Syst. Biol. 55(5):803–817, 2006

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150600981562

The Ancestral Distance Test: What Relatedness can Reveal about Correlated Evolution in Large Lineages with Missing Character Data and Incomplete Phylogenies

DAVID HEARN[1] AND MARK HUBER[2]

[1]*Plant Sciences, University of Arizona, 303 Forbes Building, Tucson, Arizona 85721, USA; E-mail: dhearn@email.arizona.edu*
[2]*Department of Mathematics, Duke University, Durham, North Carolina 27708, USA; E-mail: mhuber@math.duke.edu*

Abstract.— The ancestral distance test is introduced to detect correlated evolution between two binary traits in large phylogenies that may lack resolved subclades, branch lengths, and/or comparative data. We define the *ancestral distance* as the time separating a randomly sampled taxon from its most recent ancestor (MRA) with extant descendants that have an independent trait. The sampled taxon either has (*target sample*) or lacks (*nontarget sample*) a dependent trait. Modeled as a Markov process, we show that the distribution of ancestral distances for the target sample is identical to that of the nontarget sample when characters are uncorrelated, whereas ancestral distances are smaller on average for the target sample when characters are correlated. Simulations suggest that the ancestral distance can be estimated using the time, total branch length, taxonomic rank, or number of speciation events between a sampled taxon and the MRA. These results are shown to be robust to deviations from Markov assumptions. A Monte Carlo technique estimates *P*-values when fully resolved phylogenies with branch lengths are available, and we evaluate the Monte Carlo approach using a data set with known correlation. Measures of relatedness were found to provide a robust means to test hypotheses of correlated character evolution. [Ancestral distance; character correlation test; homeosis; Markov model; Monte Carlo simulation; phylogeny; Poisson process; rate heterogeneity; taxonomic rank; Yule tree.]

A central goal of comparative biology is to understand the evolution of form, function, and behavior. A common approach to investigate their interplay is to test for correlated evolution between traits of interest. Ridley (1983) and Felsenstein (1985) demonstrated that tests of correlated evolution need to include information about the phylogeny in order to take the nonindepdendence of traits into account. Several comparative tests have since been created (reviewed by Harvey and Pagel 1991; Maddison, 1994; Martins, 1996a; Martins and Hanson, 1996; Ridley and Grafen, 1996; Pagel, 1997). Such tests often require fully resolved phylogenies, branch lengths, and comparative data for all taxa under consideration (Table 1). Comparative studies that investigate infrequent traits that appear sporadically across large lineages of life (i.e., thousands of taxa) are limited by such requirements. We present the ancestral distance (AD) test to deal with these difficulties when analyzing two binary traits.

Characters that are common within particular sublineages but evolved independently in distantly related lineages pose particular problems. In the extreme, when all taxa in a particular sublineage have the characters of interest, correlated evolution cannot be detected by considering only the sublineage (Harvey and Pagel, 1991; Maddison, 1990). A large sample of taxa that spans distantly related lineages is required so that several independent evolutionary transitions of character states are included in the analysis. (see, e.g., Sanderson, 1993, for considerations of power). A random sample of taxa is also important, as samples that are biased towards taxa with characters of interest can alter inferred patterns of character evolution (Ackerly, 2000; Sillen-Tullberg, 1993).

Large, random samples of taxa themselves present further challenges. First, collection of complete sets of comparative data for correlation analyses can be time consuming and expensive (Vamosi et al., 2003). Second, inference of phylogeny for large data sets becomes especially difficult due to lack of appropriate loci that can resolve both deep and shallow phylogenetic events, alignment issues, problems assessing orthology and homology, and violation of assumptions of character independence when loci that are under strong selection are used for phylogenetic inferences (Phillipe et al., 1994; Rokas et al., 2003). Third, several comparative methods require ancestral state reconstruction (e.g., Maddison, 1990; Ridley, 1983). When the phylogeny is large with ancient divergence events (or trait evolution is fast relative to speciation), accurate reconstruction of ancestral states can be difficult, if not impossible (Mossel, 2003).

A trend among current systematic studies is to aim for larger and more fully resolved phylogenies. However, large-scale phylogenies are simply not available for several comparative questions, in particular, those involving rare and sporadic traits. Supermatrix and supertree methods (Driskell et al., 2004; Sanderson et al., 1998) may eventually provide the tools to piece together the entire tree of life from different types of molecular data and different phylogenetic analyses. Even with these methods, and after inference of topology, accurate estimation of branch lengths may still present hurdles. Large multilocus phylogenies have also been reconstructed, but such phylogenies typically include a nonrandom sample of taxa emphasizing taxonomic breadth rather than depth of a relatively small sample of total diversity (e.g., Soltis et al., 1999, with $\approx$0.3% of angiosperm species). Taxa with rare, sporadic traits are still underrepresented in recently published large-scale phylogenetic analyses. Several traits of reproductive and ecological importance are rare and sporadic. In plants, such traits include breeding systems (e.g., dioecy), pollination syndromes

TABLE 1. Current comparative tests compared. Column labels: Char. Model = assumes explicit model of character evolution; Const. Rate = based on constant rate model of character evolution; Spec. Model = assumes explicit model of speciation; Anc. States = reconstruction of ancestral states involved; Comp. Data = requires complete character information; Res. = requires resolved phylogeny; Lrg. = easily applicable with large taxon samples; Rob. = robust to deviations from assumptions of model; Y = yes; N = no; NA = not applicable; ? = unclear; dep. = dependent; indep. = independent; ∼ = not crucial. Text provides method references.

Method	Example software	Null hypothesis	Char. Model	Const. Rate	Spec. Model	Anc. States	Comp. Data	Res.	Lrg.	Rob.
IC	CAIC	Evolutionary changes in one trait are not correlated with changes in another.	Y	Y	N	Y	Y	Y	N	Y
CC	MacClade	Number of gains of dep. character in presence of indep. character no different than expected by chance.	N	NA	N	Y	N	Y	N	NA
Omnibus	DISCRETE	Equal probability of data under models with and without correlation.	Y	Y	N	N	Y	Y	N	?
Ancestral distance	By hand or Monte Carlo	Relatedness of taxa with dep. trait to taxa with indep. trait equals that of taxa lacking dep. trait to taxa with indep. trait.	Y	Y	Y	∼Y	N	N	Y	Y
Pairwise comparisons	By hand	Pattern of co-occurrence of trait transition due to chance.	N	N	N	N	N	N	Y	NA
Bayesian methods	Software not readily available	Estimate probabilities of correlation based on models of character evolution	Y	∼N	Y	Y	∼Y	Y	?	?

(e.g., wind, water, bat pollination), certain fruit dispersal mechanisms, leaf characters (leaf margin, shape, arrangement, venation, vestiture, nectaries, etc.), growth habits (vines, succulents, etc.), and floral traits (nectar spurs, color, etc.). Investigations of the evolutionary causes and consequences of dioecy exemplify issues associated with rare, sporadic traits (Bawa, 1980; Carlquist, 1974; Cox, 1988; Donoghue, 1989; Givnish, 1980; Rowley, 1987; Vamosi et al., 2003). Dioecy occurs in approximately 6% of angiosperm species, 7% of angiosperm genera contain dioecious species, and over half of the flowering plant families possess dioecious species (Heilbuth, 2000; Renner and Ricklefs, 1995). It is a trait that is relatively uncommon, yet widely distributed, having evolved multiple times.

Studies investigating dioecy have focused on either analyzing nested patterns of character state variation among taxonomic rank levels (e.g., Givnish, 1980; Renner and Ricklefs, 1995), or using fully resolved phylogenies (e.g., Donoghue, 1989; Vamosi et al., 2003). Both approaches have problems and benefits. Even when rank-level classifications are consistent with the underlying phylogeny, they represent a very coarse-grained estimate of relationships. Moreover, some arbitrariness is commonly acknowledged in rank classifications. Nonetheless, a well-defined statistical structure exists in rank classifications (Scotland and Sanderson, 2004), and rank-level classifications are readily available for large lineages of life. Fine-grained phylogenetic information is lost, but accurate rank classifications still provide information about nested patterns of descent. On the other hand, fully resolved phylogenies provide a fine-grained depiction of relationships, but their inference can be difficult for reasons described above. Although techniques have been offered to resolve polytomies in various ways prior to application of software (Grafen, 1989, 1992; Pagel, 1992), or to generate trees with branch lengths (Losos, 1995; Martins, 1996a), these methods require at least some phylogenetic data for taxa considered so that some information is available to reconstruct the phylogeny or set of phylogenies (Martins, 1996a).

