 2018).

In this work, we introduce a method to sample character histories directly from their joint probability 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 models, such as SSE model, 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; Hastings 1970) 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 in the plant family Onagraceae. The majority of flowering plants are hermaphrodites, and the loss of self-incompatibility (SI), the genetic system that encourages outcrossing and prevents self-fertilization, is a common evolutionary transition (Stebbins 1974; Grant 1981; Barrett 2002). Independent transitions to self-compatibility (SC) have occurred repeatedly across the angiosperm phylogeny (Igic et al. 2008) and within Onagraceae (Raven 1979).

Despite the repeated loss of SI, outcrossing is widespread and prevalent in plants, an observation that led Stebbins to hypothesize that selfing was an evolutionary dead-end (Stebbins 1957). Stebbins proposed that over evolutionary time selfing lineages will have higher extinction rates due to reduced genetic variation and an inability to adapt to changing conditions. However, Stebbins also speculated that selfing is maintained by providing a short-term advantage in the form of reproductive assurance. The ability of selfing 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). These short-term advantages led Ho and Agrawal (2017) to propose that primarily selfing lineages may at first diversify at higher rates than primarily outcrossing lineages but over time slow down due to elevated extinction rates.

Recent studies have reported higher net-diversification rates for SI lineages, supporting Stebbins' dead-end hypothesis (Ferrer and Good 2012; Goldberg et al. 2010; de Vos et al. 2014; Gamisch et al. 2015). Explicit phylogenetic tests for increased extinction rates in SC lineages have been performed in Solanceae (Goldberg et al. 2010), Primulaceae (de Vos et al. 2014), and Orchidaceae (Gamisch et al. 2015), and all of these studies reported lower overall rates of net-diversification in SC lineages compared with SI lineages. In these studies, 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 O'Meara 2016; Caetano et al. 2018). 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 a fossil-calibrated phylogeny of Onagraceae. We test the hypothesis that SC lineages have higher extinction and speciation rates yet lower net-diversification rates compared with SI lineages, and investigate the short-term versus long-term macroevolutionary consequences of the loss of SI.

Materials and Methods

Stochastic Character Mapping Method

Figure 1 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 (Fig. 1a; 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 (Fig. 1b). Finally, character histories are repeatedly simulated using rejection sampling for each branch of the tree (Fig. 1c).

A detailed pseudocode formulation of our new stochastic character mapping algorithm is provided in Algorithm 1. In this 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 differential equations similar to Maddison et al. (2007). We numerically integrate these equations for every arbitrarily small time interval along each branch, however, unlike Maddison et al. (2007), we store a vector of conditional likelihoods for the character being in each state for every small time interval (Fig. 1e). Letting $\mathcal{X}$ represent the observed tip data, Ψ an observed phylogeny, and θ_q a particular set of character evolution model parameters, the likelihood at the root of the tree is then given by:

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

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 $\mathcal{X}$ (Freyman and Höhna 2018).

We then sample a complete character history during a pre-order tree traversal. First, the root state is drawn from probabilities proportional to the marginal likelihood of each state at the root $\left(p_i = \frac{\pi_i D_{R,i}(t)}{\sum_i \pi_i D_{R,i}(t)}\right)$. Then, states are drawn for each small time interval moving toward the tip of the tree conditioned on the state of the previous small time interval (Fig. 1f). To compute the probability of a state at the end of each small time interval, we integrate numerically over a set of differential equations during this root-to-tip tree traversal (see Fig. 1f). This integration, however, is performed in forward-time, thus a different and new set of differential equations must be used (defined below). When sampling a state at the end of each small time interval, the forward-time probabilities are weighted by the backward-time conditional likelihoods computed and stored for this same time interval during the earlier post-order tree

FIGURE 1. Schematic and 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. Here, the vector $[l_{n0}, l_{n1}, l_{n2}]$ represents the conditional likelihoods of the process at node n in states 0, 1, and 2, and the vector $[l_{ni0}, l_{ni1}, l_{ni2}]$ represents the conditional likelihoods of the process in the small time interval i along the branch leading to node n. Next, during a pre-order traversal ancestral states are sampled for each time interval (f). The grey dashed loop represents the forward-time equations (Equations 4 and 6) conditioning on a state sampled during each small time interval. The result is a full character history (g) without the need for a rejection sampling step. See the main text for more details.

