Syst. Biol. 65(1):109–127, 2016

DOI:10.1093/sysbio/syv062

Detecting Macroevolutionary Self-Destruction from Phylogenies

LINDELL BROMHAM*, XIA HUA, AND MARCEL CARDILLO

Centre for Macroevolution and Macroecology, Division of Ecology, Evolution and Genetics, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia
Correspondence to be sent to: Centre for Macroevolution and Macroecology, Division of Ecology, Evolution and Genetics, Research School of Biology, Australian National University, Canberra ACT 0200, Australia; E-mail: lindell.bromham@anu.edu.au.
Lindell Bromham and Xia Hua contributed equally to this article.

Received 12 October 2014; reviews returned 26 July 2015; accepted 18 August 2015
Associate Editor: Thomas Buckley

Abstract.—Phylogenetic analyses have lent support to the concept of lineage selection: that biological lineages can have heritable traits that influence their capacity to persist and diversify, and thereby affect their representation in biodiversity. While many discussions have focused on “positive” lineage selection, where stably heritable properties of lineages enhance their diversification rate, there are also intriguing examples that seem to represent “negative” lineage selection, where traits reduce the likelihood that a lineage will persist or speciate. In this article, we test whether a particular pattern of negative lineage selection is detectable from the distributions of the trait on a phylogeny. “Self-destructive” traits are those that arise often but then disappear again because they confer either a raised extinction rate or they are prone to a high rate of trait loss. For such a trait, the reconstructed origins will tend to be dispersed across the tips of the phylogeny, rather than defining large clades of related lineages that all share the trait. We examine the utility of four possible measures of “tippiness” as potential indicators of macroevolutionary self-destruction, applying them to phylogenies on which trait evolution has been simulated under different combinations of parameters for speciation, extinction, trait gain, and trait loss. We use an efficient simulation approach that starts with the required number of tips with and without the trait and uses a model to work “backwards” to construct different possible trees that result in that set of tips. We then apply these methods to a number of case studies: salt tolerance in grasses, color polymorphism in birds of prey, and selfing in nightshades. We find that the relative age of species, measured from tip length, can indicate a reduced speciation rate but does not identify traits that increase the extinction rate or the trait loss rate. We show that it is possible to detect cases of macroevolutionary self-destruction by considering the number of tips with the trait that arise from each inferred origin, and the degree to which the trait is scattered across the phylogeny. These metrics, and the methods we present, may be useful for testing macroevolutionary hypotheses from phylogenetic patterns. [Dead-end; diversification; extinction; lineage selection; macroevolution; speciation.]

Phylogenies are increasingly being used to uncover macroevolutionary patterns, particularly through the identification of correlates of diversification rates. Identification of properties of lineages that can be stably inherited by descendant lineages and can increase lineage survival (by reducing extinction rate) or fecundity (by increasing speciation rate) has led to the revival of the concept of “species selection” (Stanley 1979; Williams 1992; Jablonski 2008; Rabosky and McCune 2010). Much of the focus has been directed at “positive” lineage selection, identifying traits that confer a higher diversification rate. But lineages might also be considered to be subject to “negative” selection, if certain stably inherited characteristics lead to a reduction in the likelihood of a lineage with that trait persisting or speciating, such that lineages with those characteristics will decline in representation.

In this article, our focus is on a particular form of negative lineage selection, which we will refer to as macroevolutionary self-destruction. This refers to lineage traits that evolve often but cause their own disappearance, either because they are rapidly lost again, or because they raise the extinction rate of the lineage. Asexuality is a classic example of a self-destructive trait that evolves often but does not tend to persist over long evolutionary timescales. Williams (1975) stated that sexually reproductive vertebrates were maladapted, in the sense that they would be rapidly replaced by an asexual variant if it arose, however any such lineage would commit “phylogenetic suicide”, keeping the frequency of asexual species low even in the face of frequent transitions to asexuality.

Traits that reduce a lineage’s chance of persisting or diversifying are sometimes referred to as “evolutionary dead-ends”. However, the term “evolutionary dead-end” has been used in many different ways to describe a range of macroevolutionary phenomena, for example traits with a short-term advantage but increased chance of extinction (Schwander and Crespi 2009; Pruitt 2013); traits that cause the extinction of a lineage before it has a chance to speciate (Helanterä et al. 2009); traits that decrease the diversification rate (Agnarsson et al. 2006; Igic and Busch 2013; Cieslak et al. 2014); character transitions that are apparently irreversible (Tripp and Manos 2008; Schneider and Michalik 2011; and lineages with no present-day descendants (Luo 2007). Here, we are specifically interested in traits that arise often but have a consistently negative impact on their own persistence, either by increasing extinction rate or due to high rates of trait reversal, or both.

Traits that cause their own disappearance have been examined in experimental systems, particularly those where short-term advantages at the individual level may have long-term disadvantages at the lineage level. For example, the rate of accumulation of deleterious mutations in parthenogenetic populations has been compared with that in closely related sexual populations (Neiman and Schwander 2011).

Similarly, the longevity of social colonies established by individuals with different behavioral traits has been compared to determine which traits increase the rate of colony extinction (Pruitt 2013). Traits that confer short-term gain but long-term pain may arise from conflict between individual advantage and population persistence (Parvinen 2005). Examples of "evolutionary suicide" might include runaway selection on sexual ornamentation or armaments that reduce the survival probability of their carriers (Gyllenberg and Parvinen 2001), or "tragedy of the commons" where selfish behavior of individuals compromises the survival of a population (Rankin et al. 2007; Kokko and Heubel 2011).

In contrast to studies that have developed population-based models or experiments, we take a comparative, phylogenetic approach, asking whether traits with negative macroevolutionary consequences can be detected from their distribution on phylogenies. These two viewpoints—microevolutionary from population studies and macroevolutionary from phylogenies—are complementary. Population-level studies provide a real-time view of processes that are open to both observation and manipulation, but such studies are limited in their taxonomic scope, and cannot evaluate long-term consequences over evolutionary timescales. While phylogenetic studies can be used to identify significant patterns and to test the relative fit of different macroevolutionary models, they may be limited in the extent to which they can reveal mechanism. However, the strengths of phylogenetic studies are their taxonomic scope, looking for similar patterns over many different lineages and traits; the timescale, allowing evaluation of processes that play out over millions of years; and their repeatability, allowing statistical analyses to be applied to detecting general patterns.

Using Tippy Phylogenetic Distributions to Identify Self-Destructive Traits

We ask whether cases of macroevolutionary self-destruction can be detected from a "tippy" distribution of a discrete trait on a phylogeny. A tippy distribution describes a trait that is dispersed and shallow on a phylogeny, mapped to many tips (terminal edges that connect each contemporary species in the tree to its nearest relative) but relatively few internal edges. We consider two (non-exclusive) ways in which macroevolutionary self-destruction could result in a tippy pattern.

The first is trait lability. Here, we take "labile" to mean "changeable", such that the trait is gained and lost often (Revell et al. 2008). A trait that is highly labile and reversible could change frequently enough that all current origins are recent, and that the trait is not shared with many close relatives. For lability alone to create a tippy pattern, the trait must have a high rate of reversal (loss of trait) relative to the rate of gain, otherwise the trait will define clades of a range of sizes, rather than being present predominantly on many isolated tips. Of course, not all labile traits may be relevant to an examination of macroevolutionary self-destruction, but some labile traits will fall into this category of macroevolutionary phenomena, for example a trait that arises often because of short-term advantages but is soon lost again due to its long-term costs.

The second is extinction rate. A trait that increases the likelihood of extinction might generate a tippy pattern because lineages with this trait would be unlikely to persist for long enough to generate deep clades. Several conditions would be necessary for a "suicidal" trait to generate a tippy distribution by raising the extinction rate. First, the trait must arise often, because a rare trait that causes extinction is unlikely to persist long enough to be observed. Second, the extinction rate must be high relative to the speciation rate, such that a lineage with the trait usually does not give rise to many daughter lineages before succumbing to extinction. Otherwise, it might lead to shallow clades of species with the trait, rather than a tippy pattern.

These two causes of self-destructive traits are non-exclusive, in that they may both act simultaneously. They are also not the only possible causes of a tippy pattern of a trait on a phylogeny. A recent environmental change might promote parallel adaptation in many different lineages. For example, Batalha et al. (2011) reported that evolution of fire resistance is found mainly on the tips of a phylogeny of woody plants from Brazilian cerrado, coincident with the relatively recent rise of flammable grasslands (Simon et al. 2009). Similarly, one explanation of the tippy distribution of plants adapted to serpentine soils is the relatively recent exposure of serpentine outcrops in some regions (Anacker et al. 2011). Evolution of resistance to insecticides might be expected to show an even more pronounced pattern, with recent independent origins in many different lineages (Wood et al. 2005; Hartley et al. 2006). In this article, we will only consider cases where a hypothesis of macroevolutionary self-destruction (frequent gain, frequent loss by extinction or reversal) can be entertained, and not consider cases of parallel adaptation to recent environmental change. A temporal (parallel gain) explanation of a tippy pattern will be most convincing when there is independent evidence of a recent relevant environmental change. Otherwise, it may be difficult to reject the hypothesis that the trait has also been gained and lost many more times in the past, even if origins in extant lineages are demonstrably recent (e.g., self-fertilization in some taxa: Stebbins 1957).

The possibility of using a tippy phylogenetic distribution as an indicator of macroevolutionary processes has been most frequently discussed in terms of the origins of asexual lineages and self-fertilization in plants, where tippy patterns are assumed to arise because the transition from outcrossing to self-reproduction occurs reasonably often but leads to a reduction in population variability (e.g., Schwander and Crespi 2009; Beck et al. 2011; Igic and Busch 2013; Wright et al. 2013). A similar argument has been made for the evolution of sociality in spiders, which leads to small, inbred populations (Agnarsson et al. 2006). Tippy distributions have also been noted where a population-level trait leads to population instability. For example, the evolution of extremely large unicolonial ant populations leads to a reduction in relatedness between individuals, so selection for altruistic behavior may break down, leading to colony collapse (Helanterä et al. 2009).

Tippiness has also been interpreted as a possible sign of the costliness of maintaining specialization (e.g., Bennett et al. 2013). Specialization has often been proposed as an "evolutionary dead-end", where it involves a directional trait change that cannot easily be reversed, although the effect of specialization on diversification rates is a matter of debate (Futuyma and Moreno 1988). There are examples of phylogenetic studies both supporting and refuting the "blind alley" of specialization (Colles et al. 2009). In particular, some phylogenetic studies have demonstrated that specialists can give rise to generalists (e.g., Tripp and Manos 2008), and that specialist lineages can persist and speciate (e.g., Kölsch and Pedersen 2008).

But while tippiness is often invoked to support macroevolutionary hypotheses, it is rarely quantified (Williams 1975; Maynard Smith 1978). Notable exceptions include Schwander and Crespi (2009), who defined "twiggyness" as the phylogenetic isolation of a character state for a given trait, encompassing both recent and ancient lineages that have few close relatives with the same trait, and Agnarsson et al. (2006), who used the term "spindliness" to describe a phylogenetic pattern of many origins that each give rise to a small clade of descendants, interpreted as the signal of frequent origin and high extinction rate. Each of these proposed measures of tippy distributions was limited by the lack of a metric that can be compared across different phylogenies, and the difficulty of generating an appropriate null model against which to test for significant departures from expectation.

In order to explore the potential for using phylogenetic patterns to detect cases of negative lineage selection, we need to establish whether we can reliably interpret a tippy phylogenetic distribution as an indication of a self-destructive trait which arises often but causes its own disappearance. The first thing we need is a way of quantifying what we mean by a tippy distribution. Then we need to explore how reliable these measures of tippiness are as a way of detecting and explaining macroevolutionary self-destruction. We consider four potential metrics of tippiness, applying them to three real-world case studies where self-destructive traits have been proposed. In 1975, George C. Williams remarked that "it may be some time before a reliable measure of nonrandomness in the phylogenetic distribution of the loss of sexuality will appear". We now have all the tools we need to make such measurements on the macroevolutionary effect of traits.