Although the number of publications with phylogenies is growing at an exponential rate, there are many species-rich groups with no phylogenies assembled (Pagel, 1999). The AD test takes advantage of currently available smaller phylogenies, instead of one global, fully resolved phylogeny. Phylogenetic information about entire groups of organisms can be missing, as only "local" information about the phylogeny surrounding a sampled taxon is required. This test was based on results from probability theory, and tested via simulations to evaluate its robustness. Using a Monte Carlo approach, its behavior was also examined using a relatively small, well-studied data set (Lutzoni and Pagel, 1997), and potential applications are discussed. A program that implements the Monte Carlo approach is available (http://ag.arizona.edu/∼dhearn/AncestralDistance/).

Framework for Proposed Test and Test Procedure

Of the issues involved in choosing a comparative test (Table 1), one of the most important is the match between the test and the biological question (Maddison, 1990). The particular question our test addresses is "Are taxa with a dependent trait more likely to be more closely related to taxa with an independent trait than expected by chance alone?" The test relies on the observation (Maddison, 1990) that when the evolution of traits is correlated, characters appear clustered in a phylogeny. For example, clusters would occur either when the dependent character is more likely to evolve in the presence of the independent character, or when losses of

the dependent character are less likely in the presence of the independent character. Alternatively, clusters appear when the two traits are developmentally linked and are both gained or lost simultaneously.

The AD test records the amount of evolutionary separation between taxa that share traits of interest. When the dependent character is more likely to evolve (higher gain) or persist (less loss) in the presence of the independent character, we show that closely related taxa are more likely to have both the independent and dependent characters than expected by chance alone.

Before describing the test procedure, we define a few new terms (see Fig. 1). The *ancestral distance* is the time separating a particular extant taxon from its most recent ancestor (MRA) that has one or more extant descendants that have an independent character. A *target sample* is a random sample (sample design discussed below) of taxa that have the dependent character, whereas a *nontarget sample* is a random sample of taxa that lack the dependent character. ADs are 0 when a sampled taxon has the independent trait. Although we use the language "dependent" and "independent," the test is not meant to imply a causal connection between the characters. Other causative factors may be responsible for generating associations between characters. Apparent increases or decreases in rates of gain or loss of the dependent character

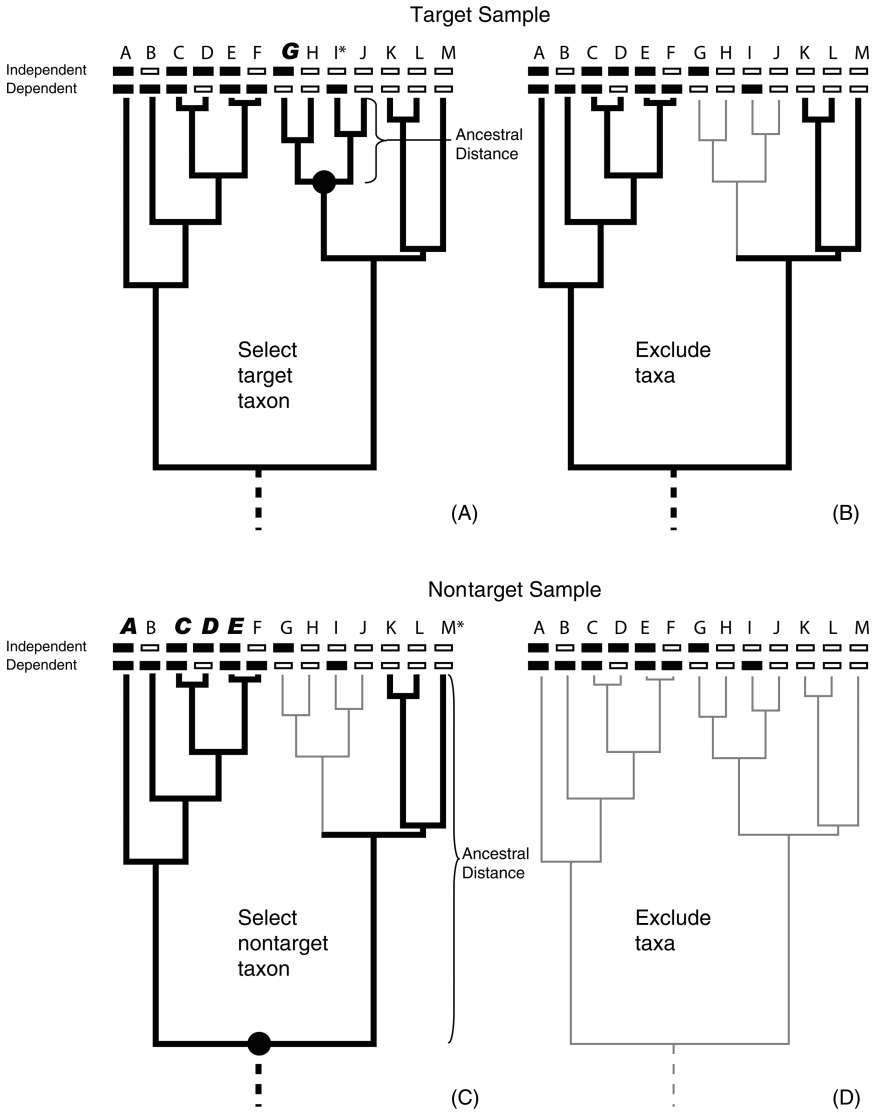

FIGURE 1. Ancestral distance method. The independent and dependent characters take two states (filled rectangle, open rectangle). (A) A random target taxon, I*, is selected that has the dependent character (filled rectangle). The time (ancestral distance) to the MRA (black dot) with one or more descendants with the independent character (i.e., taxon G in bold italics) is recorded. (B) All descendants of the MRA are removed from sampling (light grey) so that only independent origins of the dependent and independent character are sampled. (C) A random nontarget taxon, M*, is selected that lacks the dependent character (open rectangle). The time (ancestral distance) to the MRA with one or more descendants (not including the excluded taxon, G) with the independent character (i.e., taxa A, C, D, E in bold italics) is recorded. (D) All descendants of the MRA are removed from sampling. This process (A) to (D) is repeated until either no taxa remain or no taxa have the dependent character or independent character. Dotted lines at base of tree indicate that it is part of a much larger tree.

may be mediated by an uncharacterized factor that is itself connected to the independent trait (Read and Nee, 1995).

After initially constructing a list of taxa with the dependent character, the sampling procedure is as follows: (1) randomly select a taxon with the dependent character (taxon in the target sample); (2) locate a published phylogeny with this taxon or a close relative with the dependent character (if no phylogeny is available for a sampled taxon and close relatives, steps 1 and 2 can be repeated until an appropriate phylogeny is available); (3) collect comparative data for the sampled taxon and closest relatives, and record the taxon's AD; (4) exclude taxa so that only evolutionarily independent origins of the traits are sampled (see below); (5) repeat steps 1 to 4 alternating between taxa in the target and nontarget samples (Fig. 1; ordering of the target and nontarget sampling can be randomized). Finally, compare the ADs between the target and nontarget samples using a standard statistical test (e.g., Wilcoxon two-sample test or *t*-test).

The fourth step helps to ensure that taxa are sampled from evolutionarily independent origins (i.e., gains or losses due to separate mutation events) of the traits of interest. At step 4, a "weak" inference about ancestral states is required. For both the target and nontarget samples, the most recent ancestor that gave rise to the independent trait is located, and descendants from this ancestor are excluded from further sampling. Also, for the target sample, the ancestor that gave rise to the dependent trait is located, and all its descendants are likewise excluded. When origins of the independent and dependent characters are more recent than the MRA (dots in Fig. 1), all descendants of the MRA are excluded. The inference of ancestral states is "weak" in the sense that the test is conservative when all descendants of an even deeper ancestor are excluded (Fig. 1B, D). ACCTRAN parsimony (Farris, 1970) is an appropriate method to reconstruct ancestral states as it will infer the fewest (Huelsenbeck et al., 2003) and deepest origins (Maddison and Maddison, 1992).