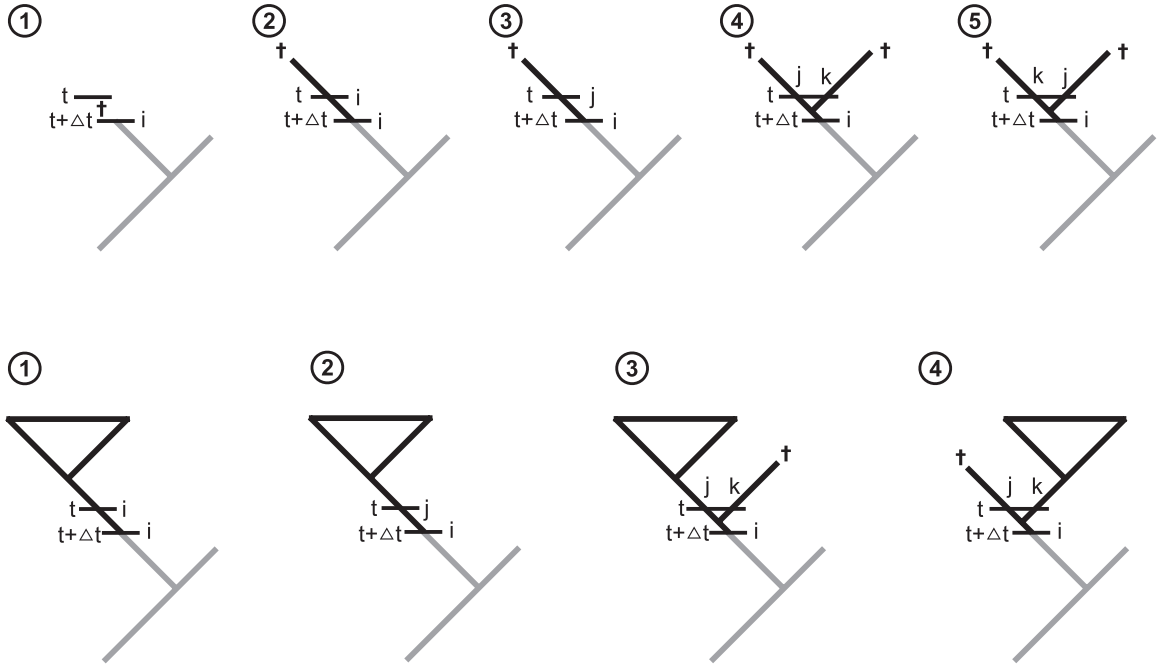

FIGURE 2. Alternative scenarios of events in a small time interval Δt looking backwards in time. The top row shows the different scenarios for a lineage that goes extinct before the present. Case 1: The lineage goes extinct in the time interval Δt. Case 2: There is no event in the time interval Δt and 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.

traversal. With this approach, we can directly sample character histories from a SSE process in forward-time, resulting in a complete stochastic character map sample without the need for rejection sampling or uniformization (see Fig. 1).

Derivation of Our Differential Equations

The two functions we integrate numerically 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.

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.

Differential Equations Backwards in Time. The original derivation of the differential equations for the SSE process are defined backward in time (Maddison et al. 2007). Here, we use a generalization of the SSE process to allow for cladogenetic events where daughter lineages may inherit different states, as derived by Goldberg and Igić (2012), see also (Magnuson-Ford and Otto 2012) and 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 2. 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 Fig. 2): (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,

Algorithm 1 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.

1. **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.
2. **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
   $D_{N,i}(t=0) \leftarrow 0$
3. **while** $t \leq t_r$ **do** // post-order tree traversal
4. **if** node L is reached **then**
5. $D_{L,i}(t) \leftarrow \sum_j \sum_k \lambda_{ijk} D_{M,j}(t) D_{N,k}(t)$ // combine descendant probabilities
6. **else**
7. $L_{N,i}(t) \leftarrow D_{N,i}(t)$ // store the conditional likelihoods for this time interval
8. $E_i(t+\Delta t) \leftarrow E_i(t) +$ // 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 (1)
9. $D_{N,i}(t+\Delta t) \leftarrow D_{N,i}(t) +$
$$\left[-\left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i\right) D_{N,i}(t) + \sum_{j \neq i} Q_{ij} D_{N,j}(t) + \sum_j \sum_k \lambda_{ijk} \left(D_{N,k}(t) E_j(t) + D_{N,j}(t) E_k(t)\right)\right] \Delta t$$ // backward-time Equation (2)
10. $t \leftarrow t + \Delta t$ // increment the current t
11. **end if**
12. **end while**
13. $s_t \sim \text{Multinomial}(n=1, D_N(t_r) \times \pi)$ // draw character state at the root
14. **while** $t \geq 0$ **do** // pre-order tree traversal
15. $D_{N,s_t} \leftarrow 1$ // condition on the sampled character state
16. $D_{N,i \neq s_t} \leftarrow 0$
17. $E_i(t-\Delta t) \leftarrow E_i(t) -$
$$\left[\mu_i - \left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i\right) E_i(t) + \sum_{j \neq i} Q_{ij} E_j(t) + \sum_j \sum_k \lambda_{ijk} E_j(t) E_k(t)\right] \Delta t$$ // forward-time Equation (4)
18. $D_{N,i}(t-\Delta t) \leftarrow D_{N,i}(t) +$
$$\left[-\left(\sum_j \sum_k \lambda_{ijk} + \sum_{j \neq i} Q_{ij} + \mu_i\right) D_{N,i} + \sum_{j \neq i} Q_{ji} D_{N,j}(t) + \sum_j \sum_k \lambda_{jik} D_{N,j}(t) E_k(t) + \sum_j \sum_k \lambda_{jki} D_{N,j}(t) E_k(t)\right]$$ // forward-time Equation (6)
19. $t \leftarrow t - \Delta t$ // decrement the current t
20. $s_t \sim \text{Multinomial}(n=1, D_N(t) \times L_N(t))$ // draw character state for time t
21. **end while**
22. **return** vector of all sampled character states s

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. With this description of

all possible scenarios we can derive the differential equation.

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

Similarly, we can consider all possible scenarios for an observed lineage. We have four different cases (bottom row in Fig. 2): (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 $D(t)$.

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

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 Fig. 2): (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

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

Unfortunately, we cannot solve Equation (3) directly because we do not know $E_j(t-\Delta t)$ and $E_k(t-\Delta t)$. Instead, we will approximate Equation (3) 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

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

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 Fig. 2), 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.

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

As before, we cannot solve Equation (5) 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.

$$
\begin{aligned}
&D_{N,i}(t-\Delta t)\\
&\approx D_{N,i}(t) + \qquad (6)\\
&\Bigg[-\left(\sum_j\sum_k \lambda_{ijk} + \sum_{j\neq i} Q_{ij} + \mu_i\right) D_{N,i}(t) && \text{Case (1)}\\
&+ \sum_{j\neq i} Q_{ji} D_{N,j}(t) && \text{Case (2)}\\
&+ \sum_j\sum_k \lambda_{jik} D_{N,j}(t) E_k(t)\\
&+ \sum_j\sum_k \lambda_{jki} D_{N,j}(t) E_k(t) \Bigg] \Delta t && \text{Case (3) and (4)}
\end{aligned}
$$

To sample character histories from an SSE process in forward-time during Algorithm (1) we calculate $E(t-\Delta t)$ using the approximation given by Equation (4) and $D(t-\Delta t)$ using Equation (6).