METHODS

Metrics for Describing Tippy Patterns

Several possible ways of detecting and describing tippy patterns have been proposed, including species age, number and distribution of inferred origins, and measures of phylogenetic dispersion. Our aim is to develop unbiased ways of measuring each of these qualities so that we can compare them between phylogenies and generate null distributions under different macroevolutionary models. This will allow us to explore the extent to which phylogenetic patterns can be used to detect and describe a particular class of traits that arise often but then disappear again due to raised chance of extinction or reversal. While this article focuses on traits that arise often then are rapidly lost, the metrics and methods described here may prove more widely applicable for testing macroevolutionary hypotheses from phylogenies (Hua and Bromham in press).

Tip age rank sum.—Given that self-destructive traits limit the persistence of lineages with that trait, lineage age has been used as a potential indicator of self-destructive traits that cause their own disappearance (e.g., Beck et al. 2011; Hugall and Stuart-Fox 2012). If a tippy pattern is caused by a trait that raises extinction rate relative to speciation rate, then lineages that evolve the trait do not persist for long. Since lineages with this trait have a limited lifespan, the presence of the trait in any contemporary species is likely to be due to a recent origin—these lineages are "dead men walking". Therefore, one potential way to evaluate whether a trait causes lineages to die young is to compare the ages of species with the trait to the ages of similar species without the trait. Age of species has been used to look at patterns of diversification and species persistence, for example examining the persistence of asexual lineages (Neiman et al. 2009; Beck et al. 2011), color polymorphism in bird populations (Hugall and Stuart-Fox 2012), adaptation to recently exposed habitats (Anacker et al. 2011), and contrasting evolutionary sources to sinks (Goldberg et al. 2005).

However, the use of species age in evaluating such hypotheses has been questioned, particularly with respect to the inference of the long-term costs of asexuality. For example, a high rate of generation of asexual clones combined with the replacement of clonal offshoots by a neutral turnover may lead to relatively short asexual tip lengths even if asexuality does not lead to a higher lineage extinction rate (Janko et al. 2008; Janko 2014). Some empirical tests of tip branch length have found no significant different between sexual and asexual taxa (Neiman et al. 2009). Similarly, while the tippy distribution of sociality in spiders has been interpreted as a sign of a reduction in long-term viability (e.g., Agnarsson et al. 2006), a phylogenetic study found a high level of sequence divergence between social spider species, which was interpreted as indicating their long-term persistence (Johannesen et al. 2007).

Within a phylogenetic context, the age of species is usually represented by the length of the edge connecting the species to the node that it shares with its nearest relative (on an ultrametric tree), which we will refer to as the tip age. We can test whether the tips with the trait of interest tend to be shorter than those without. This is equivalent to cutting off all the tips of the tree, lining them up by length, and asking if the species with the trait are randomly distributed throughout the size distribution or skewed toward the shorter tips. The tip age rank sum (TARS) test uses the Wilcoxon rank-sum test, for which we report the P value of the test (pTARS). The P value reflects the proportion of randomly distributed traits that have a greater difference in length between the tips with and without the trait than for the observed distribution. pTARS ≤ 0.05 indicates that the difference in length between the set of tips with the trait and those without is greater than in 95% of the randomly generated traits. This test is nonparametric and thus is relatively robust to differences in the absolute value of diversification rate.

Number of tips per origin.—A tippy pattern arises when there are many independent evolutionary origins of a trait that are all relatively recent and each origin gives rise to only one or few descendant species with the trait of interest. Therefore, the number of reconstructed origins will be relatively high given the number of tips with the trait. (Agnarsson et al., 2006) summarized this pattern (which they termed "spindliness") by calculating the proportion of internal and terminal nodes reconstructed as having the trait. But this measure will be strongly affected by tree shape and species sampling, so that the observations cannot be tested against a null model derived from simulating traits on the same tree, because the hypothesized links between trait values and speciation, extinction, and trait transition probabilities are expected to influence tree shape. A simpler but more robust approach is to compare the number of extant species that arise from each reconstructed origin.

We estimate the minimum number of trait origins needed to explain the observed pattern of tips using a parsimony-based criterion. Our aim here is not to reconstruct the history of the trait over the entire phylogeny but specifically to ask what is the minimum number of origins we would need to describe the pattern of traits across the tips of the tree. The advantage of a parsimony approach is that it is nonparametric and it infers the minimum number of origins needed to explain the distribution of tips with the trait, without inferring "internal" origins and subsequent losses that do not lead to tips with the trait (Bennett et al. 2013). A maximum-likelihood (ML) approach is less suited to this purpose because it tends to reconstruct many origins and losses on the internal edges of a phylogeny under higher trait transition rates. The inference of many internal gains and losses that have not resulted in extant species with the trait complicates the comparison of number of origins per observed number of species with the trait. For a self-destructive trait, there are likely to be many more gains that have happened throughout the history of the lineages then been lost again by extinction or reversal, but in order to determine whether a tippy phylogenetic pattern is indicative of macroevolutionary self-destruction, we consider only the distribution of the trait across extant tips, not past gains and losses that do not lead to extant taxa with the trait.

To test whether the observed NoTO value is significantly lower than expected, we generate a null distribution using a Brownian motion model to evolve 1000 binary traits on each tree tested, following the method of Felsenstein (2005) and Fritz and Purvis (2010). In brief, a continuous trait is first generated under Brownian motion and a threshold is then chosen to convert the continuous trait to a binary trait with the same prevalence as seen in the observed data. We then count how many traits generated under the Brownian motion model have lower NoTO values than the observed value, and report the count number as a proportion (pNoTO).

Phylogenetic clustering.—Trait lability is often described in terms of relatively low phylogenetic signal, such that closely related taxa share the same trait value no more frequently than more distantly related taxa (Blomberg et al. 2003). Therefore, we would expect a self-destructive trait that is gained often and lost often to have low phylogenetic signal. A trait that raises the extinction rate will also reduce phylogenetic clustering, so is expected to be overdispersed on the phylogeny (Purvis 2008). We evaluate the utility of two different measures of phylogenetic signal for detecting self-destructive traits. The degree of clustering can be represented by the sum of sister clade differences (SSCD). Given a binary trait that takes values of 1 or 0, values can be estimated at each node of the phylogeny from the state at the two descendant (daughter) nodes. A sister clade difference at a given node is the absolute difference between trait values of the two daughter nodes descending from that node. SSCD is the sum of the absolute difference in trait values over all pairs of daughter nodes, which will be higher for a more labile trait (Fritz and Purvis 2010). As for NoTO, we test the significance of the observed value of SSCD for a given tree by generating a null distribution using a Brownian motion model to evolve 1000 binary traits on that tree, then count how many simulated values are higher than the observed value and report the count number as a proportion (pSSCD).

The Fritz and Purvis D statistic (FPD) calculates the difference between observed SSCD and expected SSCD under Brownian motion, scaled by the difference between SSCD under random distributions of the trait across the tips of the phylogeny and SSCD under

Brownian motion (Fritz and Purvis 2010). Thus, the absolute value of FPD equals 1 if the observed trait is randomly distributed across tips, and equals zero if the trait evolves under Brownian motion. The value is derived by generating 1000 Brownian motion traits on each tree, and 1000 replicates of random distribution made by shuffling the trait values across the tips of the tree. We expect a tippy trait to have significantly positive FPD: as the trait becomes more scattered than a randomly distributed trait, the value approaches or exceeds 1. Therefore, the test for this metric is whether the FPD value is greater than zero or greater than 0.5. We use these two measures of phylogenetic clustering because, while they share some features, they may reveal different patterns in trait evolution or have different levels of sensitivity to different mechanisms.

Macroevolutionary Models

Our next step was to ask if these metrics can reliably detect cases of self-destructive traits that evolve often but do not persist due to high rates of trait loss, raised extinction, or both. We use simulations to generate trees and traits under a range of macroevolutionary scenarios. One of these models is a baseline model against which the others are compared: the other models differ from the baseline in the value of 1 or 2 parameters (Table 1). There are three self-destructive models, in which the trait has a higher rate of gain and loss (labile), or a higher rate of extinction (dead-end), or higher rate of gain and a higher extinction rate (suicide) than the baseline model. There are also two additional models that are not self-destructive, which are included for comparison because these models are often discussed in terms of tippy distributions on phylogenies: the lonely model, where the trait reduces the speciation rate compared with the baseline model, and the irreversible model, where the trait loss rate is zero.

In the labile model, the trait is gained more often than the baseline model but also lost more often. An example of this model might be costly specialization strategies which are initially advantageous in allowing access to an underutilized resource but come at the cost of decreased growth or reproduction which might limit their long-term viability as a strategy. For example, it has been suggested that the tippy pattern in plants adapted to saline or serpentine soils may reflect frequent gain and frequent loss (Anacker et al. 2011; Bromham 2014). In the dead-end model, the rate of trait gain is the same as the baseline model, but when it arises, the trait causes an increase in the extinction rate. For example, it has been suggested that when colony size in ants exceeds a threshold size, the colony collapses, leading to a "twiggy" phylogenetic distribution of unicolonial ants (Helanterä et al. 2009). The suicide model describes a trait that is gained more often than baseline and also increases the extinction rate. For example, it has been suggested that sociality in spiders has originated many times but leads to raised extinction rates so that the social lineages disappear before they have a chance to diversify (Agnarsson et al. 2006).

In addition, we include several models that have been discussed in terms of negative lineage selection and tippy patterns on phylogenies. The lonely model has the same rate of gain and loss as the baseline model, but the trait causes a drop in the speciation rate, without affecting the extinction rate. It is called lonely because it is expected to result in long lonely lineages with few relatives with the trait. For example, it has been suggested that moth pollination in flowers is an evolutionary dead-end because loss of pigment cannot be reversed, so a lineage that moves to moth pollination cannot then diversify to other pollinators (Tripp and Manos 2008). We also include the irreversible model, which differs from the baseline model only in the loss rate of the trait which is zero. We include this because irreversibility has been cited as a contributing factor to "evolutionary dead-ends". For example, irreversibility has been suggested to contribute to selfing being an evolutionary dead-end on the assumption that biparental mating is unlikely to be regained once lost (Igic and Busch 2013). It is therefore appropriate to test what phylogenetic patterns can arise from irreversibility on its own, and whether these might mimic self-destructive traits.

TABLE 1. Parameter combinations used in the simulations

Model description	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
1. Baseline: (balanced speciation and extinction, balanced gain and loss)	0.1	0.1	0.01	0.01	0.01	0.01
	0.3	0.3	0.03	0.03	0.01	0.01
	0.3	0.3	0.01	0.01	0.01	0.01
	0.1	0.1	0.03	0.03	0.01	0.01
2. Labile (balanced speciation and extinction, higher gain + higher loss	0.1	0.1	0.01	0.01	**0.1**	**0.1**
	0.3	0.3	0.03	0.01	**0.1**	**0.1**
	0.1	0.1	0.01	0.03	**0.1**	**0.1**
	0.3	0.3	0.03	0.03	**0.1**	**0.1**
3. Dead-end (increased extinction, balanced gain and loss)	0.1	0.1	0.01	**0.1**	0.01	0.01
	0.3	0.3	0.03	**0.1**	0.01	0.01
	0.3	0.3	0.01	**0.1**	0.01	0.01
	0.1	0.1	0.03	**0.1**	0.01	0.01
4. Suicide (higher extinction, higher gain)	0.1	0.1	0.01	**0.1**	**0.3**	0.01
	0.3	0.3	0.03	**0.1**	**0.3**	0.01
	0.3	0.3	0.01	**0.1**	**0.3**	0.01
	0.1	0.1	0.03	**0.1**	**0.3**	0.01
5. Lonely (reduced speciation, balanced gain and loss)	0.1	**0.001**	0.01	0.01	0.01	0.01
	0.3	**0.001**	0.03	0.03	0.01	0.01
	0.3	**0.001**	0.01	0.01	0.01	0.01
	0.1	**0.001**	0.03	0.03	0.01	0.01
6. Irreversible (balanced speciation and extinction, no trait loss)	0.1	0.1	0.01	0.01	0.01	**0**
	0.3	0.3	0.03	0.03	0.01	**0**
	0.3	0.3	0.01	0.01	0.01	**0**
	0.1	0.1	0.03	0.03	0.01	**0**

Notes: For a trait that can be present (1) or absent (0), the six parameters are the speciation rate in lineages with (λ_1) or without the trait (λ_0), the extinction rate in lineages with (μ_1) or without (μ_0) the trait, and the rate of gain (μ_{01}) and loss (μ_{10}) of the trait. Speciation rate and extinction rate can take one of two values, so there are four baseline models that have different combinations of these parameters. Each of the evolutionary models is run under all four scenarios but with one or two parameters changed, as outlined below in bold.