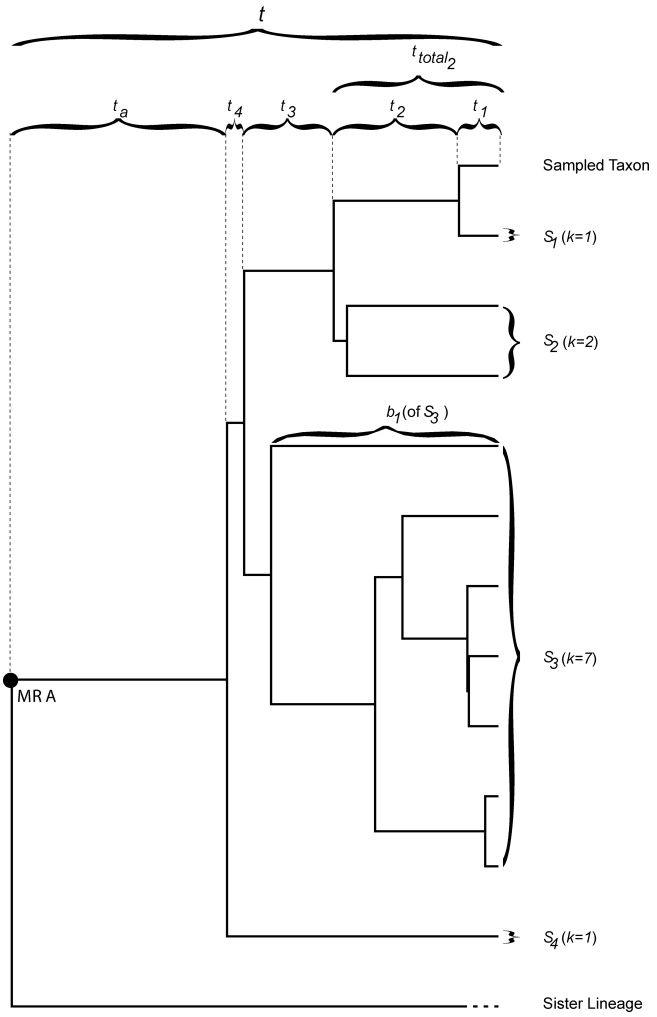

FIGURE 2. Notation for times, subclades, and branch lengths used to derive the distribution of ancestral distances. Without loss of generality, the top-most taxon is sampled (Sampled Taxon). For this particular phylogeny, the ancestral distance of the sampled taxon is t when no taxa in subclades s_1 to s_4 have the independent character, but the clade that is sister (Sister Lineage) to subclades s_1 to s_4 has one or more taxa with the independent character. The times, t_1 to t_a, represent the waiting times between speciation events that separate the sampled taxon from the MRA. The total time from the ith speciation event until present is t_{total_i}. The lengths of the $2k-1$ branches of a subclade with k terminal taxa are represented by b_1 to b_{2k-1}.

MATERIALS AND METHODS

Simulations

In Appendix 1 and Figure 2, we derive the probability distribution of ADs under a pure birth (Yule) process of phylogenesis and a Markov model of character evolution. We show that this distribution is equivalent for the target and nontarget samples when characters are uncorrelated (consequently, the proportion of taxa with ADs equal to 0 is the same for the target and nontarget samples), and that the expected frequency of taxa with ADs equal to 0 is the stationary frequency of the independent character when character evolution is fast relative to speciation (fast character evolution, slow speciation [FCSS] conditions).

We perform simulations (Table 2) to examine the robustness of these results when the Yule model of phylogenesis and Markov model of character evolution are violated. During these simulations, we measure the AD using the time, the number of speciation events (node depth), and the total branch length separating a sampled taxon from the MRA that has one or more extant descendants with the independent trait. We expect the ADs of the target sample to be the same, on average, as the ADs of the nontarget sample when characters are uncorrelated and shorter, on average, than the nontarget sample when characters are positively correlated. We expect that any random deviation from assumptions that is itself independent of the characters of interest (e.g., molecular rate heterogeneity, taxon subsampling, extinction, altered distributions of branch lengths) will affect

TABLE 2. Simulation Information. Set = grouping of simulations; Cor. Chars = characters are correlated; Taxa Indep. = states of taxa are independent and do not depend on ancestral state; No. Taxa = number of taxa in simulated phylogeny; Reps. = number of simulation replicates; Spec. Rate = rate of speciation; Var. Rate = variable substitution rate; Exp. No Diff. = no difference between mean target and nontarget ancestral distances expected; M = medium correlation; H = high correlation; E = empirical data set; NA = not applicable; V = variable; Y = yes; N = no. Set 1 explores varying sizes of phylogenies for uncorrelated characters, set 2 explores varying strengths of correlation, and set 3 applies the ancestral distance method to an empirical data set (Lutzoni and Pagel, 1997). Variable substitution rates, taxon sampling, and phylogenetic structure (ranks) are explored in set 4.

Set	Cor. Chars	Taxa Indep.	No. Taxa	Reps.	Spec. Rate	Var. Rate	Exp. No Diff.
1	N	Y	30	10,000	0.00001	N	Y
1	N	Y	3000	10,000	0.00001	N	Y
1	N	N	30	10,000	1	N	Y
1	N	N	3000	10,000	1	N	Y
2	M	N	3000	10,000	1	N	N
2	H	N	3000	10,000	1	N	N
3	E	E	30	10,000	NA	N	N
4	N	N	3000	10,000	1	Y	Y
4 (ranks)	V	N	V	1000 each parameter combination	1	N	V

the ancestral distances of the target and nontarget samples the same, on average.

For simulations involving uncorrelated characters, instantaneous rates of character state transition were randomly selected, but rate of gain in the dependent character ranged between 0.05 and 0.08, whereas loss was set between 0.5 and 0.8. These restrictions were chosen so that stationary frequencies of the characters were around 0.1. Rates of gain of the dependent character in the presence of, or loss in the absence of, the independent character were set to be 2 to 10 times greater than rates of gain in the absence of, or loss in the presence of, the independent character for medium correlation. Values were set to be 20 times higher for high correlation.

Two trees and character sets were simulated for each replicate. A single taxon with the dependent character was randomly selected (target sample) from the first tree and character set, as was a single taxon lacking the dependent character (nontarget sample) from the second tree and character set, and their ADs were recorded.

We generate branch lengths by modeling substitution as a constant-rate Poisson process (Zuckerkandl and Pauling, 1965). For these constant rate simulations, the number of substitution events, b_ℓ, in a molecular sequence of length ℓ along a branch of time duration t was therefore sampled from a Poisson distribution with parameter $\mu = \lambda \cdot \ell \cdot t$ so that $\mathbb{P}(b_\ell = i) = e^{\mu}\mu^i/i!$. We set $\lambda = 1.0$ and $\ell = 100$ for all simulations. When substitution events are modeled as a constant rate Poisson process, we expect the AD as measured by branch length to be approximately a constant multiple of the AD measured by time because the expected number of substitutions along a branch is a constant multiple of time and sequence length, ℓ: $E[b_\ell] = \mu = \lambda \cdot \ell \cdot t$.

The third set of simulations (Table 2) evaluates the AD method with empirical character data of known correlation. These data tested the hypothesis that evolutionary origins of lichenization in *Omphalina* and relatives are associated with increased rates of molecular evolution (Lutzoni and Pagel, 1997). We used the provided tree and absolute node ages, and the rate matrix estimated under maximum likelihood by Discrete to simulate character evolution. For each of the two trees per replicate, we sampled one taxon, alternating between taxa in the target sample and nontarget sample.

A fourth and final set of simulations (Table 2) examines the robustness of the method to deviations from the Markov assumptions by evaluating the effects that random taxon sampling, substitution rate heterogeneity, phylogenies collapsed to rank level, and branch times that are non-Yule (i.e., not exponentially distributed) have on AD statistics. In general, branch lengths will not be ultrametric when molecular evolution rates vary, unlike those under a constant rate Poisson process, but we still expect that the distributions of ADs for target and nontarget samples to be the same when rates vary. We modeled rate heterogeneity as a random walk on a tree, paralleling how Thorne et al. (1998) and Kishino et al. (2001) modeled substitution rate evolution to infer divergence times. Substitution rate is constant along a branch, but varies from branch to branch, and the logarithm of the rate of molecular evolution is normally distributed. Starting with rate $= e^1$ at the root, the rate, r_d, at a descendant branch changes from the ancestral branch, r_a, according to the random walk

$$r_d = e^{ln(r_a)+r\cdot t\cdot Z} = r_a \cdot e^{r\cdot t\cdot Z}$$

where $Z \sim N(0, 1)$, t is the time along the branch from the ancestral node to the descendant node, and r is the rate of the random walk. We set $r = 1.0$ for all 10,000 replicates.