Correctness of the Forward Time Equations

Validation of the Forward Time Extinction Probabilities. 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 (4) 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 "Validation of the forward time equations against `diversitree`" section.

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 3 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.

Validation of the Forward Time Equations against `diversitree`. Second, we validate our method of sampling character histories from an SSE process

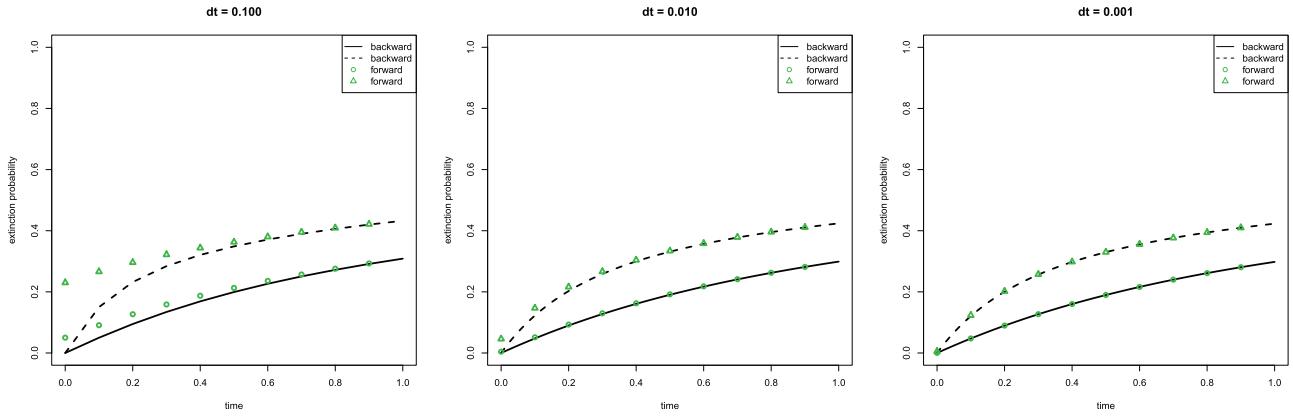

FIGURE 3. The probability of extinction computed backward and forward in time. Here, we compute the extinction probabilities $E_0(t)$ and $E_1(t)$ for a BiSSE model backward and forward in time. Details about the parameters of the BiSSE model are given in the text. We varied the step-size Δt for the numerical integration between 0.1, 0.01, and 0.001 to show that both computations give the same probabilities once Δt is small enough.

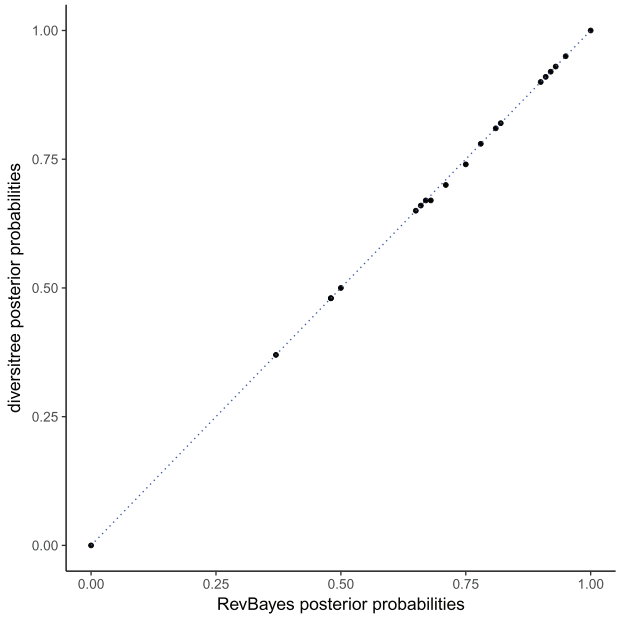

FIGURE 4. Comparing marginal posterior ancestral state estimates from `diversitree` to those calculated in `RevBayes`. Each point represents the posterior probability of a given node having the ancestral state 0. On the *y*-axis are the posterior probabilities as analytically calculated by `diversitree`. On the *x*-axis are the posterior probabilities as calculated by `RevBayes` using Algorithm (1). Our approximation given in Equation (6) yields highly similar 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.

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` cannot 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 O'Meara 2016), ChromoSSE (Freyman and Höhna 2018), 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 10,000 samples drawn by our Monte Carlo method. Figure 4 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 4 shows that using the approximation of $E(t - \Delta t)$ given by Equation (4) and the approximation to compute $D(t - \Delta t)$ in Equation (6) during Algorithm (1) 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.

Implementation, MCMC Sampling and Computation Efficiency

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.

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 are 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 (1) 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.

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). In total, eight gene regions were used (seven chloroplast loci plus the nuclear ribosomal internal transcribed spacer region) representing a total of 340 taxa (292 Onagraceae taxa and 48 Lythraceae taxa). Information about the alignments and GenBank accessions used can be found in the Supplementary Section S1.1 available on Dryad at http://dx.doi.org/10.5061/dryad.9jj428f. Phylogeny and divergence times were inferred using RevBayes (Höhna et al. 2016). Node ages were calibrated using five fossil calibrations and one secondary calibration. Details regarding the calibrations, the models of molecular evolution, and MCMC analyses are given in the Supplementary Section S1.1 available on Dryad.

Analyses of Mating System Evolution

The mating systems of Onagraceae species were scored as either SC or SI following (Wagner et al. 2007). Most of the SC/SI assignments in Wagner et al. (2007) come from detailed family-level surveys such as Raven (1979), in which the outcrossing/selfing modes of 283 Onagraceae species were examined, and Heslop-Harrison (1990), in which compatibility tests of 48 Onagraceae species were performed. Other SC/SI assignments come from the many genus- and section-level studies cited in Wagner et al. (2007) such as Lewis and Lewis (1955), Plitmann et al. (1973), and Seavey et al. (1977).