These models can be framed in the same terms as binary state speciation and extinction model (BiSSE: Maddison et al. 2007). For a trait that can be present (1) or absent (0), the six parameters are speciation rate in lineages with (λ_1) or without the trait (λ_0), the extinction rate in lineages with (μ_1) or without (μ_0) the trait, and the rate of gain (q_{01}) and loss (q_{10}). The six different macroevolutionary models we use are:

(1) Baseline: The trait does not affect the speciation rate ($\lambda_0 = \lambda_1$) or extinction rate ($\mu_0 = \mu_1$), and rates of trait gain and loss are equal ($q_{01} = q_{10}$).

(2) Labile: Gain and loss rates are equal ($q_{01} = q_{10}$) but higher than in the baseline model, but the trait does not affect speciation and extinction rates ($\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$).

(3) Dead-end: The trait increases the extinction rate ($\mu_0 < \mu_1$), but the speciation and trait transition rates are equal between lineages with and without the trait ($\lambda_0 = \lambda_1$, $q_{01} = q_{10}$).

(4) Suicide: The trait has a rate of gain greater than the loss rate ($q_{01} > q_{10}$). In addition, the trait causes a raised extinction rate ($\mu_0 < \mu_1$), but the speciation rate is the same as the baseline ($\lambda_0 = \lambda_1$).

(5) Lonely: The trait reduces the speciation rate ($\lambda_0 > \lambda_1$), but the extinction and trait transition rates are equal in lineages with and without the trait ($\mu_0 = \mu_1$, $q_{01} = q_{10}$).

(6) Irreversible: The loss rate is zero ($q_{10} = 0$) speciation and extinction rates are equal in lineages with and without the trait ($\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$).

Each of the six models was run under four combinations of parameters (see Table 1), representing high and low values for speciation rate and extinction rate. To give reasonable bounds for the simulations, these parameter values were derived from large-scale phylogenetic studies over major clades in chordates, arthropods, molluscs, and plants (Weir and Schluter 2007; Rabosky et al. 2013; De Vos et al. 2015). One or two of the parameters in the baseline model were adjusted to create each of the alternative macroevolutionary models (Models 2–6: Table 1). In addition, we repeated simulations for Models 3–6 using the higher trait transition rates in the labile model (Supplementary Table S1, available on Dryad at http://dx.doi.org/10.5061/dryad.10668) in order to compare the relative contribution of trait lability versus trait effects on diversification rates to the tippy distribution of the trait on a phylogeny.

To test the effects of different sampling scenarios on phylogenetic patterns, we repeated each of the model simulations under a number of different values for tree size, trait prevalence, and sampling fraction (Table 2). We varied the absolute number of tips with a trait as 10, 20, and 100 lineages, and the relative number of tips with and without the trait as 1:10 and 1:2. We examined the effect of incomplete sampling by analyzing both the complete phylogeny (100% of extant tips included) and an incompletely sampled tree (where only 10% of extant tips are included, randomly selected with respect to trait state).

For each of the six models under each of the four sets of parameter values (Table 1), we generated 100 simulated phylogenies under each of the five sampling scenarios (Table 2). For each simulated phylogeny, we calculated the values of the four metrics (TARS, NoTO, SSCD, and FPD) and their corresponding P values (pTARS, pNoTO, pSSCD) specific to that phylogeny. We drew the distribution of metric values from the value of each of the metrics calculated for the 100 simulated trees under each of the simulation conditions. We then assessed the power of each metric test to detect each of the macroevolutionary models under different simulation conditions by calculating how often the test gives significant result, that is, P values were no greater than 0.05 (for pTARS, pNoTO, and pSSCD), or FPD values were greater than zero or greater than 0.5. This power is different from the power of a metric test to distinguish one macroevolutionary model from another. But it can suggest the ability to distinguish different models if the power is high under one model and low under the others. Note that we keep these metric tests simple by generating null distributions on the tree being tested. However, this means of generating a null distribution may not capture the range of values expected for traits that are linked to speciation, extinction, and trait transition probabilities, because such traits would also be expected to influence tree shape.

Backward Simulation of Phylogenies

In order to provide null distributions of the four metrics, the simulated trees should be constrained to have the same size and relative number of tips with and without the trait as the observed data. This is tricky to enforce using conventional phylogeny simulation methods that “grow” trees from an initial starting lineage using a model that specifies the probability of lineage speciation and extinction, because the resulting trees can vary widely in size (number of tips) and trait representation (number of tips with a given trait). While it is possible to conduct classic tree simulations then discard any trees that do not match the size and trait representation of the observed data, this is a computationally inefficient strategy.

We solve this problem using “backwards simulation” to build trees with a given number of tips with and without the trait. We developed an algorithm that starts with the extant numbers of sampled taxa both with and without the trait, then constructs the phylogeny using a birth–death model with a binary trait that can influence speciation and extinction rates to generate trees with the appropriate proportions of traits at the tips. The algorithm makes our test for tippiness more efficient because it automatically constrains the extant numbers

TABLE 2. False-positive rate and statistical power of significance tests on the four metrics

	Trait prevalence (Sampling)	Low speciation, low extinction						High speciation, high extinction						High speciation, low extinction						Low speciation, high extinction					
		B	La	D	S	Lo	I	B	La	D	S	Lo	I	B	La	D	S	Lo	I	B	La	D	S	Lo	I
pTARS≥0.95	10:100	**0.05**	0.03	0.13	0.05	**0.98**	0.05	0.09	0.05	0.07	0.07	—	0.08	0.06	0	0.12	0.05	**0.94**	0.05	**0.05**	0	0.10	0.04	—	0.07
	10:100 (10%)	**0.05**	0.01	0.17	0.02	**0.86**	0.07	**0.04**	0.02	0.11	0.02	—	0.02	**0.03**	0.02	0.13	0.01	**0.89**	0.05	**0.01**	0.01	0.19	0.03	—	0.04
	20:200	0.07	0.01	0.17	0.02	**1**	0.03	0.09	0.03	0.07	0.05	—	0.07	**0.03**	0.03	0.11	0	**1**	0.10	0.07	0.01	0.13	0.02	—	0.11
	100:200	0.12	0.02	0.29	0.10	**1**	0.04	**0.03**	0.03	0.02	0.02	—	0.03	0.07	0.02	0.10	0.08	**1**	0.03	0.09	0.04	0.21	0.06	—	0.10
	100:1000	0.06	0.02	0.36	0.06	**1**	0.05	0.07	0.04	0.09	0.03	—	0.08	**0.03**	0.03	0.12	0.07	**1**	0.04	**0.05**	0.06	0.19	0.05	—	0.04
pNoTO≤0.05	10:100	**0**	0.77	0.05	**0.95**	0.17	0.01	**0**	0.46	0	0.33	—	0	**0**	0.47	0	0.35	0	0	**0.02**	0.74	0.06	**0.95**	—	0
	10:100 (10%)	0.11	0.55	0.17	**0.96**	0.53	0.02	**0**	0.51	0	0.46	—	0	**0**	0.51	0	0.42	0.05	0	0.11	0.62	0.24	**0.95**	—	0.05
	20:200	**0.01**	**0.99**	0.05	**1**	0.38	0	**0**	**0.90**	0	**0.83**	—	0	**0**	**0.88**	0	0.79	0	0	**0.02**	**1**	0.13	**1**	—	0.01
	100:200	**0**	**1**	0.03	**1**	0.01	0	**0**	**1**	0	**0.98**	—	0	**0**	**1**	0	**0.98**	0	0	**0**	**1**	0.11	**1**	—	0
	100:1000	0.12	**1**	0.90	**1**	**0.97**	0.02	**0**	**1**	0	**1**	—	0	**0**	**1**	0	**1**	0	0	0.33	**1**	**0.93**	**1**	—	0.08
pSSCD≤0.05	10:100	**0**	**0.94**	0.03	**0.99**	0.09	0	**0**	0.79	0	0.66	—	0	**0**	0.79	0	0.59	0	0	**0.01**	**0.95**	0.06	**0.99**	—	0
	10:100 (10%)	0.18	**0.82**	0.42	**1**	0.79	0.02	**0**	0.78	0	0.72	—	0	**0**	**0.82**	0	**0.81**	0.02	0	0.23	**0.89**	0.51	**1**	—	0.04
	20:200	**0**	**1**	0.08	**1**	0.46	0	**0**	**0.99**	0	**0.95**	—	0	**0**	**0.99**	0	**0.93**	0	0	**0.03**	**1**	0.13	**1**	—	0
	100:200	**0**	**1**	0.07	**1**	0.22	0	**0**	**1**	0	**1**	—	0	**0**	**1**	0	**1**	0	0	**0**	**1**	0.18	**1**	—	0
	100:1000	0.13	**1**	**0.96**	**1**	1	0	**0**	**1**	0	**1**	—	0	**0**	**1**	0	**1**	0	0	0.37	**1**	**0.97**	**1**	—	0.02
FPD>0	10:100	0.34	**1**	0.68	**1**	**0.92**	0.12	**0.05**	**1**	0.01	**0.98**	—	0	**0.01**	**0.99**	0.02	**1**	0.18	0	0.49	**1**	0.66	**1**	—	0.24
	10:100 (10%)	0.81	**1**	**0.96**	**1**	**1**	0.61	0.11	**1**	0.09	**1**	—	0.04	**0.01**	**1**	0.11	**1**	0.44	0.02	0.85	**1**	**1**	**1**	—	0.70
	20:200	0.46	**1**	**0.82**	**1**	**0.98**	0.22	**0.01**	**1**	0	**1**	—	0	**0**	**1**	0.02	**1**	0.24	0	0.66	**1**	**0.91**	**1**	—	0.40
	100:200	0.24	**1**	**0.92**	**1**	**0.92**	0	**0**	**1**	0	**1**	—	0	**0**	**1**	0	**1**	0.02	0	0.36	**1**	**0.96**	**1**	—	0
	100:1000	0.97	**1**	**1**	**1**	**1**	**0.83**	**0**	**1**	0	**1**	—	0	**0**	**1**	0	**1**	0.57	0	1	**1**	**1**	**1**	—	**0.98**
FPD>0.5	10:100	**0.01**	**0.96**	0.05	**0.99**	0.14	0	**0**	**0.88**	0	0.72	—	0	**0**	**0.86**	0	0.74	0	0	**0.02**	**0.97**	0.08	**0.98**	—	0
	10:100 (10%)	0.43	**0.95**	0.61	**1**	**0.91**	0.13	**0.01**	**0.94**	0	**0.90**	—	0	**0**	**0.97**	0	**0.93**	0.07	0	0.43	**0.96**	0.69	**1**	—	0.19
	20:200	**0**	**1**	0.02	**1**	0.25	0	**0**	**0.95**	0	**0.86**	—	0	**0**	**0.98**	0	**0.84**	0	0	**0**	**1**	0.06	**1**	—	0
	100:200	**0**	**1**	0	**1**	0	0	**0**	**1**	0	0.53	—	0	**0**	**1**	0	0.52	0	0	**0**	**1**	0	**1**	—	0
	100:1000	**0**	**1**	0	**1**	0.64	0	**0**	**1**	0	**1**	—	0	**0**	**1**	0	**1**	0	0	**0**	**1**	0.01	**1**	—	0

Notes: Each value is the proportion of the simulated data sets that have pTARS ≥ 0.95, pNoTO ≤ 0.05, pSSCD ≤ 0.05, FPD > 0, or FPD > 0.5. Values under the baseline model (B = baseline) show the false-positive rate of these significance tests because the baseline model is treated as the null hypothesis. Values under each of the other evolutionary models (La = labile; D = dead-end; S = suicide; Lo = lonely; I = irreversible) show the statistical power of the significance tests when the model is treated as the alternative hypothesis. False-positive rate ≤ 0.05 and power ≥ 0.8 are highlighted in bold to help detecting conditions under which each metric can reliably detect the alternative hypothesis. For each significance test, each row is under the same condition of trait prevalence and sampling fraction; each of the four column sections is under the same condition of speciation and extinction rate (see Table 1). The sampling strategy gives the number of tips with and without the trait, respectively, where 10% means 10% of extant tips are included, randomly selected with respect to trait state. With high extinction rate, trees do not coalesce under the "lonely" model, so no values are shown.

of taxa. In addition, it accounts for random incomplete sampling, which is common in the observed data. The backward simulation method has been implemented in an R package called "Phylometrics" (Hua and Bromham in press).