We also considered the performance of the method when only a taxonomic rank level classification is available. We used the hollow-curve data of Scotland and Sanderson (2004), which depict the number of species in genera, number of genera in families, and families in orders in plants. A taxonomic order was randomly selected from the hollow-curve data. Then a Yule tree with 500 taxa was simulated, and the number of families in that order were randomly sampled and all other taxa were removed from the tree using software adapted from Mesquite (Maddison and Maddison, 2003). For each of

those families, genera were similarly evolved, and for each genus, species were similarly evolved.

On each simulated tree, continuous character data were evolved under Brownian motion. The independent character was simulated as $n_d = n_a + r \cdot t \cdot Z$. The rate of the Brownian motion process is r, and t is the time separating the state at the daughter node, n_d, from the state at the ancestral node, n_a. The Brownian rate was set to 1.0 for all simulations, and the root node character state was set to 0.0. The continuous-valued states were dichotomized using threshold values to produce specific binary character frequencies. When characters were independent, the evolution of both characters was simulated separately on the same tree. For correlated characters, the state of the dependent character was sampled from a normal distribution with mean equal to the state of the independent character. The variance of the normal distribution determines the level of correlation. High correlation corresponds to low variance, whereas low correlation corresponds to high variance; with 0 variance, the state of the dependent character is equal to the state of the independent character and with high (approaching infinite) variance, the characters are effectively independent.

We evolved both a correlated set of characters and an uncorrelated set for each replicate tree. Let the difference between the AD of a randomly sampled target taxon and the AD of a nontarget taxon be d_c and d_{uc} for correlated and uncorrelated characters, respectively. Let D represent the difference between d_c and d_{uc}. D is expected to approach 0 as the correlation between the characters decreases, because d_c will approach d_{uc}. For each replicate, one D value was calculated using fully resolved phylogenies, and one D value was calculated for ADs using phylogenies that were collapsed to taxonomic ranks. The rank AD (rank in which the independent character first appeared) was recorded as 0 for species, 1 for genera, 2 for families, and 3 for orders.

Test Distributions

We ran 10,000 replicates, each with two trees (one for sampling a target taxon and one for a nontarget taxon), for all simulations except the rank-level simulations (Table 2). Using sample sizes of 10, we randomly divided the 10,000 replicates into 1000 target and 1000 nontarget samples. The 1000 differences between means of the target and nontarget samples comprise the *test distribution*. The expected difference between means for uncorrelated characters is 0, whereas the expected difference is positive for correlated traits. The null hypothesis of no correlation can be rejected if 95% or more of the 1000 differences are greater than 0.

RESULTS

Table 2 summarizes the simulations and the expected results. All results matched expectations, except when simulated trees were small. We expected the distributions of ADs to be identical for the target and nontarget samples when characters were independent. We compared AD distributions (estimated by 10,000 simulation replicates) using the Kolmogorov-Smirnov test as implemented in the R Statistical Language (R Core Development Team, 2005). The distributions of ADs of the target sample were indistinguishable from the ADs of the nontarget sample when characters were independent, as measured by time, branch length, or node depth, both under FCSS conditions (time: D = 0.0099, P-value > 0.71, length: D = 0.0086, P-value > 0.85; node depth: D = 0.006, P-value > 0.99), and non-FCSS conditions (time: D = 0.012, P-value > 0.46, length: D = 0.0109, P-value > 0.59; node depth: D = 0.0092, P-value > 0.79; Figs. 3 and 4).

We expected the test distribution to be centered at 0 when characters were independent and for it to be shifted to the right of 0 for correlated characters. The test distribution was centered at 0 when characters were uncorrelated (Fig. 4F). When characters were weakly correlated, 5.5% of the test distribution area was less than 0 (Fig. 4E). When characters were highly correlated, 0% of the test distribution was less than 0 (Fig. 4D).

The test assumes that the phylogeny is arbitrarily large, as ADs range across all positive values (Appendix 1). With small phylogenies (30 taxa) and independent characters under non-FCSS conditions, there was a detectable difference between the target and nontarget distributions when measured by time or branch lengths. However, distributions of node depths were still indistinguishable (times: D = 0.0274, P-value < 0.05; lengths: D = 0.0278, P-value < 0.001; nodes: D = 0.0165, P-value > 0.13). When FCSS conditions held, distributions were indistinguishable for small phylogenies. In all cases (small and large phylogenies, FCSS and non-FCSS conditions), test distributions were centered at 0 for independently evolving characters (Fig. 3G–I).

We predicted that the frequency that taxa had the independent character (i.e., when the AD was 0) would be π_{Y_1} (see Appendix 1) when characters were independent and FCSS conditions held. The simulated frequency was not distinguishable from π_{Y_1} ($\pi_{Y_1} = 0.0748$, target sample $\pi_{Y_1} = 0.0741$, nontarget sample $\pi_{Y_1} = 0.0743$, $z = 0.054$, $n = 10{,}000$, P-value > 0.47).

As expected, under a constant rate Poisson process, the distribution of ADs based on branch lengths was approximately a constant scaling of the distribution based on time (Fig. 5). No differences between target and nontarget distributions were detected when rates were heterogeneous and characters were uncorrelated (Fig. 6; times: D = 0.0153, P-value > 0.19; lengths: D = 0.018, P-value > 0.07; nodes: D = 0.0119, P-value > 0.47).

The AD test could not reject the hypothesis that lichenization was associated with increased molecular rate. The entire test distribution was greater than 0. In fact, all target samples had ADs equal to 0.

Finally, in the simulations that considered random taxon subsampling and taxonomic rank level (Fig. 7), correlated characters could be detected best (highest values of D) when the presence of characters was infrequent (Fig. 7A, B, cells at the lower left of each box of cells) and characters were highly correlated (Fig. 7A, B, top

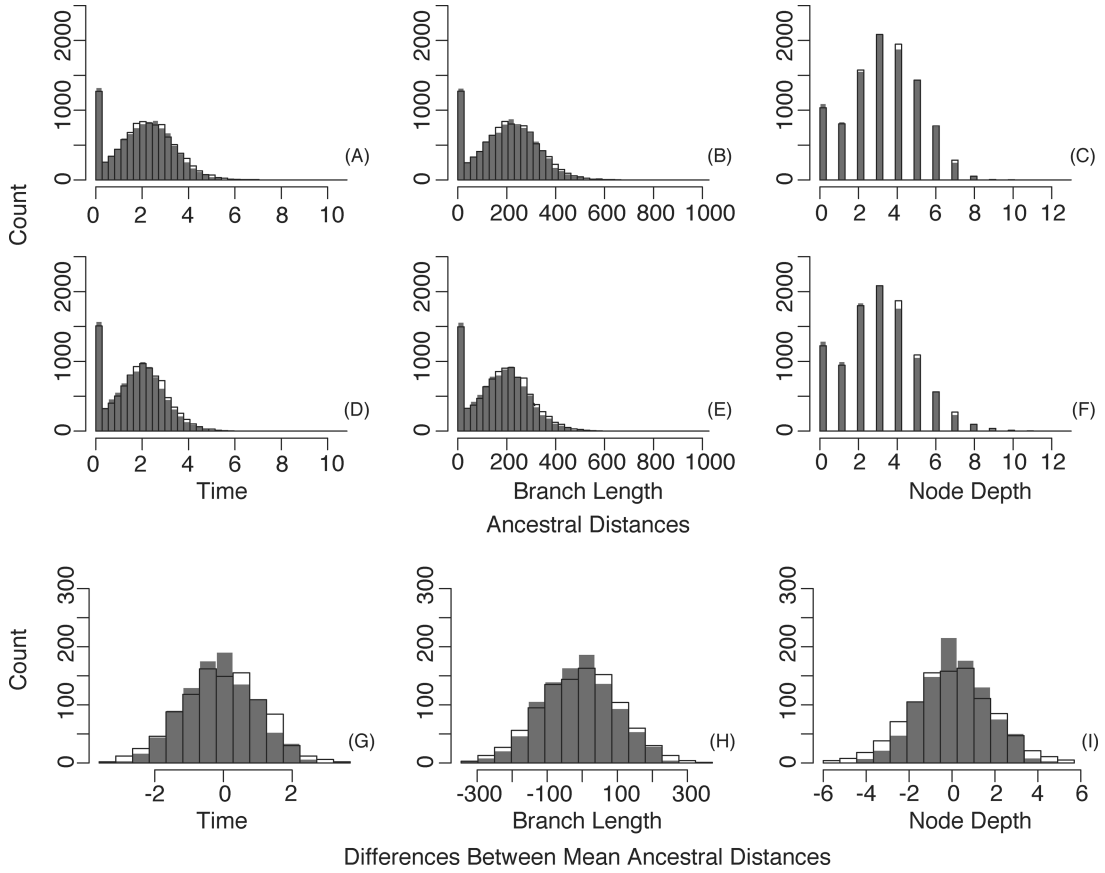

FIGURE 3. Histograms of ancestral distances for independent characters and fast character evolution relative to speciation: Set 1 of simulations. (A)–(F) Solid bars are ancestral distances of target taxa (i.e., taxa with the dependent character), and black borders represent ancestral distances of nontarget taxa (i.e., taxa lacking the dependent character). (A–C) Phylogenies of 3000 taxa; (D–F) phylogenies of 30 taxa. Ancestral distances are measured by time for (A) and (D), branch length for (B) and (E), number of speciation events (node depth) for (C) and (F). Test distributions (black borders: small phylogenies; solid bars: large phylogenies) are for times (G), branch lengths (H), and node depth (I).

rows of boxes). Correlated characters could be detected using rank depth, but with weak power (Fig. 7B; low, but positive values of D).