For the analysis of mating system evolution, all outgroup (Lythraceae) lineages were pruned off our phylogeny, leaving 292 Onagraceae species. The species sampling fraction of extant Onagraceae species was thus $\rho = 292/650 = 0.45$, which is the number of species sampled divided by the approximate total number of Onagraceae species reported in Wagner et al. (2007). We use this sampling fraction of extant Onagraceae as the uniform taxon sampling probability ρ, assuming that missing species are uniformly distributed over the phylogeny (Nee et al. 1994; Yang and Rannala 1997; Höhna et al. 2011; Höhna 2014). We assumed there was no state-dependent sampling bias since we lacked complete SC/SI assignments for all Onagraceae taxa that would indicate such a bias. Finally, we accounted for uncertainty in the phylogeny and divergence times by sampling 200 trees from the posterior distribution of trees.

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 four states that describes the joint evolution of mating system as well as an unobserved character with hidden states *a* and *b* (Fig. 5). For each of the four states we estimated speciation (λ) and extinction (μ) rates. For details on priors used and the MCMC analyses see the Supplementary Section S2.1 available on Dryad.

The system of SI found in Onagraceae is S-RNase-based gametophytic SI (Raven 1979; Franklin et al. 1995; Igic et al. 2008). This system of SI evolved once in the common ancestor of the Asteridae and Rosidae (Steinbachs and Holsinger 2002; Igic et al. 2008; Vieira et al. 2008; Niu et al. 2017; Ramanauskas and Igić 2017), the clade that contains Onagraceae. Raven (1979) writes the system of SI in Onagraceae "is gametophytic, and involves a series of S-a1leles, with inhibition of pollen-tube growth normally in the surface layers of the stigma." Furthermore, Raven writes that SI "seems to have been characteristic of the original common ancestor of Onagraceae, judged by the occurrence of SI in four of the seven tribes of the family. There is no evidence for the evolution of SI within the family once it has been lost." Following Raven (1979), we used an irreversible model that only allowed transitions from SI to SC. However, to test the assumption of irreversibility on our results, we additionally used a model that allowed for the possibility of secondary gains of SI by permitting both transitions from SI to SC and transitions from SC to SI (see Supplementary Section S2.3 available on Dryad).

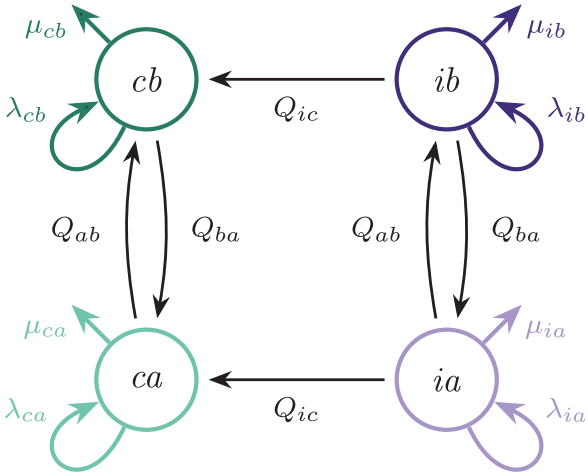

FIGURE 5. SSE model depicting states and rate parameters used to infer mating system evolution. The states are labeled *ca*, *cb*, *ia*, and *ib*, representing self-compatible hidden state *a*, self-compatible hidden state *b*, self-incompatible hidden state *a*, and self-incompatible hidden state *b*, respectively. Independent extinction μ and speciation λ rates were estimated for each of the four 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}.

Model Comparisons, Incomplete Sampling, 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 four 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 and the effect of incomplete taxon sampling, 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. To test the effect of missing data on our power to detect state-dependent diversification, the simulated data sets were pruned to have the same proportion of taxon sampling as our empirical Onagraceae data set (45%). 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 Supplementary Section S3.1 available on Dryad.

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 19,000 generations within each step. The Bayes factor was then calculated as twice the difference in the natural log marginal likelihoods (Kass and Raftery 1995).

RESULTS

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. 6), and a summary of the divergence times of major clades within Onagraceae can be found in Supplementary Table S3 available on Dryad.

Stochastic Character Maps

Since the results from the analysis allowing for secondary gains of SI were essentially identical to the results from the analysis that assumed irreversibility and disallowed secondary gains of SI, we report here only the results from the irreversible analysis. See Supplementary Section S2.3 available on Dryad for results of the analysis allowing for secondary gains.

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. 6). 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. 7). Following the loss of SI, there was an evolutionary time lag (mean 1.97 My) until net-diversification (speciation minus extinction) turned negative (Fig. 8). Since SC hidden state *b* was estimated to have positive net-diversification and SC hidden state *a* was estimated to have negative net-diversification, we calculated the time lag from the loss of SI until an evolutionary decline as the time spent following the loss of SI in hidden state *b* until transitioning to hidden state *a*. In many cases, the loss of SI occurred in an ancestral lineage with positive net-diversification (hidden state *b*) followed by multiple shifts to negative net-diversification (hidden state *a*) in descendant lineages. To account for these non-independent time lags and avoid double counting the time the ancestral lineage spent in SC state *b*, we average over all (partially) dependent events. For example, the ancestral lineage spent time t_a in state *b* and the left and right descendant lineages spent time t_l and t_r in state *b* before switching to state *a*, respectively. Then, we counted the time as $t = t_a + \frac{t_l + t_b}{2}$.

Diversification Rate Estimates

Within either hidden state (*a* or *b*) SC lineages had generally higher speciation and extinction rates compared with SI lineages (Table 1 and Fig. 6). Despite higher speciation and extinction rates, SC lineages had lower net-diversification compared with SI lineages.