Case Studies

We analyze three published data sets that have been interpreted as providing evidence for a trait that is evolving under a "macroevolutionary self-destruction" scenario (Table 3). We chose these three data sets as representing different forms of traits that could be said to cause their own disappearance, and because sufficiently large data sets were available with multiple trait transitions within a single phylogeny. In each of these case studies, a trait is proposed to evolve frequently, but then be lost again by trait loss or lineage extinction. Salt tolerance in plants has been suggested to be gained often due to the competitive advantage of being able to colonize saline conditions, but then be frequently lost due to the costs of maintaining tolerance. Color polymorphism in birds has been suggested to have multiple origins but is often lost again through fixation to generate two monomorphic daughter populations. Selfing in plants is often considered to be an evolutionary dead-end due to reduced capacity to adapt or speciate.

Salt tolerance in grasses.—Bennett et al. (2013) noted a tippy pattern in the evolution of salt tolerance in grasses, with over 70 estimated origins giving rise to 200 salt-tolerant species. They suggested that the tippy pattern of salt tolerance in grasses might indicate that salt tolerance is gained often but lost often, implying that it is a costly trait that arises often but is not maintained over long evolutionary time periods. Here, we focus on a subset of taxa from the Bennett et al. analysis: the Pooideae, a subfamily of grasses containing major crop species such as wheat and barley. The phylogeny, derived from Edwards and Smith (2010), contains 461 taxa of which 24 are identified as salt-tolerant.

Color polymorphism in birds of prey.—Color polymorphism—the presence of multiple distinct color morphs in an interbreeding population—is rare in birds but taxonomically widespread. The rarity and phylogenetic scatter of color polymorphism has been explained by its loss through fixation of one morph when the population splits into two monomorphic daughter lineages. Hugall and Stuart-Fox (2012)

TABLE 3. Summary of the three data sets examined in this study

Trait	Taxon	Proposed mechanism	Tree size	Tips with trait	Sampling fraction		Source
					With trait	Without	
Salt tolerance	Pooideae (grasses)	High trait loss or raised extinction	461	24	0.12	0.12	Bennett et al. (2013)
Color polymorphism	Accipitridae (birds of prey)	High trait loss	178	43	0.69	0.81	Hugall and Stuart-Fox (2012)
Selfing	Solanaceae (nightshades)	Raised extinction rate	356	221	0.15	0.16	Goldberg et al. (2010)

Notes: Proposed mechanism is the explanation for the "tippy" pattern given in the original publication (listed under Source). Tree size is the number of terminal taxa included in the phylogeny, and tips with trait are the number of taxa with the putative "self-destructive" trait. The sampling fraction is the estimated proportion of extant species from the taxon included in the phylogeny, as reported in the source.

suggested that color polymorphism in birds is lost at a significantly greater rate than is gained, and that it stimulates diversification. Here, we analyze one group from their study, the Accipitridae (diurnal birds of prey including hawks and eagles). This phylogeny contains 178 taxa, 43 with color polymorphism.

Selfing in nightshades.—Most flowering plants are self-incompatible, meaning that pollen and ovules from the same plant cannot be combined to make a viable offspring. However, self-compatibility (selfing), which allows uniparental reproduction, has evolved multiple times. It has been suggested that selfing has a short-term advantage in terms of increased speciation rate, but long-term disadvantage due to increase in extinction rate. Goldberg et al. (2010) suggested that lineages in the Solanaceae (nightshades, such as potato and tobacco) undergo frequent transitions from outcrossing to selfing, but that self-compatible lineages have a higher extinction rate than self-incompatible lineages. We use Goldberg et al.'s (2010) phylogeny of 356 taxa, 221 of which are recorded as self-compatible.

We performed all analyses on the published phylogeny from each study, scoring the tips as 0 for the presumed ancestral trait or 1 for the putative self-destructive trait. We derived sampling fraction (the estimated proportion of taxa included in the phylogeny) from the published study. For each case study, we calculated the value of each of the four metrics on the phylogeny provided in the original source, and tested for a significant departure from expectation (using the Wilcoxon rank-sum test for TARS, 1000 sets of Brownian motion traits for NoTO and SSCD, and 1000 sets of Brownian motion traits and random traits for FPD). Then we used simulations to explore which macroevolutionary models are compatible with the observed trait distribution in each of the case studies and how reliably the results of each significance test indicate self-destruction.

In order to condition the simulations with appropriate parameter values for each case study, we first found the ML estimates of the six parameters in the BiSSE model (Maddison et al. 2007) using the "diversitree" R package (FitzJohn 2012). The parameters are speciation rate with (λ_1) and without the trait (λ_0), extinction rate in lineages with (μ_1) and without (μ_0) the trait, and rates of trait gain (q_{01}) and loss (q_{10}). To test a range of evolutionary scenarios, we generated eight alternative sets of parameter values for each case study by constraining three, two, one, or none of the estimated rates to be equal between lineages with and without the traits (Table 4). We generated alternative models in this way because it allows us to distinguish the relative contribution of speciation, extinction, or trait evolution to generating the observed trait distribution on phylogenies. For example, constraining speciation rate and trait transition rate to be equal between lineages with and without the trait allows us to test if asymmetry in extinction rate, analogous to the dead-end model, can generate the observed trait distribution on phylogenies. Furthermore, BiSSE models with fewer parameters have higher power to detect asymmetries in speciation, extinction, or trait transition rates (Davis et al. 2013). For the nightshades case study, we generated an additional set of parameter values to account for the irreversibility of selfing as suggested by Goldberg et al. (2010).

We estimated these alternative sets of parameter values using constrained BiSSE models (FitzJohn 2012). We then simulated 100 trees under each set of parameter values, conditional on the sampling fraction and extant number of taxa with and without the trait (Table 3). Under certain parameter sets, the lineages did not coalescence into one ancestral lineage because of the negative diversification rate of lineages with the trait. In such parameter sets, the extinction rate is reduced by 0.01 each time until the lineages coalesce. Because the BiSSE estimates of extinction rate have wider standard error than the other parameters (Maddison et al. 2007; Davis et al. 2013), adjusting extinction rate in this way is likely to have relatively little influence on the likelihood of the model. We compared the observed value for the four metrics (TARS, NoTO, SSCD, and FPD) to the values for the simulated data sets. We rejected a model as being inconsistent with the data if the observed value for any of these metrics was significantly different from the values generated by the 100 simulations under that model.

Table 4. ML estimates of the six BiSSE parameters for each of three data sets: speciation rate with and without the trait (λ_1, λ_0), extinction rate with and without the trait (μ_1, μ_0), and trait gain and loss rates (q_{01}, q_{10})

Trait	Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
Salt tolerance							
a)	No constraints	9.28	0.00	2.66	0.00	0.00	0.66
b)	$\lambda_0=\lambda_1$, $\mu_0=\mu_1$, $q_{01}=q_{10}$	9.25	9.25	3.10	3.10	0.15	0.15
c)	$\lambda_0=\lambda_1$, $\mu_0=\mu_1$	9.26	9.26	3.11	3.11	6.41	116.7
d)	$\lambda_0=\lambda_1$, $q_{01}=q_{10}$	8.90	8.90	0.51	43.7 (8.00)	2.39	2.39
e)	$\mu_0=\mu_1$, $q_{01}=q_{10}$	9.60	0.00	3.11	3.11	0.54	0.54
f)	$\lambda_0=\lambda_1$	1.43	1.43	1.07	0.00	0.16	1.34
g)	$\mu_0=\mu_1$	9.16	0.00	2.44	2.44	0.78	5.16
h)	$q_{01}=q_{10}$	9.00	0.00	1.37	21.4 (0.00)	1.59	1.59
Color polymorphism							
a)	No constraints	0.62	0.50	0.09	0.00	0.05	0.45
b)	$\lambda_0=\lambda_1$, $\mu_0=\mu_1$, $q_{01}=q_{10}$	0.24	0.24	0.00	0.00	0.06	0.06
c)	$\lambda_0=\lambda_1$, $\mu_0=\mu_1$	0.24	0.24	0.00	0.00	0.12	0.42
d)	$\lambda_0=\lambda_1$, $q_{01}=q_{10}$	0.29	0.29	0.00	0.36	0.15	0.15
e)	$\mu_0=\mu_1$, $q_{01}=q_{10}$	0.23	0.33	0.00	0.00	0.05	0.05
f)	$\lambda_0=\lambda_1$	0.38	0.38	0.37	0.00	0.00	0.32
g)	$\mu_0=\mu_1$	0.15	0.53	0.06	0.06	0.06	0.42
h)	$q_{01}=q_{10}$	0.19	1.45	0.00	1.36	0.18	0.18
Selfing							
a)	No constraints	0.21	9.71	0.00	9.72	1.34	0.82
b)	$\lambda_0=\lambda_1$, $\mu_0=\mu_1$, $q_{01}=q_{10}$	4.21	4.21	4.13	4.13	0.12	0.12
c)	$\lambda_0=\lambda_1$, $\mu_0=\mu_1$	4.29	4.29	4.21	4.21	0.17	0.09
d)	$\lambda_0=\lambda_1$, $q_{01}=q_{10}$	4.22	4.22	4.20	4.09	0.13	0.13
e)	$\mu_0=\mu_1$, $q_{01}=q_{10}$	4.15	4.26	4.13	4.13	0.13	0.13
f)	$\lambda_0=\lambda_1$	4.20	4.20	4.00	4.21	0.22	0.06
g)	$\mu_0=\mu_1$	4.31	4.13	4.13	4.13	0.21	0.06
h)	$q_{01}=q_{10}$	4.01	4.36	4.00	4.24	0.13	0.13
i)	$q_{10}=0$	3.14	5.20	2.61	5.41 (5.19)	0.48	0

Notes: Parameters were estimated under a range of constraints, reflecting whether there was no effect of trait on speciation rates or extinction rate and equal rates of gain and loss. Where extinction rate was adjusted to allow coalescence, the adjusted value is given in parenthesis following the estimated value.

Results

Metrics for Describing Tippy Patterns

In the simulations conducted, the TARS measure identified the lonely model (Fig. 1), where lineages with the trait have a reduced speciation rate without reducing the extinction rate. The significance test on TARS had a low false-positive rate for rejecting the Baseline model and high statistical power to detect the lonely model (Table 2). Under the lonely model, the terminal lineages with the trait tend to be significantly longer than those without the trait. TARS values were not significantly different between the baseline model and the suicide, labile, or dead-end models. This suggests that the age of tips with the trait of interest is not a good indicator of the influence of the trait on raised extinction or high lability, and therefore does not detect self-destructive traits.

Trees generated under the suicide model (high gain, high extinction rate) and the labile model (high gain, high loss) tend to have fewer tips resulting from each reconstructed trait origin than the other models, so have significantly lower NoTO (Fig. 2). NoTO values under the two models are more different from the other models when speciation rate is high or when the ratio between the number of tips with and without the trait is high (Fig. 2). Under the suicide and labile models, NoTO was always significantly lower than for Brownian motion traits simulated on the same tree. This significance test has low false-positive rate to reject the baseline model and high statistical power to detect the suicide model and the labile model when the absolute number of tips with the trait increases (Table 2). But the statistical power to distinguish the suicide model and the labile model from the lonely and dead-end models decreases when speciation rate is low and the absolute difference in the number of tips with and without the trait increases (Table 2). NoTO did not consistently detect the dead-end (high extinction rate) model (Fig. 2, Table 2).

SSCD is larger under the suicide and labile models than for the other models (Fig. 3). SSCD values under the two models are more different from the other models when speciation rate is high or when the absolute number of tips with the trait increases (Fig. 3). Under the suicide and labile models, SSCD is always significantly larger than Brownian motion traits simulated on the same tree, suggesting that these models give rise to a significant amount of scattering of the trait on the phylogeny. But, like NoTO, this significance test only reliably identifies the two models when the absolute difference in the number of tips with and without the trait is small if speciation rate is low, or when the absolute number of tips with the trait is large (Table 2).