DISCUSSION

Strengths and Weaknesses of the Ancestral Distance Test

Measures of relatedness provide a means to detect correlated trait evolution. We introduced the ancestral distance (AD) as measured by time, node depth, branch length, or rank level, and derived a two-sample test based on its use. As comparative biologists are often limited to character data of extant species, inferences based on these data are dependent on assumptions about evolutionary processes (Pagel and Harvey, 1989), and in particular, about the branching process that gives rise to the tree of life. The AD test relies on two specific models of evolution—one depicting character evolution and one depicting phylogenesis. These models assume a constant rate of character evolution, a constant rate of species formation, and the standard Markov assumptions concerning independence of events: future character states of a lineage are based only on its current state (the process is memoryless), descendants inherit character states from their immediate ancestor, and lineages evolve independently after speciation (no coevolution). Although we have not evaluated rigorously the effects that extinctions, rate heterogeneity, and taxon sampling have on AD calculations, it is reasonable to expect that the null hypothesis (i.e., ADs of target and nontarget samples equal on average) will be appropriate when these effects are *random effects that are independent of characters of interest*. When such effects are independent of the characters, they are expected to affect both the target and nontarget ADs equally, on average. In all cases that we examined in which our model assumptions were violated (i.e., molecular evolution rates varied: Fig. 6; lineage times non-Yule, cf., Barraclough and Nee [2001]: Fig. 7A; random taxon subsampling: Fig. 7; pattern of clade formation similar to taxonomic rank structure of plants: Fig. 7B), the AD test behaved robustly (i.e., is relatively free of model assumptions).

Simulation results (e.g., Figs. 3 and 4) also suggest that the test performs well both when character transition rates are fast relative to speciation (FCSS) and phylogeny is relatively unimportant (Freckleton et al., 2002; Rheindt

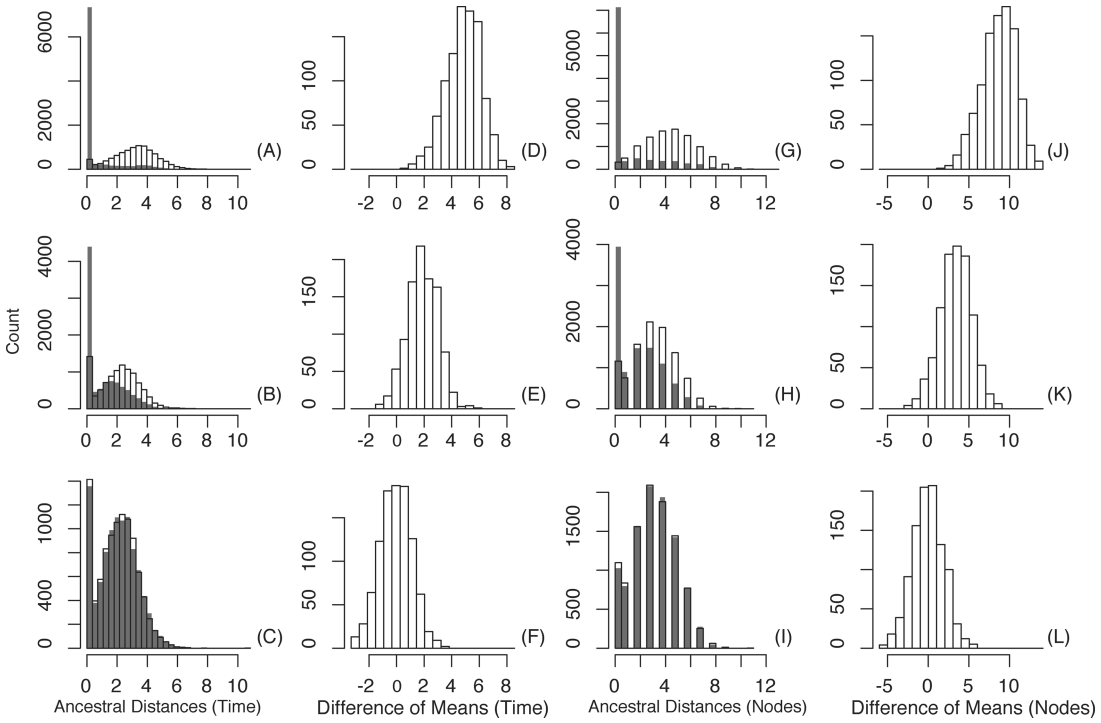

FIGURE 4. Histograms of ancestral distances for uncorrelated to highly correlated characters: set 2 of simulations. The figure is divided into four columns and three rows. The first and the third columns are histograms of ancestral distances for high (top row), medium (middle row), and no (bottom row) correlation. Solid bars are ancestral distances of target taxa, whereas black borders are ancestral distances of nontarget taxa measured by time (first column) or node depth (third column). The second and fourth columns present the test distributions approximated by 1000 samples (i.e., 10,000/10). The *x*-axes are the same within a column.

et al., 2004) as well as conditions when character evolution is slower (non-FCSS). Under FCSS conditions, the proportion of taxa in the target sample with ADs equal to 0 can be compared to this proportion of the nontarget sample. Both probability theory and simulations reveal that when characters are uncorrelated and evolved under FCSS conditions, these proportions will be equal on average, and a test of proportions can be applied.

In general, FCSS conditions will not hold, and character states of daughter taxa will depend on the states

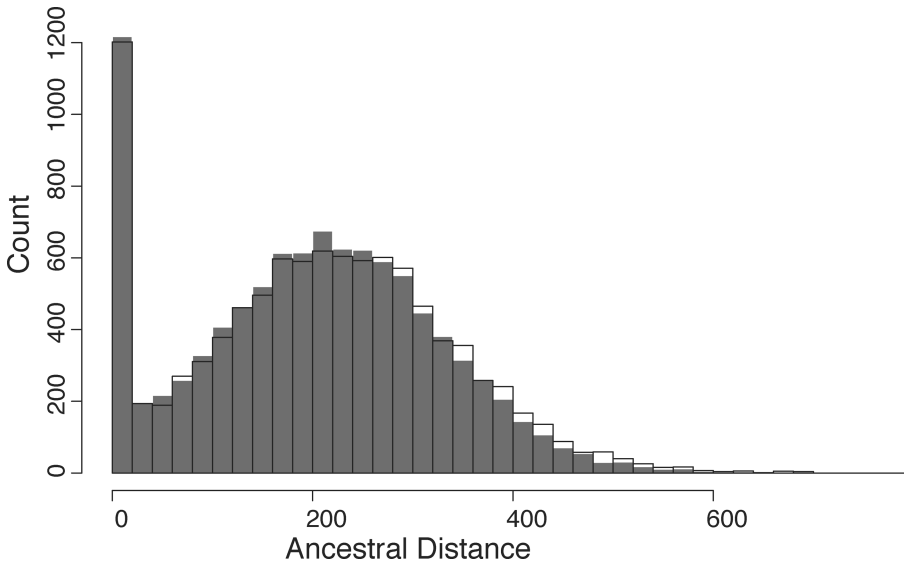

FIGURE 5. Comparison between ancestral distances calculated by time and by branch length. A constant scaling of ancestral distances as measured by time (solid bars) approximates ancestral distances as measured by branch length (black borders) under a constant rate Poisson process.

