

ON PHYLOGENETIC TESTS OF IRREVERSIBLE EVOLUTION

Emma E. Goldberg^{1,2,3} and Boris Igic^{3,4,5}

¹Department of Biology, Biology/Psychology Bldg. #144, University of Maryland, College Park, Maryland 20742

²E-mail: eeg@umd.edu

⁴Department of Biological Sciences, 840 West Taylor St., MIC 067, University of Illinois at Chicago, Chicago, Illinois 60607

⁵E-mail: boris@uic.edu

Received January 14, 2008

Accepted July 22, 2008

“Dollo’s law” states that, following loss, a complex trait cannot reevolve in an identical manner. Although the law has previously fallen into disrepute, it has only recently been challenged with statistical phylogenetic methods. We employ simulation studies of an irreversible binary character to show that rejections of Dollo’s law based on likelihood-ratio tests of transition rate constraints or on reconstructions of ancestral states are frequently incorrect. We identify two major causes of errors: incorrect assignment of root state frequencies, and neglect of the effect of the character state on rates of speciation and extinction. Our findings do not necessarily overturn the conclusions of phylogenetic studies claiming reversals, but we demonstrate devastating flaws in the methods that are the foundation of all such studies. Furthermore, we show that false rejections of Dollo’s law can be reduced by the use of appropriate existing models and model selection procedures. More powerful tests of irreversibility require data beyond phylogenies and character states of extant taxa, and we highlight empirical work that incorporates additional information.

KEY WORDS: Acquisition bias, AIC, ancestral state reconstruction, character evolution, diversification rate, Dollo’s law, extinction, irreversibility, phylogenetic methods, speciation.

An organism never returns exactly to a former state, even if it finds itself placed in conditions of existence identical to those in which it has previously lived.

Louis Dollo (1893)

Often termed “Dollo’s law,” the proposition that organisms never revert to a former evolutionary state was controversial since its inception (see Gould 1970 for a detailed review). The original formulation of Dollo’s law is so broad as to be of limited use (Simpson 1953; Hennig 1966; Bull 2000), and a narrower version—the irreversible loss of single complex characters—is almost exclusively the concept considered in the literature (Muller 1939; Simpson 1953; Kohlsdorf and Wagner 2006). Much of the debate about Dollo’s law centered on the criteria for sufficient complexity, the metrics for identifying whether mutations are

true reversals or merely analogs (e.g., homologous reversion vs. analogous novelty, exact nucleotide substitution reversal vs. compensatory mutation), speculative estimates of the general probabilities of reversal, and the status of biological laws. These are important issues, and they are reviewed elsewhere (Gould 1970; Wagner 1982; Marshall et al. 1994; McIntyre 1997). We instead focus on phylogenetic methods for testing Dollo’s law.

The broad availability of sequence data for construction of accurate phylogenies and the development of a quantitative framework for inference of character evolution opened up new avenues for testing Dollo’s law. Early phylogenetic studies of irreversibility based on parsimony reconstructions (Hennig 1966; reviewed in Maddison and Maddison 1992) often initially inferred reversal, but then found that this conclusion could be overturned by even a modest asymmetry in the difficulty of gain of a complex state over its loss (Cunningham et al. 1998; Cunningham 1999). Paradoxically, the limitations of parsimony methods (see

³ Both authors contributed equally to this article.

Harvey and Pagel 1991; Cunningham et al. 1998) were in some ways advantageous in tests of Dollo's law in that they elicited explicit statements about assumptions, critical interpretation of results, and adjustments of methods (Kohn et al. 1996; Wray 1996; Omland 1997; Lee and Shine 1998; Cunningham 1999). The statistical model-based methods for character change in a maximum likelihood (Felsenstein 1981; Harvey and Pagel 1991; Sanderson 1993; Pagel 1994; Schluter et al. 1997) or Bayesian (Huelsenbeck et al. 2000, 2003; Pagel et al. 2004) framework were subsequently introduced and widely adopted. Applications of these methods to test Dollo's law recently yielded several spectacular claims of reversion to complex states (e.g., Oakley and Cunningham 2002; Collin and Cipriani 2003; Whiting et al. 2003; Nosil and Mooers 2005; Cruickshank and Paterson 2006; Domes et al. 2007; Ferrer and Good-Avila 2007; Brandley et al. 2008). Accordingly, criticism of Dollo's law shifted from debate about what constitutes a reversal to purported evidence of true reversals, leading to the prevailing view that Dollo's law was invalidated through the use of phylogenetic methods (recently reviewed in Pagel 2004; Kohlsdorf and Wagner 2006; Domes et al. 2007).

Here, we show that phylogenetic tests of Dollo's law are frequently misled by violations of at least two standard model assumptions. First, reconstructions are almost exclusively attempted only on clades that are variable at the focal character, leading to "acquisition bias" (Felsenstein 1992; Frumhoff and Reeve 1994; Lewis 2001) and inappropriate assignment of the character state distribution at the root. Second, association of character states with different net diversification rates (species-level selection) can lead to a strong bias in both transition rate and ancestral state estimation (Janson 1992; Strathmann and Eernisse 1994; Oakley and Cunningham 2000; Igic et al. 2006; Maddison 2006; Maddison et al. 2007; Paradis 2008). By applying existing models of character evolution (Pagel 1994; Lewis 2001; Mk2, Maddison et al. 2007, BiSSE) to simulated trees, we show that commonly used methods frequently incorrectly reject Dollo's law, but that more appropriate model comparisons do not. We also reanalyze data from two empirical studies that rejected irreversibility, and we discuss how extensions of phylogenetic methods and incorporation of additional data may improve tests of directionality in character evolution and ancestral state reconstructions.

Unidirectional Character Evolution: A Thought Experiment

To motivate the simulation results we present below, we begin by describing a thought experiment concerning an arbitrary binary character that is subject to Dollo's law. The two possible states are denoted A and B, transitions from A to B occur at some positive rate, and transitions from B to A cannot occur. We first consider the evolution of such a character whose states do not differently

affect speciation or extinction rates, so that lineages in states A and B have equal net diversification rates. We then examine the implications for inference of character evolution when this restriction is removed.

STATE-INDEPENDENT DIVERSIFICATION

A clade containing a character for which A-to-B changes are irreversible can only show a mix of the two states at the tips when the root was in state A (Fig. 1. If the root state were B, the entire history of the lineage and all of its tips would simply remain in state B.) If the character states A and B are not associated with different diversification rates, the equilibrium proportions of states A and B are, however, 0 and 1, respectively. This discrepancy is the source of many of the problems we describe.

Computing the likelihood of the character states at the tips of a tree under the standard continuous-time Markov model of character evolution (Mk2 for a binary character such as presence or absence of a feature; Pagel 1994; Lewis 2001)

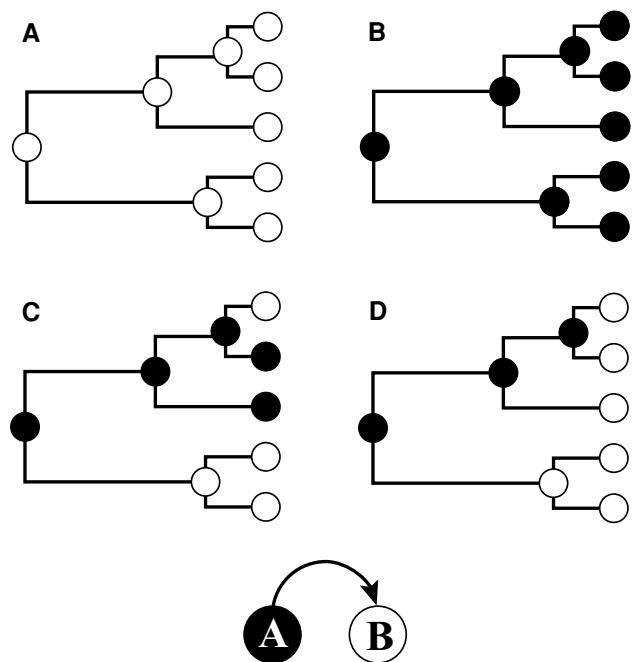


Figure 1. Inferences about character evolution require both states (A and B here) to be present in extant species. Assuming unidirectional transitions from A to B, there are four possible outcomes. (a) When the root is in state B, all tips are B. When the root is in state A, the tip states may be (b) all A, (c) some A and some B, or (d) all B. Therefore, in order for both states to be represented in the extant taxa, the root must be A. Under state-independent diversification, the outcome is determined by the product of the transition rate and elapsed time: to obtain (c), the product of the clade age and transition rate cannot be very large or very small. Under state-dependent diversification, the outcome is also affected by differences in state-specific speciation and extinction rates.

to obtain maximum-likelihood (ML) estimates of the transition rates (Felsenstein 1981; Pagel 1994; Schluter et al. 1997) is easily achieved with any of several software packages such as Discrete/Multistate/BayesTraits (Pagel et al. 2004) or Mesquite (Maddison and Maddison 2007). These calculations require an assumption about the probability of each character state at the root, which may or may not be explicitly specified by the user. Stationary probabilities are often used, based on the assumption of equilibrium in the state frequencies (Felsenstein 1981). But in the case of unidirectional evolution, this is likely to lead to erroneous conclusions because the root probability will be incorrectly weighted to state B. Although other assumptions about the root state are possible, they are either arbitrary (e.g., equal weights for the two states) or must rely on additional evidence (Pagel 1999).