FPD is significantly larger than zero under the suicide and labile models over all parameter combinations, and

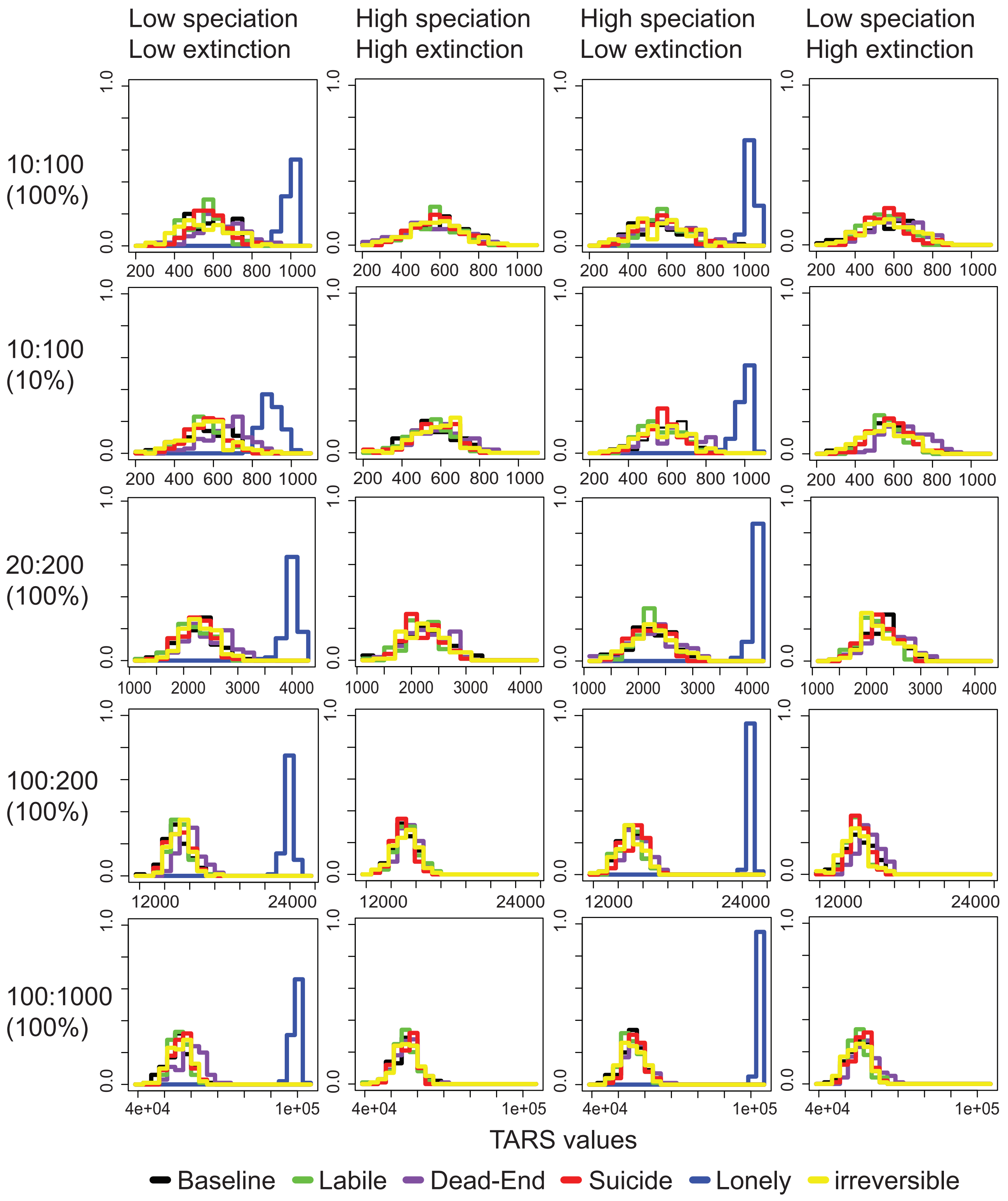

FIGURE 1. Distributions of TARS values of the six evolutionary models described in Table 1, under four combinations of high or low speciation and extinction rates and five combinations of tree size, trait prevalence and sampling fraction. The ratio of tips with and without the trait is given on the left-hand side (with: without), and the percentage of tips from the final tree included in the data set is given in brackets below. A distribution that is more distinct from others indicates that TARS could identify its corresponding evolutionary model more reliably. With high extinction rate, trees do not coalesce under the lonely model.

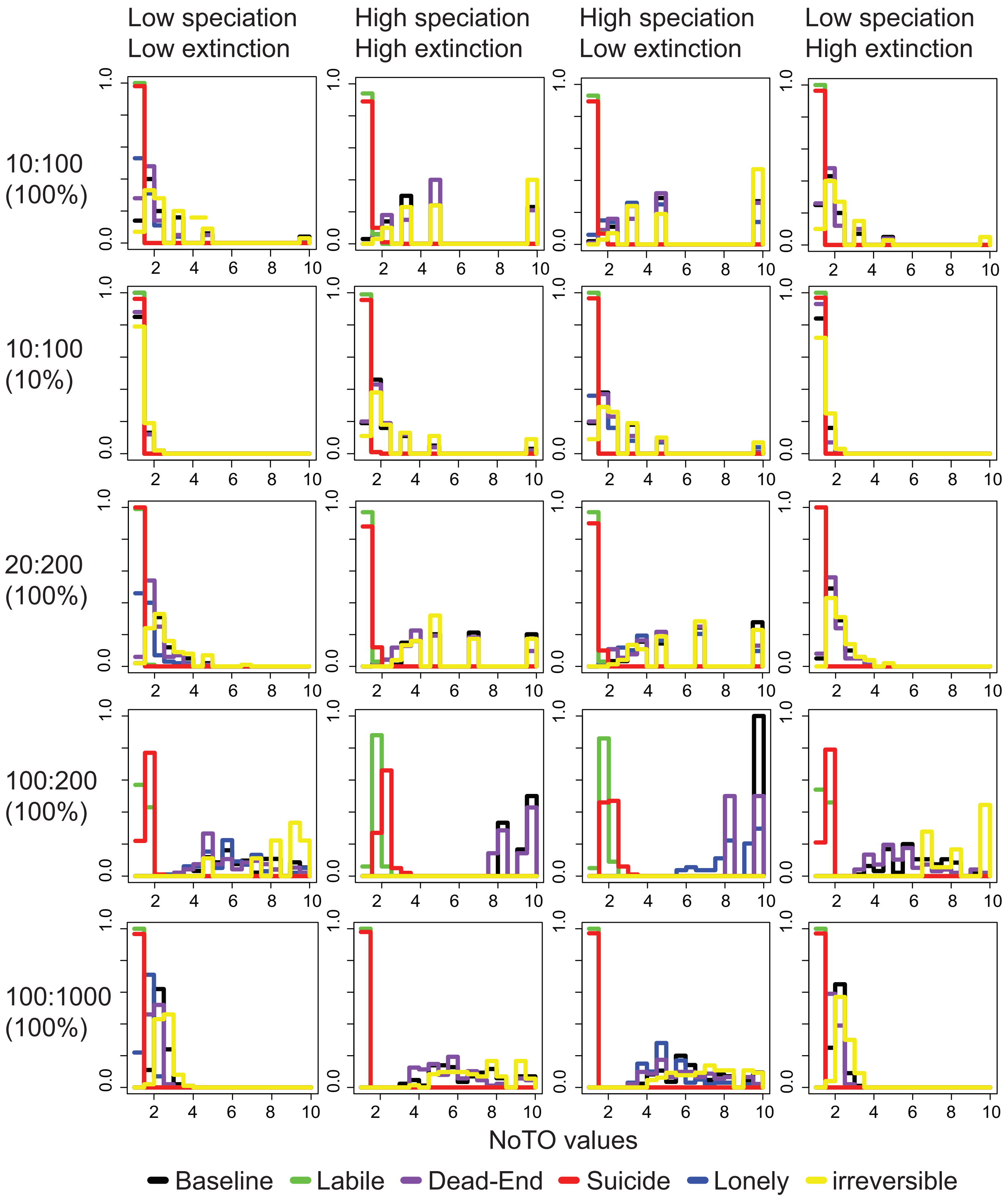

FIGURE 2. Distributions of NoTO values of the six evolutionary models described in Table 1, under four combinations of high or low speciation and extinction rates and five combinations of tree size, trait prevalence and sampling fraction. The ratio of tips with and without the trait is given on the left-hand side (with: without), and the percentage of tips from the final tree included in the data set is given in brackets below. A distribution that is distinct from others indicates that NoTO could identify its corresponding evolutionary model more reliably.

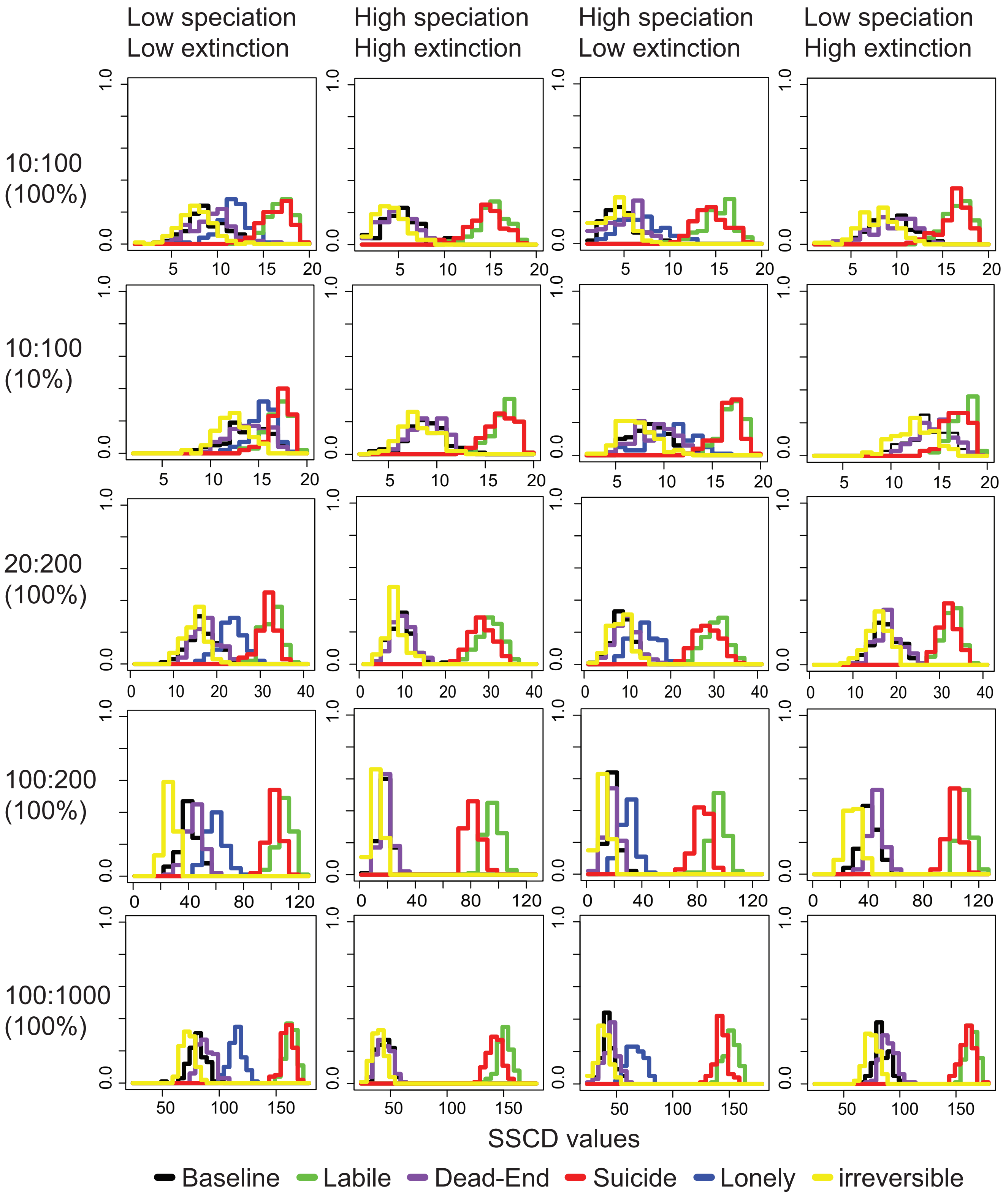

FIGURE 3. Distributions of SSCD values of the six evolutionary models described in Table 1, under four combinations of high or low speciation and extinction rates and five combinations of tree size, trait prevalence and sampling fraction. The ratio of tips with and without the trait is given on the left-hand side (with: without), and the percentage of tips from the final tree included in the data set is given in brackets below. Distribution of FPD metrics shows a similar pattern (Supplementary Fig. S5, available on Dryad at http://dx.doi.org/10.5061/dryad.10668).

is significantly larger than 0.5 when speciation rate is low or when the absolute number of tips with the trait increases (Supplementary Fig. S5, available on Dryad at http://dx.doi.org/10.5061/dryad.10668, Table 2). To identify the suicide and labile models, the significance test of FPD against zero is more powerful than the significance test of FPD against 0.5 under high speciation rate. The significance test of FPD against 0.5 is more suitable under low speciation rate because it has lower false-positive rate (Table 2).