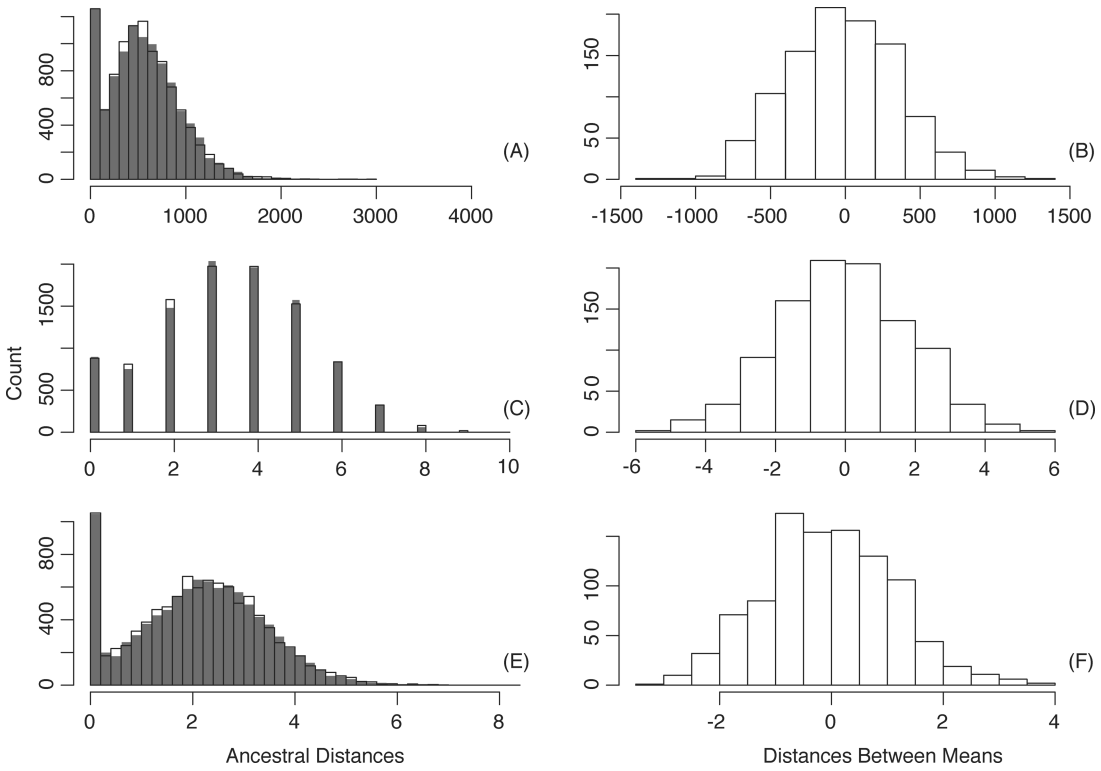

FIGURE 6. Ancestral distances when substitution rates vary and characters are uncorrelated: Set 4 of simulations, part 1. Left column presents histograms of ancestral distances for target (solid bars) and nontarget (black borders) samples when measured by branch length (A), node depth (C), and time (E), for uncorrelated characters. Right column presents the test distributions for branch lengths (B), node depths (D), and times (F).

of their ancestors. Most comparative studies will therefore require information about the pattern of descent with modification, stored in the phylogeny (Felsenstein, 1985; Freckleton et al., 2002). The math derivations and simulations indicate that ADs of the target sample are expected to be the same as the ADs of the nontarget sample when characters are uncorrelated. The null hypothesis of no difference between ADs is based on this result.

Ideally, comparative biologists would have at their disposal a fully resolved, accurate phylogeny and complete character data for all taxa of interest. In such a case, more powerful and flexible methods are available, such as the omnibus test (Pagel, 1992, 1994) for dichotomous traits and independent contrasts (IC; Felsenstein, 1985; Purvis and Rambaut, 1995) for continuous traits. However, for rare and sporadic characters that necessitate a large sample of taxa, this ideal may not yet be available due to the lack of a resolved phylogeny or complete character data. Recent Bayesian methods take into account uncertainty in the tree, branch lengths, substitution model parameters, and character mappings (Huelsenbeck et al., 2003; Huelsenbeck and Rannala, 2003), and these methods directly estimate the marginalized posterior distribution of correlation parameters under a particular model of character evolution or compare observed posterior values to expected values under a null. Bayesian analyses rely, in part, on the specification of a prior model of character evolution and phylogenesis. How many new data points are required to "swamp" the effect of the prior is an open question (e.g., Kou et al., 2005), and currently available Bayesian methods require some data for all taxa. The AD test, in contrast, gleans data from phylogenies for which comparative data are at hand and excludes taxa with no data.

The AD approach is most powerful when the frequencies of the characters are low and the number of taxa is large (Fig. 7). This is expected; when the independent character is widespread and common throughout the phylogeny, target taxa and nontarget taxa will both be close relatives of taxa with the independent character, so ADs for both target and nontarget taxa will be short. In the extreme, when all taxa have the independent character, ancestral distances will be 0 for all sampled taxa (Harvey and Pagel, 1991). This method also relies predominantly on extant character data, so unlike current Bayesian methods, it does not need to consider uncertainty in ancestral states. The AD test requires some inferences about ancestral states, but the "weak" inference is conservative when parsimony is used or when even deeper nodes are chosen for taxon exclusion (see Framework for Proposed Test and Test Procedure). The AD test allows a researcher to be ignorant of the states of the independent character before analysis begins. Collection of these data is required for close relatives of sampled taxa only. Ignorance of certain taxa with the dependent character is also possible (as long

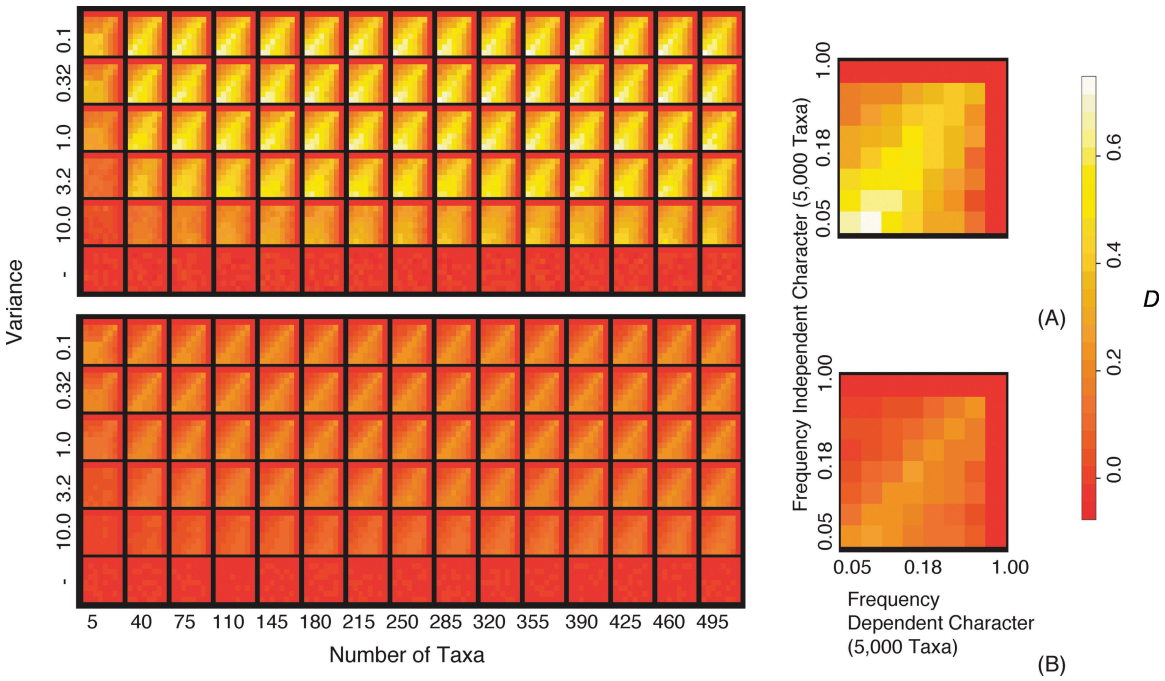

FIGURE 7. Analysis of rank-level ancestral distances and the effect of random taxon sampling: Set 4 of simulations, part 2. Both (A) and (B) summarize five dimensions. (A) Fully resolved phylogenies. (B) Phylogenies that have been collapsed to taxonomic rank level. Within each of the squares of cells surrounded by black borders (enlarged at right for phylogenies of 5000 taxa), the frequencies of the characters vary. The *x*-axis is the frequency of the dependent character and the *y*-axis is for the independent character. The rows of black-bordered squares represent different strengths of correlation as measured by variance (see text). The top row has highest correlation and the row indicated by "—" has no correlation. Columns of squares represent differently sized phylogenies. The lighter a particular cell, the greater the ability to detect correlation, the higher the value of *D* (see text).

as the ignorance is not biased). Moreover, use of smaller phylogenies can reduce problems of inference associated with large scale phylogenies: sequence alignment, long branches and saturation, sequence and locus availability and appropriateness, branch length inference, cost and feasibility, and character information availability.