Because it is impossible to obtain meaningful ML estimates of transition rates or ancestral states in a clade in which all extant taxa share the same state (see Schultz et al. 1996; Schultz and Churchill 1999), the logical requirement of a root state of A must be incorporated into tests of Dollo's law (Nosil and Mooers 2005). We address the implications of the root state assumption quantitatively in our simulation study below, and we explain how a simple change in the model selection procedure can greatly improve test accuracy.

STATE-DEPENDENT DIVERSIFICATION

Tests of character evolution require the presence of both states of a binary character, but under unidirectional transitions, one state (A in our example) is expected to become vanishingly rare. If the net diversification (speciation minus extinction) rate is sufficiently greater for lineages in state A than in state B, however, the equilibrium frequency of state A can be nonzero (Nunney 1989). Such state-dependent diversification is therefore likely to play an important role in the maintenance of a character subject to Dollo's law, but it can lead to incorrect estimates of transition rates (Maddison 2006). In our example, greater diversification of state A would lead to overestimation of the B-to-A transition rate and an incorrect rejection of Dollo's law (Fig. 2).

Many characters put forth as examples of Dollo's law (Bull and Charnov 1985; Pagel 2004) can reasonably be expected to affect speciation and/or extinction rates, so this is an important problem. A model that incorporates state-specific rates of speciation and extinction has recently been developed (BiSSE, Maddison et al. 2007), and we investigate below the extent to which it will improve phylogenetic tests of Dollo's law.

ADDITIONAL CONCERNS

The assumption of unidirectional evolution naturally raises the question of how state A came to exist at all (e.g., Dawkins 1986). Dollo's law, however, by no means asserts that a complex character cannot evolve. It simply conditionally posits that once acquired

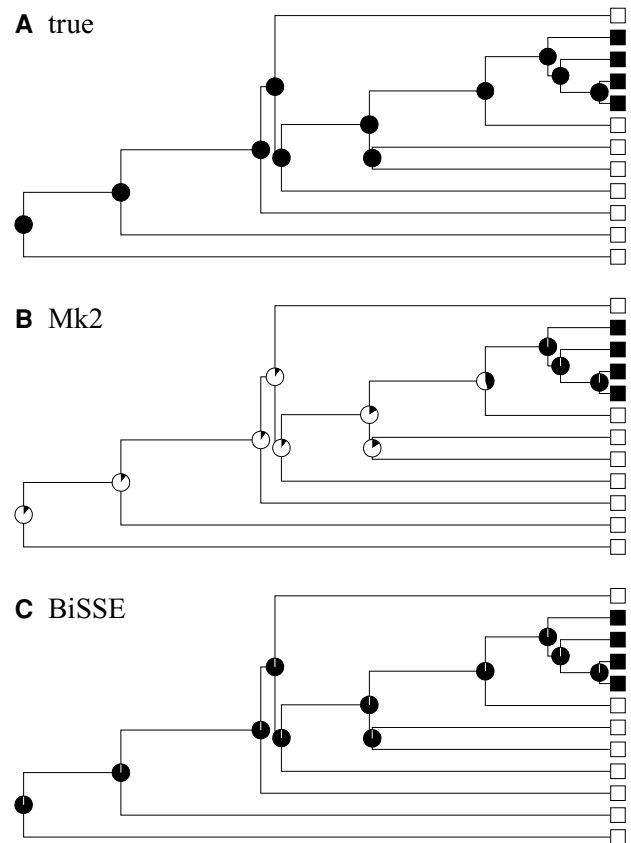


Figure 2. Ancestral reconstructions under state-dependent diversification. Transitions are unidirectional (black to white), and speciation can occur only in the black state. (A) A simulated tree showing the true states for the internal nodes. (B) The same tree, but showing Mk2 reconstructions for the nodes. Many nodes are confidently assigned to the incorrect state, and a white-to-black transition is inferred. (C) The same tree, showing BiSSE reconstructions for the nodes. All nodes are reconstructed correctly. BiSSE generally performs much better than Mk2 under state-dependent diversification, but it is not always this accurate. Using the notation black = A and white = B, the rates used in the simulation are $\lambda_A = 1$, $\lambda_B = 0$, $\mu_A = \mu_B = 0$, $q_{AB} = 1$, $q_{BA} = 0$, with elapsed time $T = 3$. The Mk2 ML rate estimates are $q_{AB} = 0.83$, $q_{BA} = 0.13$. The BiSSE ML rate estimates are $\lambda_A = 1.76$, $\lambda_B = 2.1e - 5$, $\mu_A = 0.016$, $\mu_B = 1.4e - 4$, $q_{AB} = 1.14$, $q_{BA} = 0.037$. Stationary root frequencies are used for both models.

and lost, a particular character will not be reacquired in the same (homologous) form. This is both because the probability of regain decays rapidly and, perhaps more importantly, because the original evolutionary context cannot be recreated (Gould 1970). Our procedures and recommendations assume that a nonhomologous "reversal" can be recognized as such. In our thought experiment, the assumption is that a complex state A evolved at some time prior to the date of the most recent common ancestor of the clade in question, and since that time, transitions to state B have been irreversible.

We consider only the extreme case of one-way evolution in the simulations below because violation of Dollo's law is a particularly spectacular claim. Of course, this is only a special case of asymmetry in character evolution. The biases and errors we find below apply to some extent to characters in which evolution is almost, but not entirely, irreversible. Limited simulations indicate that there is often a threshold value for the reversal rate, below which erroneous conclusions are likely and above which they are much less frequent (between 10^{-4} and 10^{-3} for the parameter values we consider below; E. E. Goldberg and B. Igić, unpubl. data).

Methods

PHYLOGENETIC TESTS OF IRREVERSIBILITY

Likelihood-ratio tests

Statistical tests of irreversibility are often carried out by using likelihood-ratio tests (LRTs) to compare nested models of character evolution (reviewed in Oakley 2003). In the case of state-independent diversification, the full model estimates two rates (the A-to-B transition rate, q_{AB} , and the B-to-A transition rate, q_{BA}). In the constrained model the reverse rate q_{BA} is fixed to zero, so only one parameter is estimated. We also briefly considered another constrained model with the rates fixed to equal each other ($q_{AB} = q_{BA}$). The full model is commonly called Mk2, and we will also refer to the first constrained model as part of the Mk2 family. The equal-rates model is commonly called Mk1 (Maddison and Maddison 2007). For state-dependent diversification, the full BiSSE model (Maddison et al. 2007) estimates six parameters: the two transition rates, and speciation and extinction rates for each state. The constrained model fixes the reverse transition rate to zero and estimates the remaining five parameters.

In either case, the maximum-likelihood values (L) under the constrained and the full models can be compared to determine whether the full model gives a sufficiently better fit to be preferred (Edwards 1972). The full model has one more parameter than the constrained model and the transition rate must be nonnegative, so $-2(\ln L_{\text{constrained}} - \ln L_{\text{full}})$ is expected to follow approximately a $(\chi_0^2 + \chi_1^2)/2$ distribution (Ota et al. 2000; adjustment for boundary condition). A small P -value from this distribution indicates rejection of the constrained model, that is, rejection of Dollo's law. A major limitation of the LRT is that it can only compare nested models. Consequently, the models it compares must make the same assumption about the root state probabilities. Because the constrained model must logically have the root fixed to state A but the unconstrained model need not, the use of an LRT is fundamentally inappropriate in tests of Dollo's law.

Ancestral state reconstructions

Commonly, empirical tests of Dollo's law have emphasized ancestral state reconstructions (ASRs; e.g., Collin and Cipriani 2003;

Whiting et al. 2003; Domes et al. 2007). This approach adds an additional layer of inference. The evolutionary rates must still be estimated under an appropriate model, and a method for reconstruction must also be chosen (Pagel 1999). Furthermore, disregard for the uncertainty associated with ML rate estimates leads to overconfidence in the reconstructions (Schultz and Churchill 1999).

Reconstructions are intuitively appealing, however, because particular branches that yield a reversal can be inferred (Schluter et al. 1997). Once ASRs have been performed, nodes at which one state is sufficiently certain can be identified (Mooers and Schluter 1999). Character state changes between such nodes, or between such a node and a tip, can then be noted. Inferring a reversal requires at least one node to be confidently reconstructed in state B, and at least one of its descendants to be in state A. Stochastic character mapping (Huelsenbeck et al. 2003) is an alternative approach for the reconstruction of ancestry, but it uses the same underlying model of character evolution.

General model selection

Of the above two commonly used methods, the LRT seems more statistically defensible than the ASR-based test (see Oakley 2003; Yang 2006) but, as currently implemented, it falls short of a proper phylogenetic test of Dollo's law. From our thought experiment above, we can see that tests for irreversibility should compare the fit of these two models: (1) two transition rates, root state not fixed, versus (2) one transition rate fixed to zero, root state fixed. Such nonnested models can be compared using, for example, the Akaike information criterion (as we do here), the Bayesian information criterion, or Bayes factors (Good 1950; Akaike 1974; Schwarz 1978).

Differences between the methods

These methods for testing Dollo's law differ philosophically. With an LRT or model comparisons, Dollo's law is rejected if the transition rate in question is significantly greater than zero. Thus, although a reversal is possible, it is not certain that one did, in fact, occur. With the ASR method, Dollo's law is rejected only if a reversal is inferred, although it is important to remember that an inferred character change did not necessarily occur.