The above results indicate that, under the range of parameter values tested in this study, the suicide and labile models are the most likely to produce a tippy pattern on the phylogeny, as indicated by significant NoTO and SSCD tests. However, when the trait transition rate is high, none of the metrics can reliably detect difference in extinction rate, or trait gain and loss rates (Supplementary Fig. S1, available on Dryad at http://dx.doi.org/10.5061/dryad.10668), presumably because the signal of association between the trait and raised extinction rate or high loss rate is broken down by the rapid rate of reversal of the trait. High extinction rate on its own (the dead-end model) does not produce a notably tippy distribution: it does so only when extinction rate is combined with high rates of trait gain (suicide model). The lonely model (reduced speciation rate) produces longer than expected tips (higher TARS) but no other indications of a tippy distribution. The irreversible model is not detected by any of these metrics.

Case Studies

Salt tolerance in grasses.—Salt tolerance is a significantly tippy trait on the Pooideae phylogeny. It has fewer tips per reconstructed origin than expected (pNoTO = 0.045) and is more scattered on the phylogeny than expected (pSSCD = 0.009; FPD = 0.71). ML values for the BiSSE parameters were estimated under a range of evolutionary constraints (Table 4). Some of these ML parameter combinations correspond to models described in Table 1: for example, model *b*, where all parameters are equal for lineages with and without the trait, is equivalent to the baseline model. Model *e* is a lonely model because speciation rate for lineages with the trait is very low. Other models describe self-destructive traits that either have a higher rate of loss than gain (*a*, *c*,*f* ,*g*) or a raised extinction rate (*d*,*h*).

By comparing the metric values calculated for the observed data to the metrics estimated for the simulated data sets under these evolutionary models, we can reject all models but two: one where salt tolerance is labile (high gain, high loss) and causes a raised extinction rate (model *d* in Table 4 and Fig. 4a), and one where salt tolerance is labile with an enhanced rate of loss, but with a lower extinction rate (model *f* in Table 4 and Fig. 4a). Therefore for the evolution of salt tolerance in Pooid grasses, we reject the baseline model (where the trait has no effect on speciation, or extinction and does not elevate loss rates). Instead, these results are compatible with salt tolerance being a self-destructive trait which is gained often but lost often either by trait reversal (model *f*) or extinction (model *d*). Under these parameter values, the significance tests on NoTO and FPD have low false-positive rates against the baseline model (model *b* in Table 5) and high power to detect self-destruction (models *a*, *c*, *d*, *f* , *g*, *h* in Table 5). These two tests cannot distinguish the self-destructive models from the lonely model. However, we would expect to identify the lonely model using TARS (Table 5), which is not significantly different from expectation for these data, therefore we reject reduced speciation rate as a good explanation for the distribution of salt tolerance in the Pooideae.

Color polymorphism in birds of prey.—Color polymorphism is significantly scattered on the phylogeny of the Accipitridae, with positive FPD (0.44), and significantly larger SSCD than expected by chance (pSSCD = 0.035). However, it does not have fewer tips per reconstructed origin than expected (pNoTO = 0.373, Fig. 2). The ML solutions for the BiSSE parameters for this data set identify models *c*,*d*, and *h* as self-destructive models where color polymorphic lineages have a higher loss rate (*c*) or higher extinction rate (*d*,*h*); models *a*,*f* , and *g* have a higher loss rate but also higher diversification rate due to color polymorphic lineages having decreased extinction (*a*,*f*) or increased speciation rate (*g*); model *e* suggests a trait that marginally increases speciation rate and *b* is the baseline (Table 4). By comparing the observed metric values to the simulated values under these evolutionary models, we can reject all models but two: the baseline model (model *b* in Table 4 and Fig. 4b), and a model where color polymorphism is easier to lose than gain, but with a lower extinction rate (model f in Table 4, Fig. 4b). So for these data, we cannot reject the null hypothesis that color polymorphism has no effect on speciation, extinction, and trait transition rates. Nonetheless, the data are also compatible with the hypothesis proposed by Hugall and Stuart-Fox (2012) that color polymorphism in birds is lost at a greater rate than is gained, and that it stimulates diversification. However, these results suggest that any increase in diversification rate in color polymorphic lineages is due to reduced extinction rate rather than a higher rate of speciation (model *f* is a better fit to the data than model *g*). Significance tests on TARS and SSCD cannot distinguish the self-destructive models (*c*,*d*,*h*) from other models (Table 5). Tests on NoTO can detect self-destructive traits, but only the test of FPD >0.5 can identify self-destructive traits that have no influence on macroevolutionary processes other than causing their own disappearance (model *c* and *d* in Table 5).

Selfing in nightshades.—Selfing has fewer tips per reconstructed origin than expected (pNoTO = 0.037) and it tends to be more scattered on the phylogeny than expected (FPD = 0.24, pSSCD = 0.067). In evolutionary models *a*–*h*, none of the ML solutions to the BiSSE parameters describe a clearly self-destructive model.

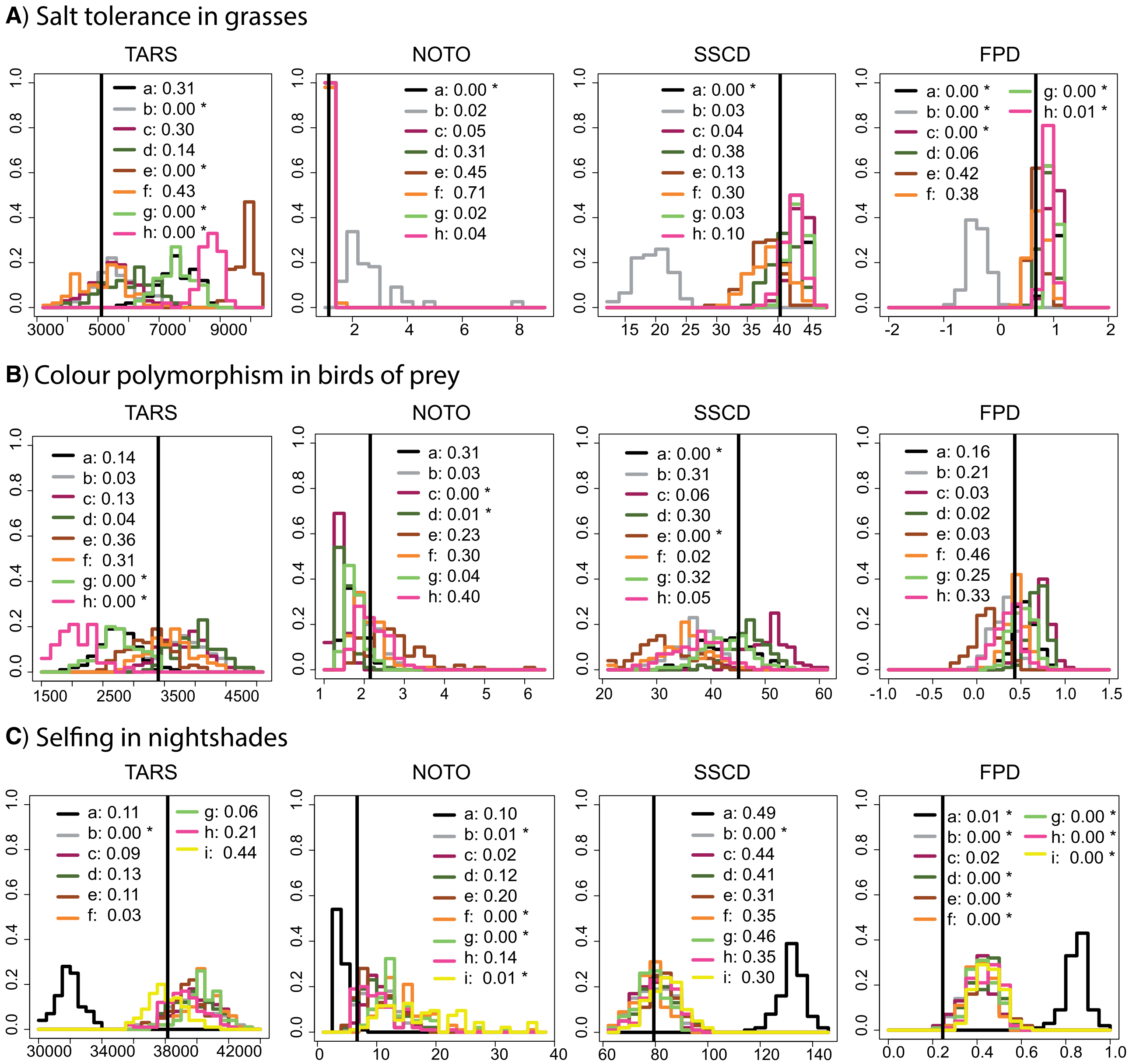

FIGURE 4. Results of the three case studies: a) salt tolerance in grasses; b) color polymorphism in birds of prey; c) selfing in nightshades. The black line represents the observed value of each of the four metrics estimated for this data set: TARS, number of tips per origin (NoTO), SSCDs and FPD. Distributions of metric values for simulated data sets under each of the models in Table 3 are plotted using different colors as indicated in the figure legend. Values in the figure legend are the proportion of the simulated data sets under the corresponding model that have metric values as extreme or more extreme than the observed. We can reject a macroevolutionary model as inconsistent with the observed data if its corresponding value in the figure legend for any of the metrics is smaller than or equal to 0.01 (marked by an asterisk) The significance level is set to 0.01 following Bonferroni correction as each macroevolutionary model is tested against the observed data using four metrics.

Most ML solutions identify either a higher rate of gain than loss (models *a*, *c*, *f*, *g*, *i*) or a higher extinction rate for lineages with the trait (models *a*,*h*, and *i*). We can reject the baseline model, where selfing has no effect on speciation, extinction, or trait loss rates, and reject all models but model *c*, where transitions from outcrossing to selfing are more frequent than the reverse (model *c* in Table 4 and Fig. 4c). Model *c* has equal speciation and extinction rates for lineages with and without the trait, and does not have an elevated trait loss rate, so it does not describe a self-destructive trait. Tests on NoTO, SSCD, and FPD have no power to detect any of these models except that the test of FPD >0.5 can reliably identify model *a*, where selfing is easier to gain than lose and induces both higher speciation and extinction rates (Table 5). If the parameter values inferred by BiSSE for this data set are correct, then the tippiness of selfing is primarily generated by high trait transition rate. A high trait transition rate could reduce the power to distinguish different models because any signal of a link between the

TABLE 5. Statistical power of significance tests on the four metrics in the three case studies

Trait	Models	pTARS ≥0.95	pNoTO ≤0.05	pSSCD ≤0.05	FPD >0	FPD >0.5
Salt tolerance						
	a)	**0.90**	**1**	**1**	**1**	**1**
	b)	**0.04**		0.86	**0.01**	0
	c)	0.03	**1**	**1**	**1**	**1**
	d)	0.15	**0.99**	**1**	**1**	**1**
	e)	**1**	**0.88**	**1**	**1**	**0.92**
	f)	0.02	0.77	**1**	**1**	**0.89**
	g)	**0.84**	**1**	**1**	**1**	**1**
	h)	**1**	**0.99**	**1**	**1**	**1**
Color polymorphism						
	a)	**1**	**0.91**	**1**	**1**	0.61
	b)	0.97	0.36	1	0.99	**0.05**
	c)	**1**	**0.99**	**1**	**1**	**0.95**
	d)	**0.93**	**0.99**	**1**	**1**	**0.97**
	e)	**1**	0.10	**1**	**0.81**	**1**
	f)	**1**	**0.96**	**1**	**1**	0.28
	g)	**1**	**0.86**	**1**	**1**	0.55
	h)	**1**	0.38	**1**	**1**	0.12
Selfing						
	a)	**1**	**1**	**1**	**1**	0
	b)	0.89	0.56	1	1	0.07
	c)	**0.85**	0.60	**1**	**1**	0.08
	d)	**0.91**	0.46	**1**	**1**	0.17
	e)	**0.94**	0.38	**1**	**1**	0.19
	f)	**0.82**	**0.83**	**1**	**1**	0.06
	g)	**0.87**	0.72	**1**	**1**	0.13
	h)	0.09	0.39	**1**	**1**	0.22
	i)	0	**0.87**	**1**	**1**	0.23

Notes: Each value is the proportion of data sets simulated under different macroevolutinary models in Table 3 that have pTARS ≥ 0.95, pNoTO ≤ 0.05, pSSCD ≤ 0.05, FPD >0, or FPD >0.5. Values under the baseline model (model b) show the false-positive rate of these significance tests. Values under each of the other evolutionary models show the statistical power of the significance tests when the model is treated as the alternative hypothesis. False-positive rate ≤ 0.05 and power ≥ 0.8 are highlighted in bold to help detecting conditions under which each metric can reliably detect the alternative model.

trait and higher rate of loss or extinction is effectively erased by the frequency of trait transition.