Sampling Design Considerations

One of the main assumptions of the AD method is that other inference methods generate unbiased phylogenies and divergence times. When this occurs, ADs using time or node depth can be compared across local phylogenies that were reconstructed using different molecular loci, because it is expected that random errors in the phylogeny that are not themselves correlated with, or due to, the characters of interest will affect target and nontarget samples the same, on average.

Thus, another assumption of the AD method is that the structure of the classification or the phylogeny is not based on (i.e., not correlated with) the characters of interest (or characters that are correlated with them). Circularity of this type (i.e., inferences about characters of interest that are reliant on the structure of the classification, which is based on the characters of interest) is an issue for other comparative methods (Hull, 1967) and may be particularly important to consider when using rank depth as the measure of AD. It is assumed that the rank classification is consistent with the underlying phylogeny. When the rank classification is inconsistent with the underlying phylogeny, separate origins of characters of interest may be considered as part of the same taxonomic grouping, or a single origin of a trait may be split into different taxonomic groupings. When this occurs, samples that are treated as separate origins may not be independent. This same problem applies to incorrect phylogenies. As our knowledge of the structure of the classification of life improves, these problems will be alleviated. Statistical power is also very low when ranks are used. However, when the correlation between traits is particularly strong, and the above assumptions are met, taxonomic rank levels can be used to detect correlation as witnessed by positive values of *D* when traits are correlated and values of *D* equal to 0 when traits are uncorrelated (Fig. 7B). Because of the highly conservative nature of the test when ranks are used, if the null is rejected, the result is very strong. Use of ranks allows a researcher to address the question, "Do taxa with the dependent and the independent characters co-occur in a higher rank (measured from order up to species) more frequently than expected by chance?"

Unless it is known that the rank classification is independent of the characters of interest, and because the power of the AD test is low when rank depth is used, use of ADs as measured by time, node depth, or branch length are recommended instead. Recent studies show

substantial molecular rate heterogeneity (e.g., Bousquet et al., 1992; Bromham, 2002; Herbert et al., 2002; Hoegg et al., 2004; Lutzoni and Pagel, 1997). Although rate variation is not problematic for the AD method (provided that evolutionary change in molecular substitution rate is not correlated with changes in rates of evolution of characters of interest; Fig. 6), the link between molecular rate heterogeneity and morphological evolution is unclear except in a few instances (e.g., Bromham, 2002; Herbert et al., 2002; Lutzoni and Pagel, 1997); therefore, it may be prudent to avoid the use of branch lengths because branch lengths may be correlated to the rates of change of characters of interest. ADs measured by node depth may be more appropriate; only the approximate phylogenetic branching pattern is required for use of node depth, and minimal consideration of time or rate is required. The AD test assumes that trees can be arbitrarily large, and ADs based on node depth are also robust when small trees are used (Fig. 3D–F).

When the assumptions of unbiased phylogenies or times, noncircularity, and random taxon samples are met, our results permit the use of several much smaller (local) phylogenies rather than a fully resolved phylogeny of all taxa in a group of interest (global). If a selected phylogeny is not of sufficient size to find the AD, either a different taxon and corresponding phylogeny can be sampled, or the distance from a sampled taxon to the root of the tree can be recorded in place of the actual AD. Effectively, this procedure will force taxa in such insufficiently large phylogenies to be more closely related to taxa with the independent character than they really are. Use of the root distance in these small phylogenies will make the test conservative because it will tend to shorten the ADs of the nontarget taxa in particular and make them more similar to the ancestral distances of the target taxa. In general, when testing for positive correlation, recording a smaller AD for the nontarget sample will produce conservative results and reduce the number of taxa that are excluded (Fig. 1).

Statistical Power, Biological Questions, and Null Models

There is a multitude of comparative tests available to the systematist. Our primary goal in designing this test was to provide a flexible method (i.e., one that can use different measurements—node depth, branch length, time, rank—depending on data availability) for use with comparative studies requiring large samples of taxa. Most methods require fully resolved phylogenies, and many require branch lengths as well (Table 1). Each test also presents a different null hypothesis that is appropriate for evaluating different biological questions (Table 1).

Pairwise comparison methods for binary characters are among the few methods that also can be applied to partially unresolved phylogenies (Maddison, 2000; Read and Nee, 1995). Read and Nee (1995) argue that for pairwise comparisons, only those taxon pairs in which both characters change state provide appropriate replicates for statistical tests. Under their null model, the state of one binary character is equally likely to be associated with either state of the other character. A binomial sign test examines whether two character states appear to be associated more often than expected by chance alone. There are, however, examples of correlated processes of evolution for which such a null is inappropriate. For example, consider two characters that are different realizations of the same developmental program expressed in different places (homeosis). When selection pays little heed to where the traits are expressed or one trait is a preadaptation for the other, one trait may be equally likely to be associated with either the presence or absence of the other trait. The null for pairwise comparisons tests would not be rejected even though the evolution of the traits is tightly linked. The AD test, however, can detect an association due to homeosis or other transference of function because instead of focusing on change in character state, it focuses on relatedness of taxa. This example of homeosis suggests that the AD test can also deal with different types of biological questions than the pairwise comparisons test.

Both the AD test and the pairwise comparisons test can be applied to unresolved phylogenies with missing comparative data, and both have relatively weak statistical power. Because the AD test is model based, a Monte Carlo approach is available to estimate P-values when fully resolved trees, branch times, and comparative data are available. Using the Monte Carlo approach, correlation can be detected on small phylogenies with few character transitions, as exemplified using the lichenization data of Lutzoni and Pagel (1997).

Methods such as IC (Felsenstein, 1985), CC (Maddison, 1990), and the omnibus test (Pagel, 1994) use information that is present across the whole tree and are therefore more powerful, whereas the sampling approach used by the AD test excludes large numbers of taxa and uses only local information around a sampled taxon (Fig. 1). Either a large sample of taxa, or the study of characters that evolve quickly (FCSS), is required by the AD test to provide sufficiently large sample sizes. The exclusion of information about taxa is both a bane and benefit. Because the AD method uses only "local" information in a phylogeny, it can be applied across thousands of species, something that is difficult for methods that require fully resolved phylogenies and complete character information (see Table 1). Furthermore, our simulation studies indicate that the AD technique may be able to detect relatively weak correlation with sample sizes as small as 10 (Figs. 3, 4, 7).

Homeosis and preadaptation in the example described above are important evolutionary processes (e.g., Sattler, 1988). For example, Olson (2003) proposed that lianas (woody vines) are evolutionary precursors to a certain type of stem succulent, the pachycaul succulent (see Rowley, 1987). He argues that stem parenchyma that lends flexibility to stems of lianas (Carlquist, 2001) is a preadaptation for parenchymatous storage tissue in pachycauls. This hypothesis, as yet, remains untested. Pachycauls represent a growth form that is relatively rare, evolved on several separate occasions, and appears in distantly related lineages, so the AD test is

appropriate for its testing. Pairwise tests are likely to reject this model of growth form evolution because many pachycauls still have lianoid stems (e.g., *Adenia karibaensis*, *Cyphostemma laza*), whereas other pachycauls that are clearly derived within lianoid lineages no longer climb (e.g., *Cyphostemma currori*, *Dendrosicyos socotrana*, *Ipomoea arborescens*). Future work will investigate the evolution of succulence across the eudicots (a large angiosperm lineage of ca. 160,000 species) with the aid of the AD technique.

The AD method was motivated out of a need based on the distributions of habit-related traits in plants. Such traits were found to have evolved on multiple occasions but are rare overall. However, often within particular "local" lineages, most taxa possessed the characters of interest, but the local lineages themselves were distantly related. This led to a problem: in a local lineage, the prevalence of characters made it difficult to detect a correlation, necessitating a huge sample of taxa. Seen at a local level, the characters were present in the majority of taxa, whereas at a global level, the characters were rare and sporadic. The option of finding a robust and fully resolved tree with taxa that had the traits of interest was pursued, but none of the currently available large phylogenies represented an adequate sample of taxa with the traits of interest. Moreover, none of the available large phylogenies consisted of a random sample of taxa, being biased toward taxonomic breadth. Hearn (2004) therefore applied the AD approach across the eudicots and found that stem succulents are significantly more closely related to tuberous plants than expected by chance. Implications for this robust and widespread pattern will be discussed in a future paper.