SIMULATION STUDIES

Simulations and analyses were performed with *C* and *Python* code written specifically for the present study (available upon request from the authors), except for rate estimation under the BiSSE model that used the *Diverse* package of Mesquite (Maddison and Maddison 2007).

State-independent diversification

Simulating a phylogenetic tree under the state-independent diversification process requires the specification of four rates (two

transition rates, q_{AB} and q_{BA} , speciation rate, λ , and extinction rate, μ), the elapsed time, T , and the character state at the root. Because we focus on Dollo's law, we set $q_{BA} = 0$ in all cases. This necessitated fixing the root state to A, as discussed above. Fixing $T = 5$ but varying λ allowed the size of the tree to vary reasonably while controlling for temporal distance from equilibrium state frequencies. For simplicity, we set $\mu = 0.2$ in all runs. We considered values of 0.1, 0.5, and 1 for q_{AB} , which are small, medium, and large relative to the elapsed time (i.e., $q_{AB}T$ less than, greater than, and much greater than 1).

Few estimates of transition rates in relation to speciation or extinction rates are available from the literature, which makes the choice of biologically meaningful transition rates difficult. But a transition rate of similar magnitude to the speciation rate is reasonable based on a case study involving the loss of self-incompatibility in the plant family Solanaceae (Igić et al. 2006, E. E. Goldberg, and B. Igić, unpubl. data), as well as the values obtained from our reanalysis of wing loss in walking sticks and loss of sexual reproduction in oribatid mites (Whiting et al. 2003; Domes et al. 2007; see Case Studies below).

For each set of parameter values, we first simulated 10,000 trees and recorded the average number of tips per tree and the proportion of trees that could not be analyzed because all the tips were in the same state (therefore indicating the relative amounts of acquisition bias). The discarded trees with all B tips have on average a high value of q_{AB} , so the remaining trees will tend to underestimate this parameter, which may affect their accuracy in assessing irreversibility. Then, again for each set of parameter values, we generated 1000 trees in which both tip states were represented. On these trees, we performed LRTs and AIC-based tests of $q_{BA} = 0$ and ASRs to identify inferred reversals (described above). In practice, fixing a transition rate to zero leads to numerical difficulties, so we instead used a near-zero value (10^{-10} or 10^{-50}).

We considered Dollo's law rejected for a tree under an LRT when $P < 0.05$. An ancestral state was considered confidently reconstructed when the probability of one state exceeded 88% (Schluter et al. 1997; Mooers and Schluter 1999; Pagel 1999) under the "global, marginal" reconstruction method (Pagel 1999).

The model with the lower AIC score is the better one; there is still "substantial" support for the other model when its AIC score is up to two or greater, there is "considerably less" support when the AIC difference is four to seven, and there is "essentially no" support when the AIC difference is more than 10 (Δ AIC, Burnham and Anderson 2002). We report histograms of the AIC differences.

For LRTs and ASRs on each simulated tree, we considered three approaches for root state assignment. First, we assigned the root state probabilities to be the stationary frequencies, as determined by the estimated transition rates. They therefore vary as different rate values are considered in the likelihood maximiza-

tion, but when q_{BA} is small (i.e., near the true value), the root state will be heavily but incorrectly weighted toward state B. This is the approach taken by the original Markov models of nucleotide evolution (Felsenstein 1981), it is currently the default setting for binary character evolution in Mesquite (Maddison and Maddison 2007), and its justification is that it applies the model of evolution consistently across the tree. Second, we assumed the root state equally likely to be A or B. This has been suggested as a reasonable option when no other information is available (Pagel 1994, 1999), but it has no specific justification other than being a flat or "uninformative" prior. And third, we fixed the root to state A. Although a root of A is logically required when transitions are one-way, this assumption is not generally appropriate in an LRT because the two-way transition model should allow either root state (but see Nosil and Mooers 2005). We use it here to illustrate the rare case in which a tremendous amount of prior knowledge directly supports a root state of A, to demonstrate the source of inaccuracies, and for comparison with the ASR results.

For model selection, we computed $AIC = 2k - 2 \ln L$, where k is the number of parameters estimated and L is the likelihood, on each tree for each of two models: (1) two transition rates, stationary root frequencies, and (2) $q_{BA} = 0$, root fixed to state A. For the Mk2 family of models, $k = 2$ and 1, respectively.

Finally, we also attempted to correct for acquisition bias directly, using the method described by Felsenstein (1992) in the context of building phylogenies from restriction site data and later applied by Lewis (2001) for discrete characters. Briefly, this conditions the tree likelihood on the presence of both character states by dividing the usual likelihood by the probability that the character is variable, which is computed with the aid of a dummy character that has the same value for all tips.

State-dependent diversification

Methods for simulating phylogenetic trees under state-dependent diversification were identical to those described above, except that four additional rates must be specified (a total of six: two transition rates, q_{AB} and q_{BA} , a speciation rate for each state, λ_A and λ_B , and an extinction rate for each state, μ_A and μ_B). We manipulated values of λ_A and λ_B to effect differing strengths of state-dependent diversification while yielding trees of reasonable size. We again set $\mu_A = \mu_B = 0.2$ in all cases, except for one in which $\mu_A = \mu_B = 0.4$ to avoid unmanageably large trees; additional simulations revealed no obvious artifacts of this difference (results not shown). BiSSE has not yet been used in LRT or ASR tests of Dollo's law, so we used only Mk2 for these tests, again considering the three root state assumptions.

We then used the better-founded AIC-based model selection to show the effects of state-dependent diversification, even on a good test of irreversibility. The description of the BiSSE method (Maddison et al. 2007) assumes the stationary distribution at the

root, but we used a modified version of the Diverse package of Mesquite (Maddison and Maddison 2007; R.G. FitzJohn, pers. comm.) that allowed fixing the root state. For each tree, we compared the two models under Mk2 ($k = 2$ and 1) and then under BiSSE ($k = 6$ and 5).

CASE STUDIES

To illustrate further how model violations may affect conclusions about complex character evolution, we reanalyzed two empirical datasets using the same models as our simulation studies. In both cases, phylogenetic trees and character state data were kindly provided by the authors of the original studies. Whiting et al. (2003) described the possible reevolution of wings in stick insects using an analysis of 39 taxa (14 winged, 25 wingless). Domes et al. (2007) described a possible regain of sexual reproduction in a dataset of 23 oribatid mites species (8 sexual, 15 asexual). Sample sizes for each system are small, and we used only a single best tree in each case, to which we applied NPRS (Sanderson 1997). Consequently, our reanalyses are not intended to be definitive. They simply serve to illustrate AIC-based model selection in this context and to emphasize the dangers of relying on ASRs when more than one model is well supported.

Results

SIMULATION STUDIES

State-independent diversification

With the stationary frequency root assignment, every tree for all parameter sets falsely rejected Dollo's law in the LRT (Table 1

column 5). This is rather alarming because a frequently used and accepted method both invariably and confidently yields the wrong answer. The ASR test performed well for the low transition rate but poorly for the intermediate and high transition rates (Table 1 column 8), so the reliability of this common test is also a cause for concern.

Assuming equal probabilities for the root state yields much better performance than the stationary frequency root assignment in both the LRT and ASR tests (Table 1 columns 6 and 9) because it is only half-incorrect, but it still fails frequently for the high transition rate.

The results acquired by fixing the root state to the correct, "complex" state A are the most accurate (Table 1 columns 7 and 10) because the acquisition bias resulting from the situation in Figure 1a is eliminated, but again errors are substantial for the high transition rate because of the large number of discarded trees (with all B tips, as in Fig. 1d). For the lowest transition rate, error rates in the LRT are at approximately the expected 5% level. Although fixing the root state yields better results in our simulation study, where we know the true ancestral states and rates, it is not a defensible method in an empirical study unless there is very strong evidence about the root state, beyond the phylogeny and extant character states.

In general, for a given value of q_{AB} and controlling for elapsed time, fewer trees were discarded for parameter sets yielding greater average tree size because trees with more tips simply have more chances to retain species in state A (Table 1 columns 3 and 4). High transition rates may make rate estimation more

Table 1. State-independent diversification results. For each run, pairing parameters q_{AB} (A-to-B transition rate) and λ (speciation rate), we report the average number of tips per tree and the proportion of trees discarded (i.e., those that could not be analyzed because all their tips had the same state) for 10,000 trees. Then, for 1000 trees that had both states represented at the tips, we report the proportion of trees that rejected the hypothesis of $q_{BA}=0$ through an LRT. This test was performed for three possible priors applied to the root: stationary frequencies obtained from the transition rates, equal frequencies, and fixing the root to state A. For the same 1000 trees and three root priors, we also report the proportion of trees on which a reversal (B-to-A transition) was inferred by ancestral reconstruction. In all cases, the other parameter values used in the simulations were $q_{BA}=0$ (B-to-A transition rate), $\mu=0.2$ (extinction rate), and $T=5$ (elapsed time).