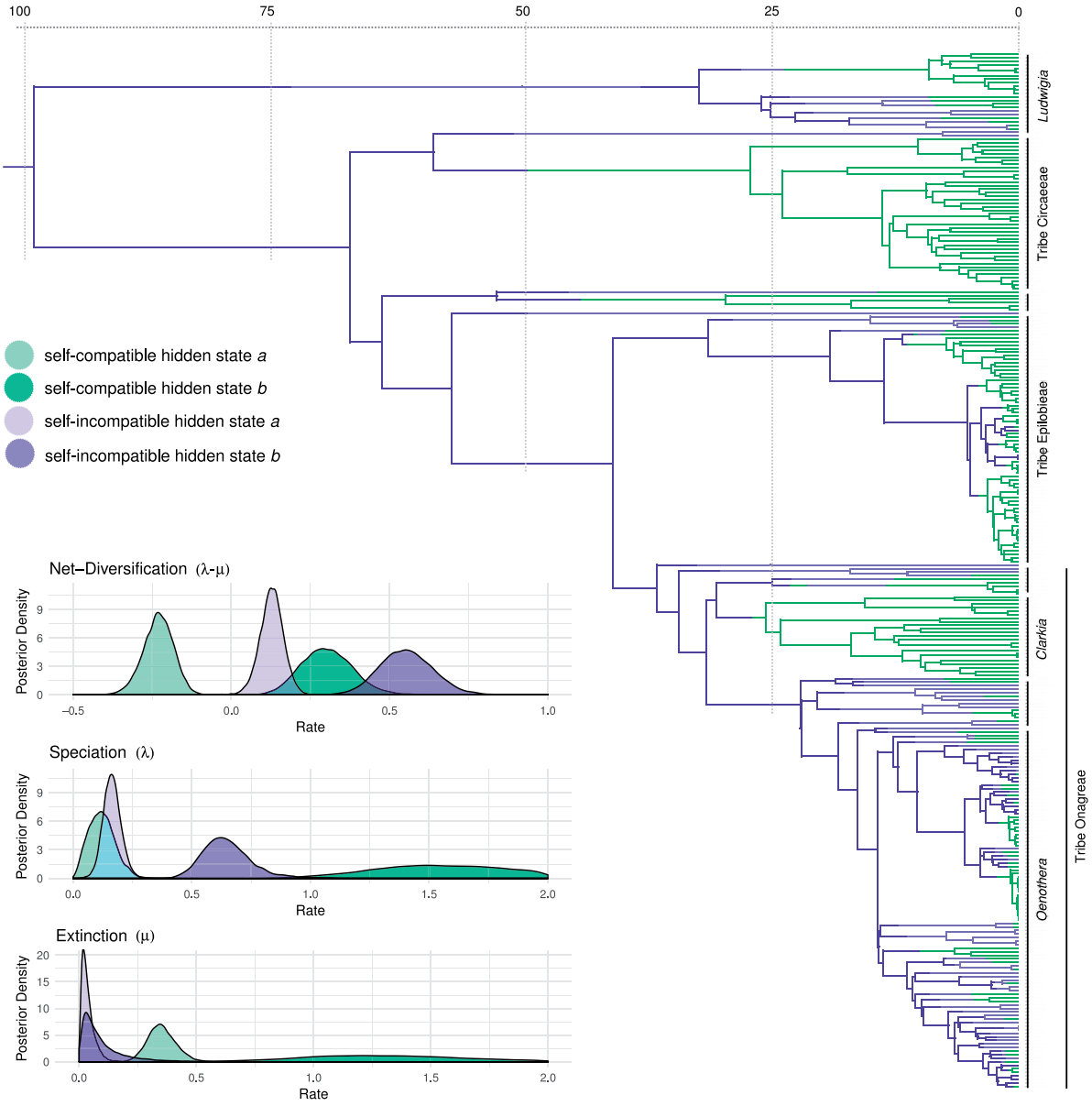

FIGURE 6. Maximum a posteriori reconstruction of mating system evolution and shifts in diversification rates in Onagraceae. Divergence times in millions of years are indicated by the axis at the top. Note that in this marginal summary reconstruction some transitions are displayed such as the loss and 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. 7). 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 with self-incompatible lineages. (The color version of this figure is available online.)

Net-diversification was found to be negative for most but not all extant SC lineages.

Model Comparisons, Incomplete Sampling, and Error Rates. For the Onagraceae data set, the state-dependent diversification model of mating system evolution (Fig. 5) 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 data sets showed that the false positive error rate was low even despite the poor taxon sampling present in our empirical data set (Fig. 9). The false positive rate for "strong" support ($2ln\text{BF} > 6$; Kass and Raftery 1995) was

TABLE 1. Posterior parameter estimates of the HiSSE mating system evolution model depicted in Figure 5

Parameters	X	Mating system	Hidden state	Mean	95% HPD interval
Speciation	λ_{ca}	SC	a	0.12	0.02–0.23
	λ_{ia}	SI	a	0.16	0.09–0.24
	λ_{cb}	SC	b	1.66	0.98–2.41
	λ_{ib}	SI	b	0.65	0.45–0.85
Extinction	μ_{ca}	SC	a	0.35	0.25–0.48
	μ_{ia}	SI	a	0.04	0.00–0.09
	μ_{cb}	SC	b	1.36	0.65–2.19
	μ_{ib}	SI	b	0.10	0.00–0.29
Net-diversification	r_{ca}	SC	a	−0.23	−0.32 to −0.14
	r_{ia}	SI	a	0.13	0.05–0.19
	r_{cb}	SC	b	0.30	0.15–0.46
	r_{ib}	SI	b	0.55	0.39–0.71
Transition	Q_{ic}	SI → SC	a/b	0.22	0.16–0.28
	Q_{ab}	SI/SC	$a \rightarrow b$	0.01	0.00–0.02
	Q_{ba}	SI/SC	$b \rightarrow a$	0.33	0.16–0.52