DISCUSSION

Studies of patterns of traits on phylogenies have provided an empirical approach that, to some extent, side-steps theoretical debates about whether natural selection can be said to act at multiple levels. Phylogenetic studies of relative representation of lineages can identify consistent biases in net diversification that indicate that some traits have a predictable effect on lineage success (e.g., Owens et al. 1999; Cardillo et al. 2003; Phillimore et al. 2006; Kraaijeveld et al. 2011). Traits that are shallow and scattered on the phylogeny have often been considered to indicate that their short-term benefits (so they are gained often) are ultimately overtaken by long-term costs (so they do not persist). But in order to use tippy patterns to detect these macroevolutionary phenomena, we need a clear expectation about the kinds of processes that can lead to tippy patterns, and we need a way of detecting when a pattern is significantly more tippy than we would expect in the absence of those phenomena.

In this article, we tested four metrics that have the potential to detect a tippy pattern on a phylogeny, and we explored the conditions under which a tippy pattern can be generated. We applied a novel method for phylogeny simulation to efficiently estimate the power of the four metrics to reveal the macroevolutionary mechanisms causing the tippy pattern. We then applied these metrics to investigate whether we can detect real-world instances of macroevolutionary self-destruction by identifying phylogenetic patterns consistent with traits that have evolved often but have an enhanced rate of extinction or loss or both. We have shown that considering both the relative number of reconstructed origins and the phylogenetic scatter of a trait can indicate a self-destructive trait that is gained often, then rapidly lost through extinction or reversal. However, we have also shown that the tippy pattern on phylogenies is strongly influenced by trait lability, and thus is not in itself a reliable indicator of a trait that increases the extinction rate of the lineage.

Metrics of Tippiness

Age of species has frequently been invoked as an indicator of a trait that confers an increased extinction rate, particularly with reference to asexuality (Neiman and Schwander 2011). Our findings cast doubt on the general applicability of tip length as an indicator of a trait that causes its own disappearance. TARS, which compares the relative distribution of lengths for tips with the trait of interest to those without, did not identify any of the models where the trait causes a raised rate of extinction or loss. Our results are consistent with the failure to find a clear and consistent signal of reduced tip length in phylogenetic studies of asexual taxa (Schwander and Crespi 2009). However, traits that reduce the speciation rate (lonely model) have consistently longer tip lengths than lineages without the trait, so may be identified by significantly higher TARS measures. TARS was not significantly different from null expectation for any of the three case studies.

Tip length does not provide a clear signal of self-destructive traits in macroevolution because it reflects both the speciation rate and extinction rate. If a lineage has a high speciation rate, tips will tend to be shorter because they are divided by speciation events. If a lineage has a high extinction rate, tips will tend to be shorter because they do not persist for long (though on an ultrametric tree, high extinction rate may increase the length of terminal edges). For example, the mean age of asexual species could be reduced by a process of frequent production of asexual populations from a sexual progenitor, even in the absence of any disadvantage limiting the persistence of asexual clones

(Janko et al. 2008). It should also be noted that tip length is only partly determined by the age of the divergence from the closest relative of the species in the phylogeny: it is also determined by the rate of molecular evolution, which may be affected by many different species traits (Bromham 2009), some of which may vary with the trait of interest (Schwander and Crespi 2009). For example, if a self-destructive trait such as selfing decreases the effective population size, then it may result in increased substitution rates, lengthening tips. So while consistent differences in the length of tips with different traits may be useful for exploring some interesting macroevolutionary patterns, it does not seem to be a useful guide for finding the class of self-destructive traits that are our focus here. As stated in a review on the macroevolutionary patterns of asexual lineages "Ideally, the age distribution of asexual lineages could be compared with the distributions that are predicted by theoretical models … So far, both predictions and observations fall well short of the goal of quantitative comparison" (Butlin 2002).

The NoTO metric and the degree of phylogenetic clustering, as reflected in SSCD and FPD, are more reliable and informative measures of tippy patterns that are caused by self-destructive traits that are gained often and lost often due to raised extinction rate or high trait loss. We demonstrate that a trait can be said to be significantly tippy on a phylogeny if pNoTO and pSSCD approaches 0, and if FPD approaches 1. However, none of the metrics can reliably identify the dead-end model, where a trait increases the extinction rate. This finding is consistent with other studies that have suggested that estimating extinction rate from phylogenies is problematic (Morlon et al. 2010; Rabosky 2010; Stadler 2011).

While we cannot generalize our results beyond the BiSSE models and the parameter space tested or set a fixed significance level over all combinations of values for speciation rate, extinction rate, and trait transition rate, we find that for the models and the parameter combinations we examined, we can detect self-destructive traits with reasonable power by identifying traits that have a significantly lower NoTO, a significantly higher SSCD, and a value of FPD greater than 0.5. These methods are able to detect traits that influence lineage persistence and diversification on relatively small phylogenies (<300 tips), which may provide an advantage over some other phylogenetic methods (Davis et al. 2013). The power of these tests on NoTO, SSCD, and FPD to distinguish the suicide and labile models is mediated by speciation rate, trait transition rate, and the number of sampled lineages with the self-destructive trait. While the power to correctly reject the suicide and labile models increases as speciation rate decreases (and decreases with incomplete sampling), the power to correctly reject the other models increases as speciation rate increases. The power of all tests increases when the absolute number of lineages with the trait increases, and decreases with the increasing trait lability.

Case Studies

The present study was stimulated by our observation that salt tolerance has a surprisingly tippy distribution in many flowering plant families (Bromham 2014; Moray et al. 2015). While we hypothesized that this tippy pattern could be due to the frequent gain and loss of a trait that carried short-term benefits but long-term costs, we lacked the means to provide a formal test of this hypothesis (Bromham 2014). We have now developed tools that allow us to ask whether a pattern is significantly more tippy than we would expect by chance, and to look for models of macroevolution that are compatible with the observations. Our analyses in this article support the hypothesis that salt tolerance is grasses is a significantly tippy trait, with more origins than would be expected under a Brownian motion model of trait evolution, and significantly scattered on the phylogeny. The models that provide the best explanation of the distribution of salt tolerance in the Pooideae are a high rate of gain and even higher rate of loss, or high rate of gain and loss and a raised extinction rate, suggesting that salt tolerance evolves often but is rarely maintained in grass lineages over long evolutionary time periods.

Our results are compatible with the suggestion that the phylogenetically scattered pattern of color polymorphism in birds of prey could indicate that it is a self-destructive trait that is gained often and lost often, consistent with the hypothesis that the loss of color polymorphism accompanies speciation events (Hugall and Stuart-Fox 2012). However, we also note that the same trait distribution could be generated without a self-destructive macroevolutionary process and power analyses suggest that the significant result of the test on SSCD and the positive FPD value cannot be used as evidence for self-destruction, as there is a risk of false-positive results, due to a high chance of rejecting a baseline model. We find no evidence to support a higher rate of speciation in color polymorphic lineages, though a model where color polymorphic lineages have a reduced extinction rate cannot be rejected, providing an alternative explanation of higher diversification rates in color polymorphic birds of prey.

In contrast, we do not find strong evidence that the significantly tippy distribution of selfing in the Solanaceae is likely to have been generated by macroevolutionary self-destruction. The unconstrained model, where all parameters are free to vary, reflects the conclusions of Goldberg et al. (2010) that selfing has a high rate of gain and a high speciation rate, but also a high extinction rate. However, this model is not a significantly better fit to the data from one where selfing has a high rate of gain but has no effect on speciation or extinction rates. Power analyses also suggest that the significant results of tests using these metrics (NoTO, SSCD, FPD) may not have the power to discriminate a self-destructive model for selfing in the Solanaceae, due to the high trait transition rate estimated by BiSSE. However, this interpretation of the results relies on correct inferences of BiSSE parameter values, which will be compromised if the assumptions of the BiSSE model are violated for this case study by state changes being linked to speciation events (i.e., selfing accompanies speciation). The Binary State Speciation and Extinction Node-Enhanced State Shift (BiSSE-NESS) model may be a more appropriate method in this case (Magnuson-Ford and Otto 2012).

Our results suggest that when we have three indicators of tippiness—significantly lower NoTO, a significantly higher SSCD, and a value of FPD greater than 0.5—we have probably identified a trait that causes its own disappearance through raised extinction or high trait loss. But we may seek different explanations if only one or two of these measures are significant. In these case studies, we identified salt tolerance in grasses as a macroevolutionary self-destructive trait, but in the other two case studies the situation is more complicated. Color polymorphism in birds of prey does have a high loss rate, but it also stimulates diversification. Selfing in Solanaceae is highly labile but does not seem to have a significantly enhanced loss rate or extinction rate. In cases where less than three of the indicators of tippiness are significant, parameter estimation under a range of evolutionary models and comparison between the observed metric values to simulated values under these evolutionary models may help identify the underlying macroevolutionary processes.

However, estimating evolutionary models using model-fitting approaches such as the BiSSE method (Maddison et al. 2007) may produce spurious associations between trait state and diversification rates due to phylogenetic pseudoreplication and violation of model assumptions (Rabosky and Goldberg 2015). Because they do not treat independent shifts in trait state as replications, these approaches may inadvertently infer significant association between trait state and diversification rates even if there is a single shift in diversification rate that co-occurs with a trait state (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015). For a tippy trait that is scattered on the phylogeny and has few tips per reconstructed origin, pseudoreplication may be less of a problem because it is unlikely that the many independent origins of the trait would co-occur with shifts in diversification by chance (Rabosky and Goldberg 2015). But, this is not true when other model assumptions are violated. For example, Rabosky and Goldberg (2015) showed that simulating labile traits on a phylogeny simulated under a pure-birth model tends to decrease the type I error of the BiSSE method, but simulating labile traits on an observed phylogeny tends to increase the type I error.

The use of model-fitting approaches in macroevolutionary biology requires some way of assessing the fit of the model to the data. Selecting the best-fitting model does not guarantee an adequate description of the processes that generated the observed data. Gatesy (2007) used the memorable analogy of a large man shopping for underwear in the petites department of a women's clothing store: the best-fit may simply not fit well enough. Similarly, choosing the best-fit model from a likelihood analysis, such as a BiSSE method, does not tell you whether the model is a good description of the processes that produced the data, only that it is the best of the models tested. A comparison of the relative fit of models can lead to the rejection of all models but one, but ideally we need a method that allows us to reject all models as inadequate, to avoid accepting a best-fit model that is a poor fit to the data.

There are several ways to formally test model inadequacy. The "backward" test starts with the observed frequency of traits for a given number of taxa, then uses a range of macroevolutionary models, specifying speciation, extinction, and trait transition rates for lineages with and without the trait, to simulate the evolution of phylogenies that would lead to that pattern of taxa (Hua and Bromham in press). We can use these simulated phylogenies to generate a distribution of metric values expected under any given macroevolutionary model for an observed set of taxa. We can then ask whether the metric value from the observed phylogeny falls within this range. If it does not, we conclude that the model is unlikely to have produced the observed data. Using multiple phylogenetic metrics that describe different aspects of tree shape will increase the power of the backward test. In this way we can reject all models as incompatible with the data, rather than having to accept the best-fit model.

However, the backward test is limited to considering phylogenies with the observed number of tips with and without the trait. Ideally, we also want to ask how likely it is that the model would produce the observed tips, without constraining the outcome. We can use a "forward" test to explore the range of outcomes we could expect under a given macroevolutionary model. By simulating tree growth under a given model, we can ask how often it produces the observed number of tips with and without the trait. If the forward simulations rarely or never produce a tree that matches the observed tip states, then we can conclude that the model is unlikely to describe the process that produced our data.