q_{AB}	λ	Mean tips per tree	Fraction discarded	Fraction of trees with $q_{BA}=0$ rejected			Fraction of trees with reversal inferred		
				Stationary root freqs.	Equal root freqs.	Root is state A	Stationary root freqs.	Equal root freqs.	Root is state A
0.1	0.8	43	0.14	1.00	0.15	0.07	0.05	0.08	0.02
0.1	1.0	114	0.08	1.00	0.09	0.05	0.06	0.07	0.02
0.1	1.2	303	0.05	1.00	0.05	0.03	0.05	0.03	0.02
0.5	0.8	44	0.50	1.00	0.18	0.10	0.61	0.14	0.05
0.5	1.0	113	0.39	1.00	0.11	0.07	0.62	0.13	0.05
0.5	1.2	306	0.29	1.00	0.05	0.07	0.46	0.07	0.03
1.0	0.8	43	0.89	1.00	0.38	0.22	0.88	0.35	0.25
1.0	1.0	114	0.82	1.00	0.30	0.15	0.96	0.35	0.21
1.0	1.2	306	0.73	1.00	0.24	0.14	0.96	0.29	0.18

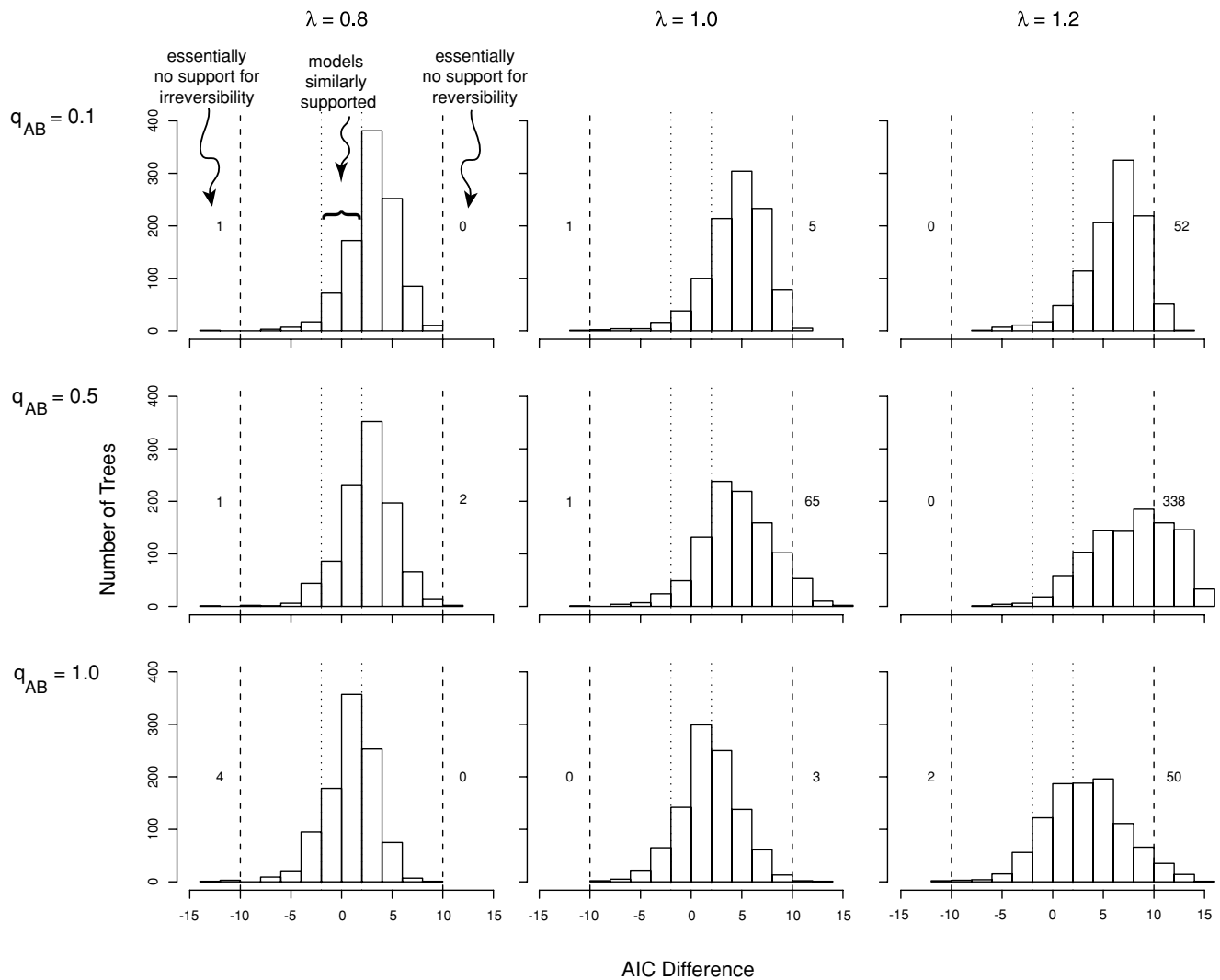


Figure 3. AIC-based tests of irreversibility under state-independent diversification. Each panel shows results for 1000 trees and corresponds to a line in Table 1 with parameter values indicated in the margins. On the horizontal axis of each panel is the difference in AIC scores between (1) the two-rate model with stationary root frequencies, and (2) the model with $q_{BA} = 0$ and the root fixed to state A, both computed under Mk2. The absolute value of this quantity is the ΔAIC score for the poorer model. Lower AIC scores indicate better model performance, so the horizontal axis is negative when the two-rate model does better and positive when the one-rate model does better. Vertical dotted lines mark $\Delta AIC = 2$, below which there is no strong preference for either model. Vertical dashed lines mark $\Delta AIC = 10$, above which there is essentially no support for the poorer model. The number of trees with $\Delta AIC > 10$ is shown near the dashed lines (some trees may fall beyond the visible segment of the horizontal axis).

difficult and ancestral state reconstructions more ambiguous, but here they led to more confidently incorrect conclusions because the discard rate (and hence acquisition bias) was greater for larger values of q_{AB} .

We do not show the results for direct acquisition bias correction because we consider our analyses preliminary, but we provisionally find that the improvement in the LRT with the stationary root assumption is substantial (roughly 30% incorrect, as compared with Mk2's 100% failure rate). Fixing the root state to A reduces the error rate to the expected 5%.

In addition, we performed LRTs of Mk2 versus Mk1 for these same trees (results not shown). The fraction of trees rejecting Mk1 varied from approximately 10% for small trees with a low value of q_{AB} (0.1) and stationary root frequencies, to nearly 100% for large trees with a high transition rate ($q_{AB} = 1.0$) and the root fixed to A. Failure to reject Mk1 indicates that use of more parameter-rich models is not justified (Mooers and Schluter 1999), but it is not sufficient basis for concluding that Dollo's law is violated.

Under the AIC-based comparison (Fig. 3), the one-way transition model was preferred over the two-way transition model

Table 2. State-dependent diversification results. Speciation rates λ_A and λ_B are specified separately for the two states. Analysis and presentation of results here are identical to Table 1. In all cases, the other parameter values used in the simulations were $q_{BA}=0$, $\mu_A=\mu_B=0.2$, and $T=5$, except for the third row, where $\mu_A=\mu_B=0.4$ to avoid unreasonably large trees.

q_{AB}	λ_A	λ_B	Mean tips per tree	Fraction discarded	Fraction of trees with $q_{BA}=0$ rejected			Fraction of trees with reversal inferred		
					Stationary root freqs.	Equal root freqs.	Root is state A	Stationary root freqs.	Equal root freqs.	Root is state A
0.1	1.0	1.0	114	0.08	1.00	0.08	0.05	0.05	0.05	0.01
0.1	1.4	0.6	574	0.03	1.00	0.14	0.11	0.06	0.07	0.04
0.1	1.6	0.0	583	0.01	1.00	0.81	0.72	0.05	0.06	0.05
0.5	1.0	1.0	113	0.39	1.00	0.11	0.06	0.58	0.11	0.04
0.5	1.4	0.6	162	0.20	1.00	0.25	0.21	0.34	0.15	0.10
0.5	1.6	0.0	317	0.06	1.00	0.90	0.86	0.15	0.14	0.13
1.0	1.0	1.0	114	0.82	1.00	0.35	0.18	0.96	0.37	0.24
1.0	1.4	0.6	55	0.61	1.00	0.49	0.37	0.88	0.39	0.28
1.0	1.6	0.0	56	0.31	1.00	0.87	0.78	0.30	0.21	0.17

for the majority of trees. The preference was stronger for lower q_{AB} and larger trees; also in these cases, fewer trees were unable to distinguish the models. Only 0.1% of all trees found “essentially no support” for the model of irreversibility. False rejections of Dollo’s law are therefore unlikely with this method, at least within these ranges of parameter values. With large trees and an intermediate transition rate, the two-way model was ruled out a substantial proportion of the time, but in other cases the AIC-based model selection was not particularly powerful.

State-dependent diversification

Irreversibility is incorrectly rejected by an LRT much more frequently when diversification is state dependent (Table 2, columns 6–8) because more rapid speciation in state A causes the B-to-A transition rate to be overestimated. Fixing the root state to A reduces the error rate but does not fix the problem.

The ASR test of irreversibility is less frequently incorrect when diversification is strongly state dependent (Table 2, columns 9–11) because overestimation of the reverse transition rate leads to less-confident reconstructions. The decrease in inaccuracy is especially prominent with the stationary root assumption because greater speciation in state A pushes the equilibrium frequency more toward A and so the root state assumption becomes less incorrect.

In the absence of state-dependent diversification, the AIC model comparison results are quite similar under Mk2 and BiSSE (Fig. 4, first column). Incorrect dismissal of irreversibility is almost nonexistent, but correct dismissal of reversibility is not common (though perhaps better under Mk2).