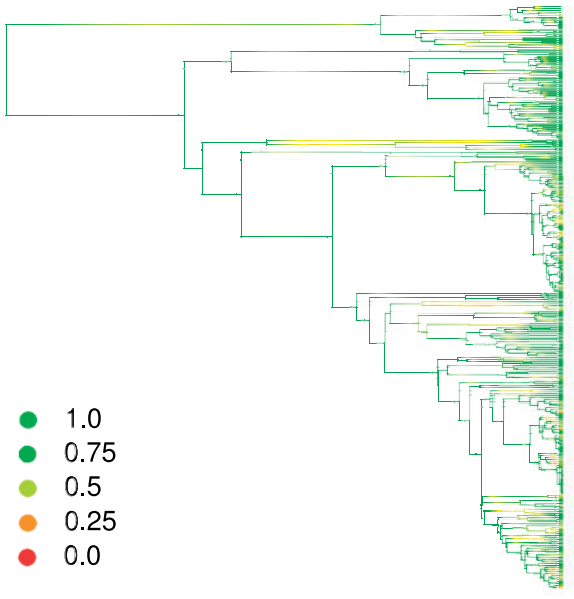

FIGURE 7. Posterior probabilities of the maximum a posteriori reconstruction of mating system evolution and shifts in diversification rates in Onagraceae. Marginal posterior probabilities of the character states shown in Figure 6. Uncertainty was highest along long branches where there was relatively little information regarding the timing of transitions. (The color version of this figure is available online.)

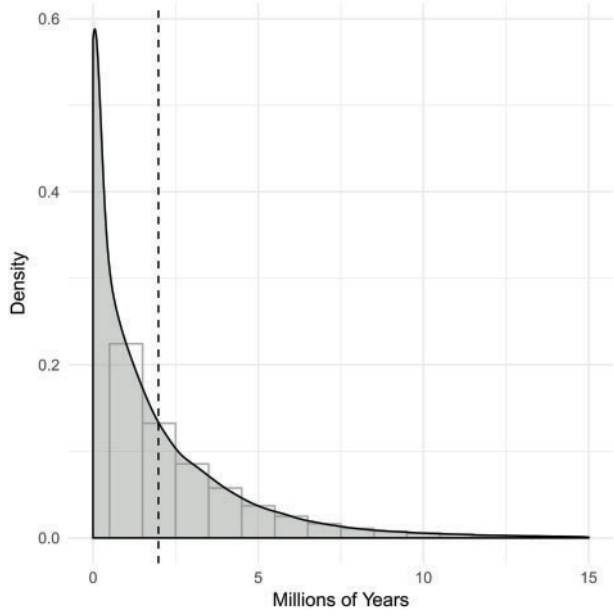

FIGURE 8. The time lag from the loss of self-incompatibility until the onset of evolutionary decline. The time in millions of years after the loss of self-incompatibility until the net-diversification rate became negative measured over 10,000 stochastic character map samples. The mean time lag until evolutionary decline was 1.97 million years (indicated by a dashed line).

0.05, and the false positive rate for "very strong" support ($2lnBF > 10$; Kass and Raftery 1995) was 0.0.

Discussion and Conclusion

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 a negative net-diversification rate. The time lag until this evolutionary decline was measured as the time spent following the loss of SI in hidden state *b* (positive net-diversification) until transitioning to hidden state *a* (negative net-diversification). By mapping the time spent in each hidden state, the stochastic character maps quantified the speed of the evolutionary decline in

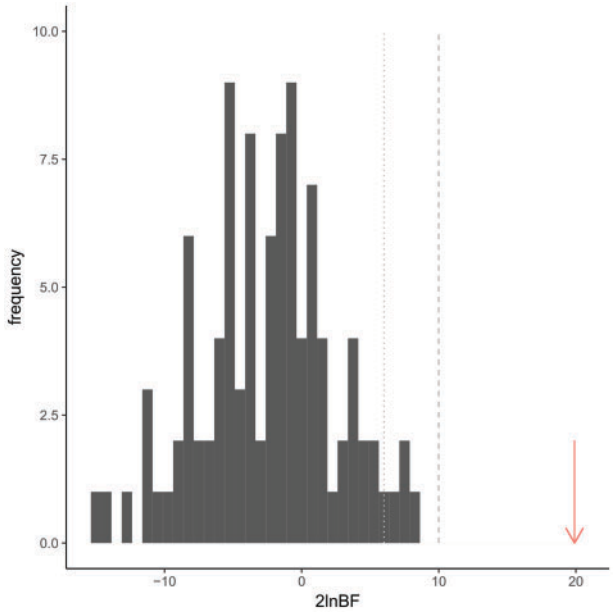

FIGURE 9. Bayes factors ($2lnBF$) comparing the fit of the state-dependent diversification model of mating system evolution with the state-independent diversification model. The red arrow indicates the "decisive" support found for the empirical Onagraceae data ($2lnBF = 19.9$; Jeffreys 1961). The dark grey bars represent Bayes factors calculated for 100 data sets simulated under a state-independent diversification model and pruned to have the same proportion of missing species as the empirical Onagraceae data set. The dotted light grey line indicates "strong" support ($2lnBF > 6$; Kass and Raftery 1995), and the dashed light grey line indicates "very strong" support ($2lnBF > 10$; Kass and Raftery 1995). Even with the poor taxon sampling present in the Onagraceae data set, our power to reject false positives was high.

SC lineages. These results are robust to phylogenetic uncertainty (by averaging over a posterior distribution of trees), to assumptions of mating system irreversiblity (Supplementary material available on Dryad for results allowing for secondary gains of SI), and to the effect of missing species sampling (false positive error rate calculated using simulations).