Both the forward and backward tests of model inadequacy give us the opportunity to reject the best-fit model as an unlikely explanation of the data. However, when we have more than one model that we cannot reject, we can use likelihood to compare the relative fit of the non-rejected models, using standard likelihood comparisons such as the Aikake Information Criterion. If a given model provides a good fit using one metric but a poor fit using the others, then we might reject that model as inconsistent with the data.

We can further minimize the influence of statistical artifacts on the inference of macroevolutionary mechanisms by applying the statistical test to the same trait in multiple independent clades (Rabosky and Goldberg 2015). Detecting the same tippy pattern of a particular trait in many different clades would strengthen the conclusions concerning macroevolutionary processes, such as the identification of traits under negative lineage selection. The metric-based tests we have proposed may be a more efficient way of detecting macroevolutionary self-destructive traits than model-fitting approaches, and have the added advantage of using common statistical measures (pTARS, pNoTO, pSSCD, and FPD) that are comparable between studies in a meta-analysis.

CONCLUSIONS

We have demonstrated that phylogenetic analyses can be used to identify self-destructive traits, which have a consistently negative impact on the persistence and diversification of lineages with that trait. While tippy patterns are often taken as a sign of a suicidal trait that increases the extinction rate, or a costly trait that is frequently lost, the extent to which either of these processes can be reliably detected using tippy phylogenetic patterns has not previously been tested. We show that calculating the number of tips per inferred origin, combined with measures of phylogenetic scatter, can identify "self-destructive" traits that are gained often then rapidly lost, either by extinction or trait reversal. When combined with macroevolutionary modeling, we may be able to distinguish these two alternative versions of macroevolutionary self-destruction in some cases. The simple measures we present in this article have the advantages that they are easy to calculate on any phylogeny and are comparable between phylogenies, and yet when combined with simulations that provide appropriate null models of trait distributions, these metrics provide reasonably powerful tests of traits that have a consistently negative influence on their own chances of persistence. We hope that the methods we have developed will make it easier to compare different possible macroevolutionary explanations for interesting trait distributions on phylogenies.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.10668.

FUNDING

This work was supported by the Australian Research Council.

REFERENCES

Agnarsson I., Avilés L., Coddington J.A., Maddison W.P. 2006. Sociality in Theridiid spiders: Repeated origins of an evolutionary dead end. Evolution 60:2342–2351.

Anacker B.L., Whittall J.B., Goldberg E.E., Harrison S.P. 2011. Origins and consequences of serpentine endemism in the California flora. Evolution 65(2):365–376.

Batalha M.A., Silva I.A., Cianciaruso M.V., de Carvalho G.H. 2011. Trait diversity on the phylogeny of cerrado woody species. Oikos 120(11):1741–1751.

Beck J.B., Windham M.D., Pryer K.M. 2011. Do asexual polyploid lineages lead short evolutionary lives? A case study from the fern genus *Astrolepis*. Evolution 65(11):3217–3229.

Bennett T.H., Flowers T.J., Bromham L. 2013. Repeated evolution of salt-tolerance in grasses. Biol. Lett. 9(2):20130029.

Blomberg SP, Garland T. Jr., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution 57(4):717–745.

Bromham L. 2009. Why do species vary in their rate of molecular evolution? Biol. Lett. 5:401–404.

Bromham L. 2014. Macroevolutionary patterns of salt tolerance in angiosperms. Ann. Bot. 115(3):333–341.

Butlin R. 2002. The costs and benefits of sex: New insights from old asexual lineages. Nat. Rev. Genet. 3(4):311–317.

Cardillo M., Huxtable J.S., Bromham L. 2003. Geographic range size, life history and rates of diversification in Australian mammals. J. Evol. Biol. 16:282–288.

Cieslak A., Fresneda J., Ribera I. 2014. Life-history specialization was not an evolutionary dead-end in Pyrenean cave beetles. Proc. R. Soc. B Biol. Sci. 281(1781):20132978.

Colles A., Liow L.H., Prinzing A. 2009. Are specialists at risk under environmental change? Neoecological, paleoecological and phylogenetic approaches. Ecol. Lett. 12(8):849–863.

Davis M.P., Midford P.E., Maddison W. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evol. Biol. 13(1):38.

De Vos J.M., Joppa L.N., Gittleman J.L., Stephens P.R., Pimm S.L. 2015. Estimating the normal background rate of species extinction. Conserv. Biol. 29(2):452–462.

Edwards E.J., Smith S.A. 2010. Phylogenetic analyses reveal the shady history of C4 grasses. Proc. Natl Acad. Sci. U. S. A. 107(6):2532–2537.

Felsenstein J. 2005. Using the quantitative genetic threshold model for inferences between and within species. Phil. Trans. R. Soc. Lond. B 360:1427–1434.

FitzJohn R.G. 2012. Diversitree: Comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3(6):1084–1092.

Fritz S.A., Purvis A. 2010. Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. Conserv. Biol. 24:1042–1051.

Futuyma D.J., Moreno G. 1988. The evolution of ecological specialization. Annu. Rev. Ecol. Syst. 19(1):207–233.

Gatesy J. 2007. A tenth crucial question regarding model use in phylogenetics. Trends Ecol. Evol. 22:509.

Goldberg E.E., Roy K., Lande R., Jablonski D. 2005. Diversity, endemism, and age distributions in macroevolutionary sources and sinks. Am. Nat. 165(6):623–633.

Goldberg E.E., Kohn J.R., Lande R., Robertson K.A., Smith S.A., Igic B. 2010. Species selection maintains self-incompatibility. Science 330(6003):493–495.

Gyllenberg M., Parvinen K. 2001. Necessary and sufficient conditions for evolutionary suicide. Bull. Math. Biol. 63(5): 981–993.

Hartley C.J., Newcomb R.D., Russell R.J., Yong C.G., Stevens J.R., Yeates D.K., La Salle J., Oakeshott J.G. 2006. Amplification of DNA from preserved specimens shows blowflies were preadapted for the rapid evolution of insecticide resistance. Proc. Natl Acad. Sci. U. S. A. 103(23):8757–8762.

Helanterä H., Strassmann J.E., Carrillo J., Queller D.C. 2009. Unicolonial ants: Where do they come from, what are they and where are they going? Trends Ecol. Evol. 24(6):341–349.

Hua X., Bromham L. in press. Phylometrics: An R package for detecting macroevolutionary patterns, using phylogenetic metrics and backward tree simulation. Methods Ecol. Evol.

Hugall A.F., Stuart-Fox D. 2012. Accelerated speciation in colour-polymorphic birds. Nature 485:631–634.

Igic B., Busch J.W. 2013. Is self-fertilization an evolutionary dead end? New Phytol. 198(2):386–397.

Jablonski D. 2008. Species selection: Theory and data. Annu. Rev. Ecol. Evol. Syst. 39:501–524.

Janko K. 2014. Let us not be unfair to asexuals: Their ephemerality may be explained by neutral models without invoking any evolutionary constraints of asexuality. Evolution 68(2):569.

Janko K., Drozd P., Flegr J., Pannell J.R. 2008. Clonal turnover versus clonal decay: A null model for observed patterns of asexual longevity, diversity and distribution. Evolution 62(5):1264–1270.

Johannesen J., Lubin Y., Smith D.R., Bilde T., Schneider J.M. 2007. The age and evolution of sociality in *Stegodyphus* spiders: A molecular phylogenetic perspective. Proc. Roy. Soc. Lond. B Biol. Sci. 274(1607):231–237.

Kokko H., Heubel K.U. 2011. Prudent males, group adaptation, and the tragedy of the commons. Oikos 120(5):641–656.

Kölsch G., Pedersen B.V. 2008. Molecular phylogeny of reed beetles (Col., Chrysomelidae, Donaciinae): The signature of ecological specialization and geographical isolation. Mol. Phylogenet. Evol. 48(3):936–952.

Kraaijeveld K., Kraaijeveld-Smit F.J., Maan M.E. 2011. Sexual selection and speciation: The comparative evidence revisited. Biol. Rev. 86(2):367–377.

Luo Z.-X. 2007. Transformation and diversification in early mammal evolution. Nature 450:1011–1019.

Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56(5): 701–710.

Maddison W.P., FitzJohn R.G. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64(1):127–136.

Magnuson-Ford K., Otto S.P. 2012. Linking the investigations of character evolution and species diversification. Am. Nat. 180(2):225–245.

Maynard Smith J. 1978. The evolution of sex. Cambridge: Cambridge University Press.

Moray C., Hua X., Bromham L. 2015. Salt tolerance is evolutionarily labile in a diverse set of angiosperm families. J. Evol. Biol. 15:90.

Morlon H., Potts M.D., Plotkin J.B. 2010. Inferring the dynamics of diversification: A coalescent approach. PLoS Biol. 8(9):e1000493.

Neiman M., Meirmans S., Meirmans P.G. 2009. What can asexual lineage age tell us about the maintenance of sex? Ann. N.Y. Acad. Sci. 1168(1):185–200.

Neiman M., Schwander T. 2011. Using parthenogenetic lineages to identify advantages of sex. Evol. Biol. 38(2):115–123.

Owens I.P.F., Bennett P.M., Harvey P.H. 1999. Species richness among birds: Body size, life history, sexual selection or ecology? Proc. Roy. Soc. Lond. B 266:933–939.

Parvinen K. 2005. Evolutionary suicide. Acta Biotheor. 53(3):241–264.

Phillimore A.B., Freckleton R.P., Orme C.D.L., Owens I.P. 2006. Ecology predicts large-scale patterns of phylogenetic diversification in birds. Am. Nat. 168(2):220–229.

Pruitt J.N. 2013. A real-time eco-evolutionary dead-end strategy is mediated by the traits of lineage progenitors and interactions with colony invaders. Ecol. Lett. 16(7):879–886.

Purvis A. 2008. Phylogenetic approaches to the study of extinction. Annu. Rev. Ecol. Syst. 39:301–319.

Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64(6):1816–1824.

Rabosky D.L., McCune A.R. 2010. Reinventing species selection with molecular phylogenies. Trends Ecol. Evol. 25(2):68–74.

Rabosky D.L., Santini F., Eastman J., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 4:1958.

Rabosky D.L., Goldberg E.E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64(2):340–355.

Rankin D.J., Lopez-Sepulcre A., Foster K.R., Kokko H. 2007. Species-level selection reduces selfishness through competitive exclusion. J. Evol. Biol. 20(4):1459–1468.

Revell L.J., Harmon L.J., Collar D.C. 2008. Phylogenetic signal, evolutionary process, and rate. Syst. Biol. 57(4):591–601.

Schneider J., Michalik P. 2011. One-shot genitalia are not an evolutionary dead end—Regained male polygamy in a sperm limited spider species. BMC Evol. Biol. 11(1):197.

Schwander T., Crespi B.J. 2009. Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. Mol. Ecol. 18(1):28–42.

Simon M.F., Grether R., de Queiroz L.P., Skema C., Pennington R.T., Hughes C.E. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. Proc. Natl Acad. Sci. U. S. A. 106(48):20359–20364.

Stadler T. 2011. Simulating trees with a fixed number of extant species. Syst. Biol. 60(5):676–684.

Stanley S.M. 1979. Macroevolution: Pattern and process. San Fransisco: W. H. Freeman and Company.

Stebbins G.L. 1957. Self fertilization and population variability in the higher plants. Am. Nat. 91(861):337–354.

Tripp E.A., Manos P.S. 2008. Is floral specialization an evolutionary dead-end? Pollination system transitions in *Ruellia* (Acanthaceae). Evolution 62(7):1712–1737.

Weir J.T., Schluter D. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. Science 315(5818):1574–1576.

Williams G.C. 1975. Sex and evolution. Princeton: Princeton University Press.

Williams G.C. 1992. Natural selection: Domains, levels and challenges. Oxford: Oxford University Press.

Wood T.E., Burke J.M., Rieseberg L.H. 2005. Parallel genotypic adaptation: When evolution repeats itself. Genetica 123:157–170.

Wright S.I., Kalisz S., Slotte T. 2013. Evolutionary consequences of self-fertilization in plants. Proc. Roy. Soc. B Biol. Sci. 280(1760):20130133.