When state A has a higher speciation rate, Mk2 becomes more likely to reject irreversibility, but BiSSE does not (Fig. 4, second and third columns). Mk2 may also be more likely to come

to the wrong conclusion for a higher transition rate, although this is difficult to disentangle from the effect of tree size.

The greatest diversification rate difference we consider is admittedly extreme ($\lambda_A = 1.6$, $\lambda_B = 0$), but our results serve to demonstrate both the potential danger of ignoring the possibility of state-dependent diversification in tests of Dollo’s law and the efficacy of BiSSE in accounting for it.

Overall

Our simulation results show that likelihood ratio and ancestral reconstruction tests frequently reject Dollo’s law when character evolution truly is irreversible (Tables 1 and 2). By varying our assumptions about the root state, we show that much of this error is due to conflict between the equilibrium frequencies of a unidirectionally evolving character (under only A-to-B transitions, all lineages are in state B at equilibrium) and the logical requirement that the root be in state A in order for both states to be extant in the clade.

When diversification rate does not depend on character state, using AIC to select among logically appropriate models leads to almost no incorrect rejection of Dollo’s law (Fig. 3). Under state-dependent diversification, this method frequently fails when used with Mk2, but false rejection errors are again nearly eliminated with the use of BiSSE (Fig. 4). With either Mk2 or BiSSE, the AIC-based method does not, however, always confidently select the correct model, so the power and general utility of this method and other techniques for model selection in this context remain to be seen.

CASE STUDIES

Application of the model selection approach to testing Dollo’s law in stick insects (data from Whiting et al. 2003) and oribatid mites (data from Domes et al. 2007) is summarized in Table 3.

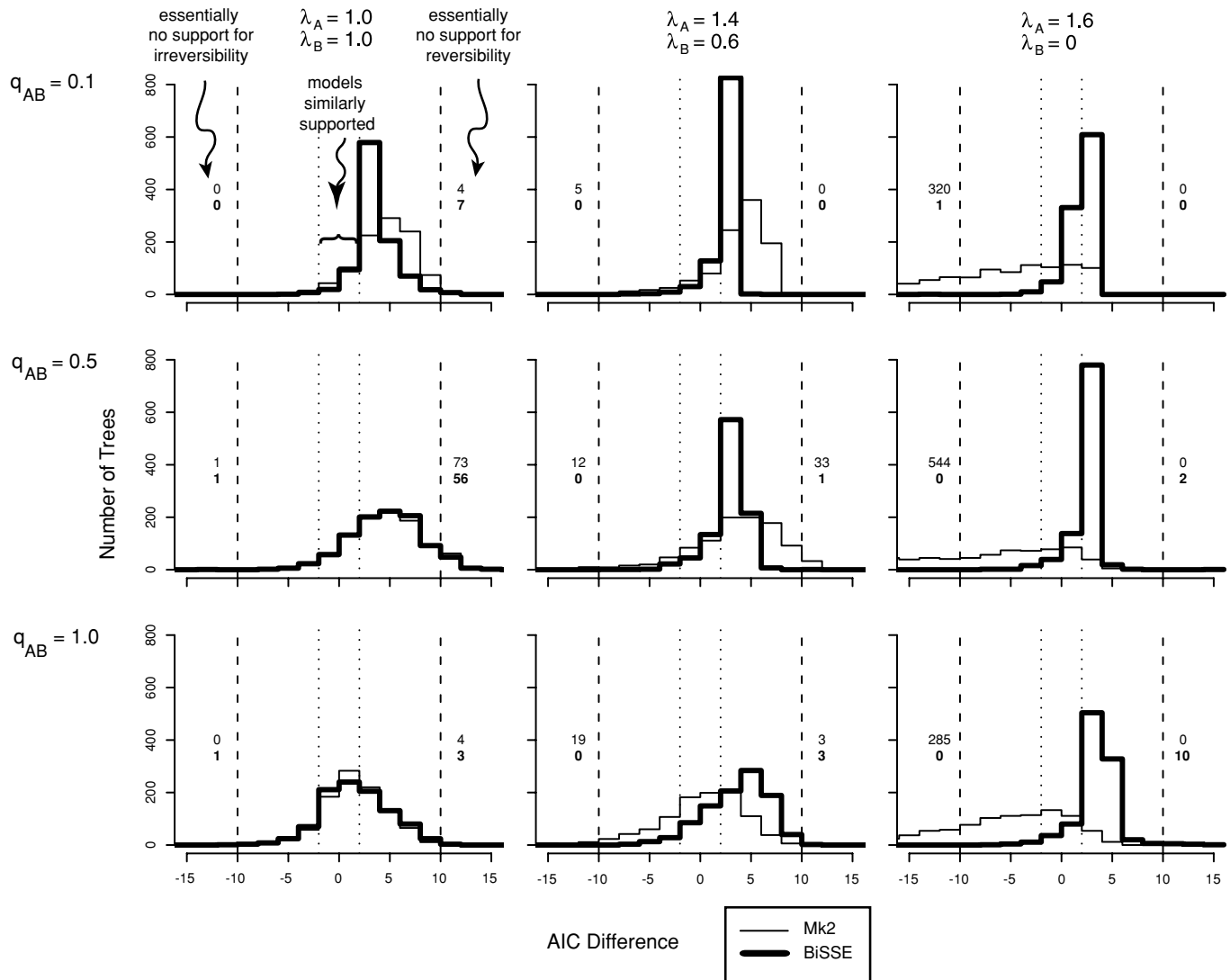


Figure 4. AIC-based tests of irreversibility under state-dependent diversification. Axes are the same as in Figure 3, and each panel again shows the results for 1000 trees and corresponds to a line in Table 2, with parameter values indicated in the margins. The thin-lined histograms show results of model comparisons under Mk2, and the thick-lined histograms show results under BiSSE. Numbers near $\Delta AIC = 10$ (regular weight for Mk2, bold for BiSSE) give the number of trees incorrectly dismissing unidirectional transitions (left) or correctly dismissing bidirectional transitions (right).

First, using the BiSSE model, the four combinations of state-independent and state-dependent diversification and unidirectional and bidirectional transitions were considered. Under state-independent diversification, the two speciation rates were constrained to be equal, as were the two extinction rates. Under unidirectionality, the root was fixed to the “complex” state (denoted A) and q_{BA} was fixed to 0. Note that the BiSSE model must be used even for the state-independent diversification models in order for the likelihood values to be comparable. For both stick insect and mite datasets, there was considerably less support for the models including state-dependent diversification. Unidirectional evolution was preferred for the stick insects,

but bidirectional evolution was equally well supported for the mites.

Because the complication of state-dependent diversification was not clearly warranted, we then assessed irreversibility using Mk2, which may have more power because it estimates fewer parameters. For the stick insects, the model allowing reversals from wingless to winged states received considerably less support than the irreversible model. For the mites, the models were about equally well supported, indicating that there is presently no definitive evidence for regain of sexual reproduction. To illustrate the dangers of assessing modes of evolution from ancestral reconstructions before finding the statistically best-supported

Table 3. Model comparison tests of Dollo's law for stick insects (Whiting et al. 2003) and mites (Domes et al. 2007). For the stick insects, state A is winged and state B is wingless. For the mites, state A is sexual reproduction and state B is asexual reproduction. For each system, two sets of AIC-based comparisons were performed. First, under the BiSSE model, the four combinations of state-(in)dependent diversification and (ir)reversible transitions were considered. Rate estimates are reported in units of inverse rate-smoothed nucleotide-based distance normalized to a crown-group age of 1, followed by the log-likelihoods for each model and the corresponding Δ AICs. In all cases, extinction rates were 10^{-4} or less and are not shown. Based on the BiSSE Δ AIC results, state-dependent diversification garners considerably less support. Consequently, the Mk2 model was then used to assess irreversibility. For the walking sticks, there is considerably less support for the model allowing reversals. For the mites, both models receive approximately equal support. Ancestral state reconstructions for the two Mk2 models are shown in Figure 5.

System	Model	Root	λ_A	λ_B	q_{AB}	q_{BA}	$\ln L$	Δ AIC
Stick insects	BiSSE	Stationary		0.37	2.74	0.50	-24.32	3.74
		Fixed		0.51	2.23	0	-23.45	0
		Stationary	3.82	0.37	2.74	0.50	-24.32	7.74
		Fixed	3.50	0.51	2.23	0	-23.45	4.00
	Mk2	Stationary	—	—	1.62	0.59	-22.02	4.14
		Fixed	—	—	1.81	0	-20.95	0
Oribatid mites	BiSSE	Stationary		2.28	0.95	0.73	-18.44	0.84
		Fixed		1.91	1.14	0	-19.02	0
		Stationary	1.62	2.28	0.95	0.74	-18.43	4.82
		Fixed	2.15	1.92	1.14	0	-19.02	4.00
	Mk2	Stationary	—	—	1.05	0.67	-12.67	1.06
		Fixed	—	—	1.13	0	-13.14	0

model, we show reconstructions for each system under the two Mk2 models (Fig. 5).