While the mean time from the loss of SI until evolutionary decline was 1.97 million years, there was a large amount of variation in time estimates (Fig. 8). This variation could be due to differences in the realized selfing/outcrossing rates of different SC 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 (Lynch et al. 1995a,b; Wright et al. 2008). 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 the potential for extinction. SC lineages with high outcrossing rates and less inbreeding, on the other hand, likely have larger effective population sizes and lower genetic load (Wright et al. 2008), thus delaying the onset of higher extinction rates. A limitation of our analysis is that our species data was coded simply SI or SC, a more nuanced exploration of the macroevolutionary impact of selfing would require hard to measure selfing/outcrossing rates from a large number of species across the phylogeny.

These results confirm theory about the macroevolutionary consequences of selfing (Stebbins 1957; Grant 1981). These consequences include the increased probability of going extinct due to the accumulation of harmful mutations (Lynch et al. 1995a,b; Wright et al. 2008) 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 in Onagraceae 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 Onagraceae lineages undergo rapid bursts of increased speciation that eventually decline, doomed by intensified extinction and thus supporting Stebbins' hypothesis of selfing as an evolutionary dead-end (Stebbins 1957). These results provide 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. Similar results were previously found in Primulaceae by de Vos et al. (2014), where SC non-heterostylous lineages were found to "live fast and die young" compared with SI heterostylous lineages.

Our findings corroborate previous analyses performed in the plant families Solanceae (Goldberg et al. 2010), Primulaceae (de Vos et al. 2014), and Orchidaceae (Gamisch et al. 2015) where SC lineages were also found to have lower net-diversification rates than SI lineages. Our results, however, are the first to use a HiSSE model 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). Interpretating the hidden states of an SSE model can be challenging (Caetano et al. 2018). Depending on the diversification rates estimated, there

were different but equally valid ways to make sense of the hidden states in our analysis: (1) if the diversification rates varied between SC and SI, but not between hidden states *a* and *b*, we could conclude that shifts in mating system explained all diversification rate heterogeneity; (2) if the diversification rates did not vary between SC and SI, but did vary between hidden states *a* and *b* we could conclude that there were background rate changes unassociated with mating system and that mating system evolution was not associated with rate shifts; or (3) if the diversification rates varied both between SC/SI and between hidden states *a*/*b*, then depending on the phylogenetic pattern of the hidden states they could represent the different long- and short-term consequences of the loss of SI. Our results are congruent with this last interpretation, and we interpret the phylogenetic pattern of the hidden states to represent the temporal decay of diversification rates in SC lineages. It is important to note, however, that our HiSSE-based analysis allowed for any of those three outcomes unlike BiSSE-based analyses.

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 SI 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 stochastic character mapping is used with an SSE model in which some or even all states are hidden (no observed states), then 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. 2018) 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.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.9jj428f.

Funding

This work was supported by the National Science Foundation [GRFP DGE 1106400, DDIG DEB 1601402, and DEB 1655478 to W.A.F.]. Computations were performed on the Savio computational cluster provided by the Berkeley Research Computing program at the University of California, Berkeley.

Acknowledgments

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.

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. Nat. Rev. Genet. 3:274–284.
Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.
Bollback J.P. 2006. Simmap: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.
Caetano D., O'Meara B., Beaulieu, J. 2018. Hidden state models improve the adequacy of state-dependent diversification approaches using empirical trees, including biogeographical models. bioRxiv 302729.
Darwin C. 1876. The effects of cross and self fertilization in the vegetable kingdom. NY: Appleton. 482 pp.
de Vos J.M., Hughes C.E., Schneeweiss G.M., Moore B.R., Conti E. 2014. Heterostyly accelerates diversification via reduced extinction in primroses. Proc. R. Soc. B 281:20140075.
Donoghue M.J. 2005. Key innovations, convergence, and success: macroevolutionary lessons from plant phylogeny. Paleobiology. 31:77–93.
Eriksson O., Bremer B. 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. J. Mol. Evol. 17:368–376.
Ferrer M.M., Good S.V. 2012. Self-sterility in flowering plants: preventing self-fertilization increases family diversification rates. Ann. Bot. 110(3):535–553.
FitzJohn R., Maddison W., Otto S. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595–611.
FitzJohn R.G. 2010. Quantitative traits and diversification. Syst. Biol. 59:619–633.
FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.
Franklin F., Lawrence M., Franklin-Tong V. 1995. Cell and molecular biology of self-incompatibility in flowering plants. Int. Rev Cytol 158:1–64.
Freyman W.A. 2015. SUMAC: Constructing phylogenetic supermatrices and assessing partially decisive taxon coverage. Evol. Bioinform. Online. 11:263–266.
Freyman W.A., Höhna S. 2018. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Syst. Biol. 67:195–215.
Gamisch A., Fischer G.A., Comes H.P. 2015. Multiple independent origins of auto-pollination in tropical orchids (bulbophyllum) in light of the hypothesis of selfing as an evolutionary dead end. BMC Evol. Biol. 15:192.
Goldberg E.E., Igić B. 2012. Tempo and mode in plant breeding system evolution. Evolution. 66:3701–3709.
Goldberg E.E., Kohn J.R., Lande R., Robertson K.A., Smith S.A., Igić B. 2010. Species selection maintains self-incompatibility. Science. 330:493–495.
Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Syst. Biol. 60:451–465.
Grant, V. 1981. Plant Speciation. New York: Columbia University Press.
Hartfield M. 2016. Evolutionary genetic consequences of facultative sex and outcrossing. J. Evol. Biol. 29:5–22.
Hastings W.K. 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika. 57:97–109.