Discussion

REVERSAL ON IRREVERSIBILITY

Many recent studies have found losses of complex characters to be reversible, in violation of Dollo's law (Oakley and Cunningham 2002; Collin and Cipriani 2003; Whiting et al. 2003; Nosil and Mooers 2005; Cruickshank and Paterson 2006; Domes et al. 2007; Ferrer and Good-Avila 2007; Brandley et al. 2008; reviewed in Pagel 2004; Kohlsdorf and Wagner 2006; Domes et al. 2007). Each study relied on phylogenetic tests that yield confidently incorrect results when asymmetry in transition rates is great (Tables 1 and 2), as is expected for complex characters. No study of Dollo's law performed to date dealt with both the implicit root state assumption and the effects of character state-dependent differences in diversification rates. Both of these problems were recognized in other work, but their severity in the case of irreversibility has not previously been acknowledged. Likewise, existing methods can address these problems, but the solutions have not been widely adopted or heavily tested.

The acquisition bias resulting from analysis of only those clades that show a mix of states was previously recognized in both empirical (Oakley and Cunningham 2000; Schluter 2000, p. 43) and theoretical contexts (Felsenstein 1992; Frumhoff and Reeve 1994; Schultz et al. 1996; Lewis 2001). A solution in the Mk2 framework exists (Lewis 2001), but it has not been widely used.

When evolution of a character is unidirectional, much of this bias is removed when the logical requirement for the root to be in the "complex" state is incorporated (Fig. 1; see also Nosil and Mooers 2005). To test Dollo's law, however, this model should be compared to a bidirectional model with an unfixed root (Figs. 3 and 4).

Character states that are likely to confer unequal probabilities of speciation and extinction (e.g., sexual vs. asexual reproduction, outcrossing vs. selfing in hermaphrodites, ecological generalists vs. specialists, winged flight vs. winglessness, feeding vs. non-feeding larvae) are of particular interest to evolutionary biologists, and phylogenetic distributions of character states are the outcome of these processes as well as of character state transitions (Maddison et al. 2007). Indeed, the potential of species-level selection to trump individual-level selection in determining character state frequencies was strongly championed by Stanley (1975) and Gould and Eldredge (1977). Although problems with the assumption of state-independent diversification rates were previously recognized (e.g., Stireman 2005; Igic et al. 2006; Maddison 2006), statistical methods for understanding such characters became available only very recently (BiSSE, Maddison et al. 2007). Here, we find that BiSSE performs substantially better than Mk2 in tests of irreversibility under state-dependent diversification (Fig. 4).

TOWARD BETTER TESTS OF IRREVERSIBILITY AND ASYMMETRY IN CHARACTER EVOLUTION

Our results reveal serious flaws in the presently employed methods used for tests of Dollo's law. Below, we discuss how the

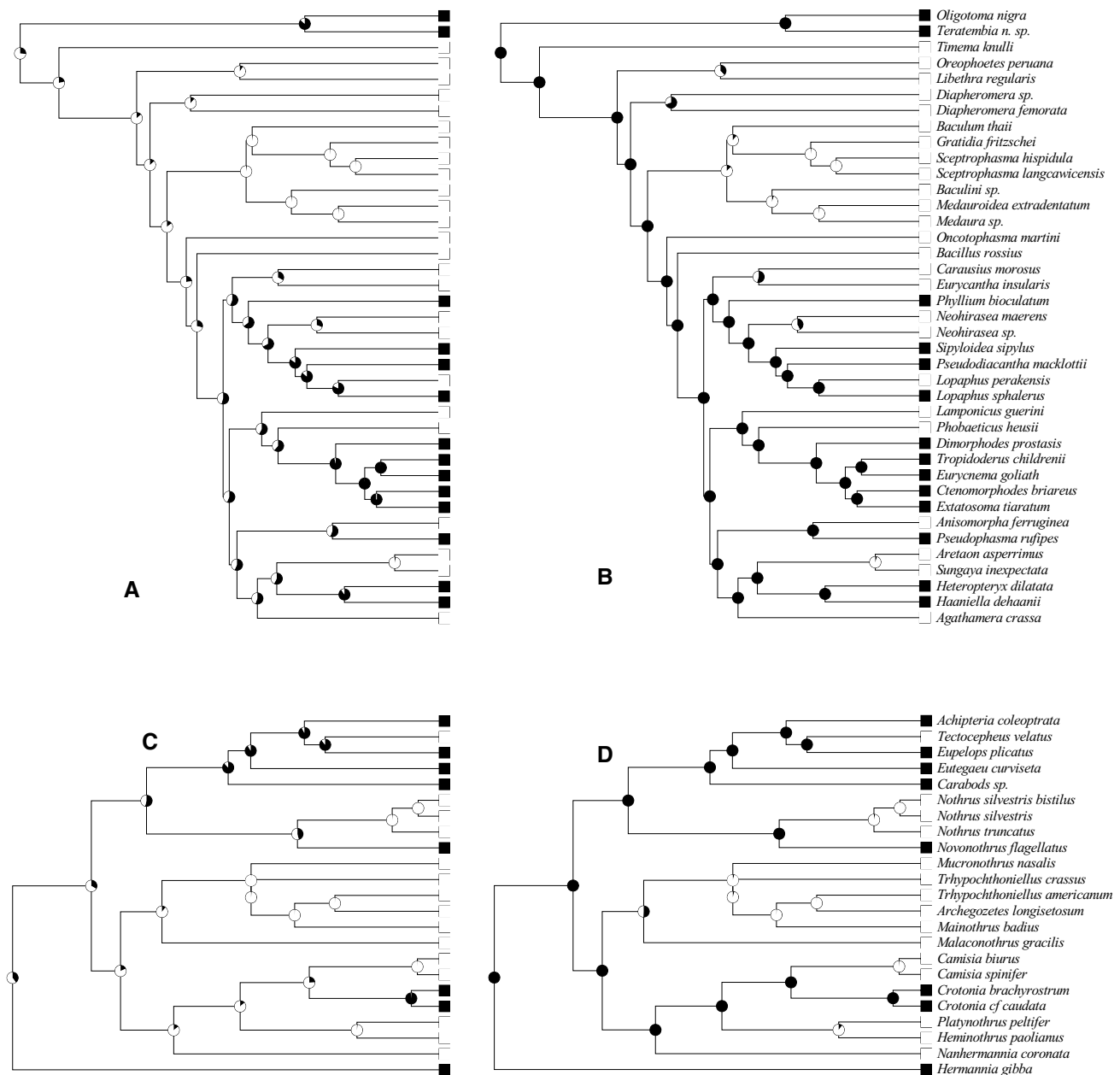


Figure 5. Ancestral state reconstructions of winged and wingless states for stick insects (A, B; data from Whiting et al. 2003) and of sexual and asexual reproduction for oribatid mites (C, D; data from Domes et al. 2007). The complex state (winged or sexual) is black, and the simpler state (wingless or asexual) is white. Two Mk2 models were used in each system: (A, C) two transition rates, stationary root, and (B, D) one transition rate fixed to zero, root fixed to complex state. Based on the results in Table 3; for the stick insects, there is substantially less support for (A) than for (B), whereas for the mites, there is approximately equal support for (C) and (D).

definition of Dollo's law may affect the framing and scope of hypothesis tests, propose appropriate phylogenetic tests of irreversibility, and stress the need for additional empirical data that can greatly improve accuracy and power.

Defining irreversibility

The fate of Dollo's law is ultimately context-dependent and scale-dependent (Wagner 1982). For example, Bull and Charnov (1985)

adopt a phenotypic definition, so that any incarnation of a complex state, regardless of its genetic basis, is deemed a reversal. Because one can now examine the genetic basis of phenotypic traits, a more appropriate level of analysis should consider the genotype. But is the involvement of the same genetic pathways or genes sufficient to infer a reversal, or must one pinpoint the exact nucleotide reversals? The latter form may be of limited use and not particularly testable, and amounts to "death by a thousand

qualifications" (cf. Gould 1970, p. 197), even without the additional burden of specifying the functional and epigenetic interdependencies of characters (Wagner 1982).

The bulk of the difficulty with defining a "Dollo character" stems from the fact that Dollo's arguments specifically addressed the cases in which a phylogeny can be established, despite the convergent reversal of one part of an organism, as he was primarily concerned with identifying such a convergence by its "indestructible past" (Dollo 1893; Gould 1970). There is also the question of whether significant reversals in quantitative or meristic traits should be considered violations of Dollo's law, although at least Hennig (1966) and Gould (1970) argue against their consideration. We generally side with Bull and Charnov's (1985) assessment: the appropriate levels of analysis and the suitable characters remain unclear, and it is unlikely that a single all-encompassing solution will be broadly accepted. But empirical studies can achieve transparency if reversals are explicitly defined and justified within the context of each study. Our principal concern is that no test of irreversibility performed to date used an appropriate testing framework, regardless of the exact definition employed.

Phylogenetic methods

Relying solely on ancestral state reconstructions under a single model to assess irreversibility frequently leads to incorrect rejections of Dollo's law, and contradictory conclusions may be reached under equally well-supported models. Likelihood-ratio tests of irreversibility come to incorrect conclusions even more commonly and are logically flawed unless independent evidence about the root state is available. Improved methods for phylogenetic tests of Dollo's law are clearly essential.

A first step is the use of more general model selection methods (e.g., Akaike information criterion, Bayesian information criterion, Bayes factors; Good 1950; Akaike 1974; Schwarz 1978), which allow the appropriate comparison between the two transition rate model with unfixed root and the single nonzero rate model with fixed root. When the character state has no effect on diversification rate, AIC-based tests of Dollo's law with Mk2 models perform well. Under state-dependent diversification, however, BiSSE must be used to avoid incorrect rejection of irreversibility. We illustrate a working procedure in our case studies. The relative performance of various model selection procedures, their power to dismiss the two-rate model, and their error rates for traits that are indeed reversible remain unknown.

Fixing the root state when transitions are unidirectional removes much but not all of the acquisition bias. Acquisition bias can be more directly attacked by conditioning the tree likelihood on the presence of both tip states (Felsenstein 1992). The modification is not difficult for Mk2 (Lewis 2001), but a similar correction for BiSSE will be more challenging because its likeli-

hood computation includes both the character states and the tree shape. These corrections are, however, still vulnerable to incorrect assumptions about the root state. More generally, an explicit description of the nature of the bias underlying the selection of the dataset at hand could be incorporated. The presence of more than one state at the tips is a minimum constraint, but additional features that may be commonly sought, such as sufficient frequency of each state, reasonable intermingling of the tip states, phylogenetic overdispersion, or reasonable clade size, could also be incorporated. Another common practice is including a few token outgroup taxa in an effort to drive the root toward the expected state. All of these forms of incomplete and selective sampling in both the focal group and outgroup could seriously mislead rate estimation and state reconstruction. The magnitude of biases from such complex selectivity may best be investigated by simulation.

We have concentrated here on the extreme case of asymmetry—a completely irreversible character state change. When character evolution is less asymmetric, all of these issues are still present to some degree. In fact, the difficulties of rate estimation may be even more insidious. If unidirectionality (one transition rate, fixed root) is correctly dismissed, a bidirectional model with stationary root frequencies may consequently be accepted. When one transition rate is very small, however, the equilibrium root assumption may still be almost as incorrect as when one transition rate is zero, and the magnitude of the small transition rate may therefore be greatly overestimated. One means of diagnosing this problem is to compare a group of models that each include the two transition rates but make a range of assumptions about the root state frequencies. Caution and additional data are still required when a prior has a strong effect on conclusions.

Although we have focused only on likelihood methods, extension of the proposed models and methods to a Bayesian framework could be particularly appropriate for tests of Dollo's law. As with the application of parsimony weights (Kohn et al. 1996; Wray 1996; Omland 1997; Lee and Shine 1998), the placement of priors on the asymmetry of character transitions in a Bayesian analysis (Schultz and Churchill 1999) could dramatically alter conclusions about character reversibility. Such a prior need not be highly subjective and, ideally, it would be a quantitative mechanism for incorporating the findings of previous studies of the system. Similarly, fixing the character state at the root could be carried out more justifiably in a Bayesian context, where a variety of external information may be expressed as a prior distribution. This is extensible to ancestral nodes other than the root. The role of prior expectations in the inferences of binary character evolution is lucidly reviewed by Schultz and Churchill (1999).

Our simulation study uses known trees, allowing us to concentrate on the performance of character evolution models. In practice, empirical studies should integrate the rapid advances in incorporation of phylogenetic uncertainty (e.g., Huelsenbeck et al.

2000; Pagel and Lutzoni 2002; Pagel et al. 2004). Nevertheless, emphasis on these sophisticated methods should not overshadow the fundamental assumptions and limitations of the underlying evolutionary models.

Empirical study beyond phylogeny

Phylogenetic methods will continue to improve, but the most powerful assessments of the importance of irreversibility require other sources of information. The use of fossil data can improve inference of character evolution, but it is unlikely to be a panacea. For any past time at which both character states occurred, fossil taxa of a focal group with either character state may be recovered, making it very difficult to assign ancestry at a particular node, even in the presence of a remarkable fossil record. Observations of changes in state frequencies over time could be fit to models of character change to test for irreversibility, but data of sufficient precision may be impossible to obtain. Most importantly, many complex characters of great interest to evolutionary biologists, and subject to Dollo's law, are generally inestimable from fossils. Nevertheless, even if limited and uncertain information about the character states can be gleaned from the fossil record, it could be used to inform the root state assumption. For example, Wagner (1996) and Hunt (2007) each use superb data from closely related species to show evidence for an ancestral morphological shift in marine gastropods and ostracods, respectively. Other work (e.g., Polly 2001; Webster and Purvis 2002; Finarelli and Flynn 2006) incorporates fossils in studies of character evolution to anchor the values for highly uncertain reconstructions of continuous characters and to test model reconstructions.

In contrast to many phylogenetic studies whose conclusions are considerably weakened by our results, a few expressed strong doubt about the results of their reconstructions of potentially irreversible traits (Takebayashi and Morrell 2001; Stireman 2005; Igic et al. 2006). In the context of testing whether selfing is a dead-end, Takebayashi and Morrell (2001) pointed out that the tip state frequencies in their analyses were disproportionately determining the inferred transition rates. Subsequently, they correctly expressed reservations about the apparent regains of outcrossing. Stireman (2005); concerned with the evolution of ecological specialists and generalists, provided an extensive logical argument that correctly called into question the entire enterprise of reconstructions when net diversification rates differ.

Several recent empirical studies provide a cause for optimism. Each is relevant to the study of irreversibility, finds direct genetic evidence for the identity of ancestral states, and shows the power of integrating molecular genetic and phylogenetic methods. Yun et al. (1999) and Inderbitzin et al. (2005) use data on the genetic organization of the locus that regulates sexual reproduction in a fungus (*Stemphylium*, Ascomycota) to pinpoint breeding system transitions from outcrossing to selfing with remarkable

accuracy. Adding a layer of complexity, it seems that this system also involves character state transitions caused by lateral genetic transfer (Inderbitzin et al. 2005). The unique genetic properties of breeding system loci were also used to infer the history of loss of self-incompatibility in the nightshade family (Solanaceae) and find unidirectional transitions (Igic et al. 2006). Igic et al. (2004, 2006) skeptically viewed their initial results, which favored many reversals, and provided independent genetic evidence against the original "naive" reconstructions. Specifically, information from the broadly occurring trans-specific polymorphism at the self-incompatibility locus was used as evidence to establish the ancient ancestry and irreversibility of self-incompatibility, at least within the last ca. 30 million years. In a study of flower color transitions, Zufall and Rausher (2004) demonstrated a possible approach for establishing ancestral states. They used a well-understood pigment-producing pathway to determine the history of flower color transitions in a group of morning glories (Convolvulaceae). Likewise, the use of molecular genetic data has allowed potentially better estimation of ancestry for other characters not explicitly concerned with Dollo's law (e.g., Hoekstra and Edwards 2000; Mark Welch et al. 2004).

At least two studies use geographic data to garner additional evidence. Culver et al. (1995) describe karst windows exposing previously cave-dwelling amphipod populations to light, apparently leading to the regain of vision. Wiens et al. (2007) take a different tack, using climate reconstructions to argue that character state reconstructions of the developmental program in marsupial frogs may be flawed, and thus may support a reversal. Their case is complicated by the difficulties of climate reconstruction, incomplete sampling, and possible state-dependent diversification, but it is novel and potentially promising.

The experimental approaches taken by Oakley and Cunningham (2000) and Teotonio and Rose (2000, 2001) are sure to be remain useful for tests of any future reconstruction methods, which may, for example, merge BiSSE, acquisition bias correction, a rate heterogeneity model, and stochastic character mapping. However, simulations and experimental populations will rarely incite the level of interest commanded by convincing empirical studies from nature.

ACKNOWLEDGMENTS

We thank J. P. Huelsenbeck, B. Hutchins, J. R. Kohn, S. Lidgard, K. A. McKean, L. Popovic, D. Merl, T. Price, R. H. Ree, W. F. Fagan's lab, and the FM-ED Group for discussions, and J. B. Joy, P. Saperstein, and J. A. Finarelli for comments that improved the manuscript. A. Ø. Mooers and T. H. Oakley wrote especially helpful reviews and shared thoughts in further correspondence. M. F. Whiting and K. Domes kindly provided character state data and phylogenies for stick insects and oribatid mites, respectively. S. P. Otto shared thoughts on incorporating acquisition bias into BiSSE and R. G. FitzJohn provided a key patch for Mesquite. W. F. Briskin wrote the first version of PieTree, which produced Figures 2 and 5.