Heslop-Harrison Y. 1990. Stigma form and surface in relation to self-incompatibility in the Onagraceae. Nordic J. Bot. 10:1–19.
Ho E.K., Agrawal A.F. 2017. Aging asexual lineages and the evolutionary maintenance of sex. Evolution 71(7):1865–1875.
Hobolth A., Stone E.A. 2009. Efficient simulation from finite-state, continuous-time Markov chains with incomplete observations. Ann. Appl. Stat. 3:1204–1231.
Höhna S. 2014. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. PLoS One. 9:e84184.
Höhna S., Coghill L.M., Mount G.G., Thomson R.C., Brown J.M. 2018. P^3: Phylogenetic posterior prediction in RevBayes. Mol. Biol. Evol. 35:1028–1034.
Höhna S., Heath T.A., Boussau B., Landis M.J., Ronquist F., Huelsenbeck J.P. 2014. Probabilistic graphical model representation in phylogenetics. Syst. Biol. 63:753–771.
Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65:726–736.
Höhna S., Landis M.L., Huelsenbeck J.P. 2017. Parallel power posterior analyses for fast computation of marginal likelihoods in phylogenetics. bioRxiv 104422.
Höhna S., Stadler T., Ronquist F., Britton T. 2011. Inferring speciation and extinction rates under different species sampling schemes. Mol. Biol. Evol. 28:2577–2589.
Huelsenbeck J.P., Nielsen R., Bollback J.P. 2003. Stochastic mapping of morphological characters. Syst. Biol. 52:131–158.
Igic B., Lande R., Kohn J.R. 2008. Loss of self-incompatibility and its evolutionary consequences. Int. J. Plant Sci. 169:93–104.
Irvahn J., Minin V.N. 2014. Phylogenetic stochastic mapping without matrix exponentiation. J. Comput. Biol. 21:676–690.
Jeffreys H. 1961. Theory of probability. 3rd ed. Oxford, UK: Oxford University Press.
Johnson M.T., FitzJohn R.G., Smith S.D., Rausher M.D., Otto S.P. 2011. Loss of sexual recombination and segregation is associated with increased diversification in evening primroses. Evolution. 65:3230–3240.
Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.
Lande R., Schemske D.W. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. Evolution. 39:24–40.
Landis M.J., Freyman W.A., Baldwin B.G. 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. Evolution 72(11):2343–2359.
Lewis H., Lewis M.E. 1955. The genus *Clarkia*. University of California Publications in Botany 20:241–392.
Lynch M., Conery J., Burger R. 1995a. Mutation accumulation and the extinction of small populations. Am. Nat. 146:489–518.
Lynch M., Conery J., Burger R. 1995b. Mutational meltdowns in sexual populations. Evolution. 49:1067–1080.
Maddison W.P., FitzJohn R.G. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64:127–136.
Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.
Magnuson-Ford K., Otto S.P. 2012. Linking the investigations of character evolution and species diversification. Am. Nat. 180:225–245.
Mayrose I., Zhan S.H., Rothfels C.J., Magnuson-Ford K., Barker M.S., Rieseberg L.H., Otto S.P. 2011. Recently formed polyploid plants diversify at lower rates. Science. 333:1257–1257.
Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092.
Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. Biol. Sci. 344:305–311.
Ng J. Smith S.D. 2014. How traits shape trees: new approaches for detecting character state-dependent lineage diversification. J. Evol. Biol. 27:2035–2045.
Nielsen R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51:729–739.
Niu S.-C., Huang J., Zhang Y.-Q., Li P.-X., Zhang G.-Q., Xu Q., Chen L.-J., Wang J.-Y., Luo Y.-B., Liu Z.-J. 2017. Lack of S-RNase-based gametophytic self-incompatibility in orchids suggests that this system evolved after the monocot-eudicot split. Front. Plant Sci. 8:1106.
Plitmann U., Raven P.H., Breedlove D.E. 1973. The systematics of Lopezieae (Onagraceae). Ann. Mo. Bot. Gard. 60:478–563.
Rabosky D.L., Goldberg E.E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64:340–355.
Ramanauskas K., Igić B. 2017. The evolutionary history of plant T2/S-type ribonucleases. PeerJ. 5:e3790.
Raven P.H. 1979. A survey of reproductive biology in Onagraceae. New Zeal. J. Bot. 17:575–593.
Rodrigue N., Philippe H., Lartillot N. 2008. Uniformization for sampling realizations of Markov processes: applications to Bayesian implementations of codon substitution models. Bioinformatics. 24:56–62.
Sanderson M.J., Donoghue M.J. 1996. Reconstructing shifts in diversification rates on phylogenetic trees. Trends Ecol. Evol. 11:15–20.
Seavey S.R., Wright P., Raven P.H. 1977. A comparison of Epilobium minutum and E. foliosum (Onagraceae). Madroño. 24:6–12.
Stebbins G.L. 1957. Self fertilization and population variability in the higher plants. Am. Nat. 91:337–354.
Stebbins G.L. 1974. Flowering plants: evolution above the species level. Cambridge, MA: Belknap Press.
Steinbachs J., Holsinger K. 2002. S-rnase–mediated gametophytic self-incompatibility is ancestral in eudicots. Mol. Biol. Evol. 19:825–829.
Sytsma K., Litt A., Zjhra M. 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., Smith J.M. 1995. The major evolutionary transitions. Nature. 374:227.
Van Dyk D.A., Meng X.-L. 2001. The art of data augmentation. J. Comput. Graph. Stat. 10:1–50.
Vieira J., Fonseca N.A., Vieira C.P. 2008. An S-RNase-based gametophytic self-incompatibility system evolved only once in eudicots. J. Mol. Evol. 67:179–190.
Wagner W.L., Hoch P.C., Raven P.H. 2007. Revised classification of the Onagraceae. Systematic Botany Monographs 83:1–240.
Wright S.I., Ness R.W., Foxe J.P., Barrett S.C. 2008. Genomic consequences of outcrossing and selfing in plants. Int. J. Plant Sci. 169:105–118.
Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.-H. 2010. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.
Yang Z., Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo Method. Mol. Biol. Evol. 14:717–724.