LITERATURE CITED

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19:716–723.
- Brandley, M. C., J. P. Huelsenbeck, and J. J. Wiens. 2008. Rates and patterns in the evolution of snake-like body form in squamate reptiles: evidence for repeated re-evolution of lost digits and long-term persistence of intermediate body forms. *Evolution* 62:2042–2064.
- Bull, J. J. 2000. Evolutionary biology. *Deja vu. Nature* 408:416–417.
- Bull, J. J., and E. L. Charnov. 1985. On irreversible evolution. *Evolution* 39:1149–1155.
- Burnham, K. B., and D. Anderson. 2002. *Model selection and multi-model inference: a practical information-theoretic approach*, 2nd ed. Springer-Verlag, New York.
- Collin, R., and R. Cipriani. 2003. Dollo's law and the re-evolution of shell coiling. *Proc. R. Soc. Lond. B* 270:2551–2555.
- Cruickshank, R. H., and A. M. Paterson. 2006. The great escape: do parasites break Dollo's law? *Trends Parasitol.* 22:509–515.
- Culver, D. C., T. C. Kane, and D. W. Fong. 1995. *Adaptation and natural selection in caves: the evolution of Gammarus minus*. Harvard Univ. Press, Cambridge, MA.
- Cunningham, C. W. 1999. Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses. *Syst. Biol.* 48:665–674.
- Cunningham, C. W., K. E. Omland, and T. H. Oakley. 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* 13:361–366.
- Dawkins, R. 1986. *The Blind Watchmaker*. Norton, New York.
- Dollo, L. 1893. Les lois de l'évolution. *Bulletin de la Société Belge de Géologie de Paléontologie et d'Hydrologie* 7:164–166.
- Domes, K., R. A. Norton, M. Maraun, and S. Scheu. 2007. Re-evolution of sexuality breaks Dollo's law. *Proc. Natl. Acad. Sci. USA* 104:7139–7144.
- Edwards, A. F. W. 1972. *Likelihood*. Cambridge Univ. Press, Cambridge.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 6:368–376.
- . 1992. Phylogenies from restriction sites: a maximum-likelihood approach. *Evolution* 46:159–173.
- Ferrer, M. M., and S. V. Good-Avila. 2007. Macrophylogenetic analyses of the gain and loss of self-incompatibility in the Asteraceae. *New Phytol.* 173:401–414.
- Finarelli, J. A., and J. J. Flynn. 2006. Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): the effects of incorporating data from the fossil record. *Syst. Biol.* 55:301–313.
- Frumhoff, P., and H. Reeve. 1994. Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. *Evolution* 48:172–180.
- Good, I. 1950. *Probability and the weighing of evidence*. Charles Griffin, New York.
- Gould, S. J. 1970. Dollo on Dollo's law: irreversibility and the status of evolutionary laws. *J. Hist. Biol.* 3:189–212.
- Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151.
- Harvey, P. H., and M. D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford Univ. Press, Oxford, UK.
- Hennig, W. 1966. *Phylogenetic Systematics*. Univ. of Illinois Press, Champaign, IL.
- Hoekstra, H. E., and S. V. Edwards. 2000. Multiple origins of XY female mice (genus *Akodon*): phylogenetic and chromosomal evidence. *Proc. R. Soc. Lond. B* 267:1825–1831.
- Huelsenbeck, J. P., B. Rannala, and J. P. Masly. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* 288:2349–2350.
- Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic mapping of morphological characters. *Syst. Biol.* 52:131–158.
- Hunt, G. 2007. Morphology, ontogeny, and phylogenetics of the genus *Poseidonamicus* (Ostracoda: Thaerocytherinae). *Paleobiology* 81:607–631.
- Igic, B., L. Bohs, and J. R. Kohn. 2004. Historical inferences from the self-incompatibility locus. *New Phytol.* 161:97–105.
- . 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proc. Natl. Acad. Sci. USA* 103:1359–1363.
- Inderbitzin, P., J. Harkness, B. G. Turgeon, and M. L. Berbee. 2005. Lateral transfer of mating system in *Stemphylium*. *Proc. Natl. Acad. Sci. USA* 102:11390–11395.
- Janson, C. H. 1992. Measuring evolutionary constraints: a Markov model for phylogenetic transitions among seed dispersal syndromes. *Evolution* 46:136–158.
- Kohlsdorf, T., and G. P. Wagner. 2006. Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in *Bachia* (Gymnophthalmidae: Squamata). *Evolution* 60:1896–1912.
- Kohn, J. R., S. W. Graham, B. Morton, J. J. Doyle, and S. C. H. Barrett. 1996. Reconstruction of the evolution of reproductive characters in Pontederiaceae using phylogenetic evidence from chloroplast DNA restriction-site variation. *Evolution* 50:1454–1469.
- Lee, M. S. Y., and R. Shine. 1998. Reptilian viviparity and Dollo's law. *Evolution* 52:1441–1450.
- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50:913–925.
- Maddison, W. P. 2006. Confounding asymmetries in evolutionary diversification and character change. *Evolution* 60:1743–1746.
- Maddison, W. P., and D. R. Maddison. 1992. *MacClade*. Sinauer Associates, Sunderland, MA.
- . 2007. *Mesquite: a modular system for evolutionary analysis*. Version 2.0b3 URL <http://mesquiteproject.org/>.
- Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–710.
- Mark Welch, J. L., D. B. Mark Welch, and M. Meselson. 2004. Cytogenetic evidence for asexual evolution of bdelloid rotifers. *Proc. Natl. Acad. Sci. USA* 101:1618–1621.
- Marshall, C. R., E. C. Raff, and R. A. Raff. 1994. Dollo's law and the death and resurrection of genes. *Proc. Natl. Acad. Sci. USA* 91:12283–12287.
- McIntyre, L. 1997. Gould on laws in biological science. *Biol. Philos.* 12:357–367.
- Mooers, A. Ø., and D. Schluter. 1999. Reconstructing ancestor states with maximum likelihood: support for one- and two-rate models. *Syst. Biol.* 48:623–633.
- Muller, H. 1939. Reversibility in evolution considered from the standpoint of genetics. *Biol. Rev.* 14:261–280.
- Nosil, P., and A. Ø. Mooers. 2005. Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution* 59:2256–2263.
- Nunney, L. 1989. The maintenance of sex by group selection. *Evolution* 43:245–257.
- Oakley, T. H. 2003. Maximum likelihood models of trait evolution. *Comments Theor. Biol.* 8:1–17.
- Oakley, T. H., and C. W. Cunningham. 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution* 54:397–405.
- . 2002. Molecular phylogenetic evidence for the independent evolutionary origin of an arthropod compound eye. *Proc. Natl. Acad. Sci. USA* 99:1426–1430.
- Omland, K. E. 1997. Examining two standard assumptions of ancestral state reconstruction: repeated loss of dichromatism in dabbling ducks (Anatini). *Evolution* 51:1636–1646.
- Ota, R., P. J. Waddell, M. Hasegawa, H. Shimodaira, and H. Kishino. 2000. Appropriate likelihood ratio tests and marginal distributions for

- evolutionary tree models with constraints on parameters. *Mol. Biol. Evol.* 17:798–803.
- Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B* 255:37–45.
- . 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* 48:612–622.
- . 2004. Limpets break Dollo's law. *Trends in Ecology and Evolution* 19:278–280.
- Pagel, M., and F. Lutzoni. 2002. Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. Pp. 151–164 in *Biological Evolution and Statistical Physics*, Springer-Verlag, Berlin.
- Pagel, M., A. Meade, and D. Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* 53:673–684.
- Paradis, E. 2008. Asymmetries in phylogenetic diversification and character change can be untangled. *Evolution* 62:241–247.
- Polly, P. D. 2001. Paleontology and the comparative method: ancestral node reconstructions versus observed node values. *Am. Nat.* 157:596–609.
- Sanderson, M. J. 1993. Reversibility in evolution: a maximum likelihood approach to character gain/loss bias in phylogenies. *Evolution* 47:236–252.
- . 1997. A non-parametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14:1218–1231.
- Schluter, D. 2000. *The ecology of adaptive radiation*. Oxford Univ. Press, Oxford, UK.
- Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51:1699–1711.
- Schultz, T. R., and G. A. Churchill. 1999. The role of subjectivity in reconstructing ancestral character states: a Bayesian approach to unknown rates, states, and transformation asymmetries. *Syst. Biol.* 48:651–664.
- Schultz, T. R., R. B. Coccoft, and G. A. Churchill. 1996. The reconstruction of ancestral character states. *Evolution* 50:504–511.
- Schwarz, G. E. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- Simpson, G. G. 1953. *The major features of evolution*. Columbia Univ. Press, New York.
- Stanley, S. M. 1975. A theory of evolution above the species level. *Proc. Natl. Acad. Sci. USA* 72:646–650.
- Stireman, J. O. 2005. The evolution of generalization? Parasitoid flies and the perils of inferring host range evolution from phylogenies. *J. Evol. Biol.* 18:325–336.
- Strathmann, R. R., and D. J. Eernisse. 1994. What molecular phylogenies tell us about the evolution of larval forms. *Am. Zool.* 34:502–512.
- Takebayashi, N., and P. L. Morrell. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *Am. J. Bot.* 88:1143–1150.
- Teotonio, H., and M. R. Rose. 2000. Variation in the reversibility of evolution. *Nature* 408:463–466.
- . 2001. Perspective: reverse evolution. *Evolution* 55:653–660.
- Wagner, G. P. 1982. The logical structure of irreversible systems transformations: a theorem concerning Dollo's law and chaotic movement. *J. Theor. Biol.* 96:337–346.
- Wagner, P. J. 1996. Contrasting the underlying patterns of active trends in morphologic evolution. *Evolution* 50:990–1007.
- Webster, A. J., and A. Purvis. 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. *Proc. R. Soc. Lond. B* 269:143–149.
- Whiting, M. F., S. Bradler, and T. Maxwell. 2003. Loss and recovery of wings in stick insects. *Nature* 421:264–267.
- Wiens, J. J., C. A. Kuczynski, W. E. Duellman, and T. W. Reeder. 2007. Loss and re-evolution of complex life cycles in marsupial frogs: does ancestral trait reconstruction mislead? *Evolution* 61:1886–1899.
- Wray, G. A. 1996. Parallel evolution of nonfeeding larvae in echinoids. *Syst. Biol.* 45:308–322.
- Yang, Z. 2006. *Computational Molecular Evolution*. Oxford Univ. Press, Oxford, UK.
- Yun, S. H., M. L. Berbee, O. C. Yoder, and B. G. Turgeon. 1999. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proc. Natl. Acad. Sci. USA* 96:5592–5597.
- Zufall, R. A., and M. D. Rausher. 2004. Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature* 428:847–850.

Associate Editor: D. Posada