ACKNOWLEDGEMENTS

Special thanks go to Lucinda McDade for her mentorship. Two anonymous reviewers provided detailed comments that greatly improved the paper. Thank you also to Karen Schumaker, Robert Robichaux, David Maddison, and Brian Enquist for critical feedback. Thank you to David Maddison for particularly useful criticism early on and to Wayne Maddison for some helpful connections to other comparative tests. Thank you also to Kathleen Pryer for providing the time to work on the manuscript. This work was partially supported by NSF CAREER grant no. 0548153 to Huber and a NSF DDIG (0105127) to Hearn.

REFERENCES

Ackerly, D. D. 2000. Taxon sampling, correlated evolution, and independent contrasts. Evolution 54:1480–1492.

Barraclough, T. G., and S. Nee. 2001. Phylogenetics and speciation. Trends Ecol. Evol. 16:391–399.

Bawa, K. S. 1980. Evolution of dioecy in flowering plants. Annu. Rev. Ecol. Syst. 11:15–39.

Bousquet, J., S. H. Strauss, A. H. Doerksen, and R. A. Price. 1992. Extensive variation in evolutionary rate of rbcl gene sequences among seed plants. Proc. Natl. Acad. Sci. USA 98:7844–7848.

Bromham, L. 2002. Molecular clocks in reptiles: Life history influences rate of molecular evolution. Mol. Biol. Evol. 19:302–309.

Carlquist, S. 1974. Island biology. Columbia University Press, New York.

Carlquist, S. 2001. Comparative wood anatomy: Systematic, ecological, and evolutionary aspects of dicotyledon wood. Springer, Berlin.

Cox, P. A. 1988. Hydrophilous pollination. Annu. Rev. Ecol. Syst. 19:216–280.

Donoghue, M. J. 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. Evolution 43:1137–1156.

Driskell, A. C., C. Ane, J. G. Burleigh, M. M. McMahon, B. C. O'Meara, and M. J. Sanderson. 2004. Prospects for building the tree of life from large sequence databases. Science 306:1172–1174.

Edwards, A. W. F. 1970. Estimation of the branch points of a branching diffusion process. J. R. Stat. Soc. B Methods 32:155–174.

Farris, J. S. 1970. Methods for computing Wagner trees. Syst. Zool. 19:83–92.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. Am. Nat. 160:712–726.

Givnish, T. J. 1980. Ecological constraints on the evolution of breeding systems in seed plants: Dioecy and dispersal in gymnosperms. Evolution 34:959–972.

Grafen, A. 1989. The phylogenetic regression. Phil. Trans. R. Soc. Lond. B Biol. Sci. 326:119–157.

Grafen, A. 1992. The uniqueness of the phylogenetic regression. J. Theor. Biol. 156:405–423.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford, UK.

Hearn, D. J. 2004. Growth form evolution in *Adenia* (Passifloraceae) and a model of the evolution of succulence. PhD Dissertation. University of Arizona, Department of Ecology and Evolutionary Biology.

Heilbuth, J. C. 2000. Lower species richness in dioecious clades. Am. Nat. 156:221–241.

Herbert, P. D. N., E. A. Remiglio, J. K. Colbourne, D. J. Taylor, and C. C. Wilson. 2002. Accelerated molecular evolution in halophilic crustaceans. Evolution 56:909–926.

Hoegg, S., S. Vences, H. Brinkmann, and A. Meyer. 2004. Phylogeny and comparative substitution rates of frogs inferred from sequences of three nuclear genes. Mol. Biol. Evol. 21:1188–1200.

Hoel, P. G., S. C. Port, and C. J. Stone. 1987. Introduction to stochastic processes. Houghton Mifflin Company, Boston.

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

Huelsenbeck, J. P., and B. Rannala. 2003. Detecting the correlation between characters in a comparative analysis with uncertain phylogeny. Evolution 57:1237–1247.

Hull, D. L. 1967. Certainty and circularity in evolutionary taxonomy. Evolution 21:174–189.

Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18:352–361.

Kou, S. C., X. Sunney Xie, and J. S. Liu. 2005. Bayesian analyses of single-molecule experimental data. Appl. Stat. 54:469–506.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Losos, J. B. 1995. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst. Biol. 43:117–123.

Lutzoni, F., and M. Pagel. 1997. Accelerated evolution as a consequence of transitions to mutualism. Proc. Natl. Acad. Sci. USA 99:11422–11427.

Maddison, D. R. 1994. Phylogenetic methods for inferring the evolutionary history and processes of change in discretely valued characters. Annu. Rev. Entomol. 39:267–292.

Maddison, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? Evolution 44:539–557.

Maddison, W. P. 2000. Testing character correlation using pairwise comparisons on a phylogeny. J. Theor. Biol. 202:195–204.

Maddison, W. P., and D. R. Maddison. 1992. Excerpts from MacClade. Sinauer Associates, Sunderland, Massachusetts.

Maddison, W. P., and D. R. Maddison. 2003. Mesquite: A modular system for evolutionary analysis, vision 1.0. http://mesquiteproject.org.

Martins, E. P. 1996a. Conducting phylogenetic comparative studies when the phylogeny is not known. Evolution 50:12–22.

Martins, E. P. 1996b. Phylogenies and the comparative method in animal behavior. Oxford University Press, Oxford, UK.
Martins, E. P., and T. F. Hanson. 1996. The statistical analysis of interspecific data: A review and evaluation of the phylogenetic comparative methods. Page 22–75 *in* Phylogenies and the comparative method in animal behavior, E. P. Martins, ed.). Oxford University Press, Oxford, UK.
Mossel, E. 2003. On the impossibility of reconstructing ancestral data and phylogenies. J. Comput. Biol. 10:669–676.
Olson, M. E. 2003. Stem and leaf anatomy of the arborescent Cucurbitaceae *Dendrosicyos socotrana* with comments on the evolution of pachycauls from lianas. Plant Syst. Evol. 239:199–214.
Pagel, M. 1992. A method for the analysis of comparative data. J. Theor. Biol. 156:431–442.
Pagel, M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. P. R. Soc. Lond. B Biol. Sci. 255:37–45.
Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.
Pagel, M. D. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
Pagel, M. D., and P. H. Harvey. 1989. Comparative methods for examining adaptation depend on evolutionary models. Folia Primatol. 53:203–220.
Phillipe, H., A. Chenuil, and A. Adoutte. 1994. Can the Cambrian explosion be inferred through molecular phylogeny? Development (Suppl.) 120:15–25.
Purvis, A., and A. Rambaut. 1995. Comparative analysis by independent contrasts (CAIC): An Apple Macintosh application for analysing comparative data. Comput. Appl. Biosci. 11:247–251.
R Development Core Team. 2005. R: A language and environment for statistical computing. http://www.R-project.org. Vienna, Austria.
Read, A. F., and S. Nee. 1995. Inference from binary comparative data. J. Theor. Biol. 173:99–108.
Renner, S. S., and R. E. Ricklefs. 1995. Dioecy and its correlates in the flowering plants. Am. J. Bot. 82:596–606.
Rheindt, F. E., T. U. Grafe, and E. Abouheif. 2004. Rapidly evolving traits and the comparative method: How important is testing for phylogenetic signal? Evol. Ecol. Res. 6:377–396.
Ridley, M. 1983. The explanation of organic diversity. Oxford University Press, Oxford, UK.
Ridley, M., and A. Grafen. 1996. How to study discrete comparative methods. Pages 76–103 *in* Phylogenies and the comparative method in animal behavior (E. P. Martins, ed.). Oxford University Press, Oxford, UK.
Rokas, A., N. King, J. Finnerty, and S. B. Carroll. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. Evol. Dev. 5:346–359.
Rowley, G. 1987. Caudiciform and pachycaul succulents: Pachycauls, bottle-, barrel-, and elephant-trees and their kin; a collector's miscellany. Strawberry Press, California.
Sanderson, M. J. 1993. Reversibility in evolution: A maximum likelihood approach to character gain/loss bias in phylogenies. Evolution 47:236–252.
Sanderson, M. J., and G. Bharathan. 1993. Does cladistic information affect inferences about branching rates? Syst. Biol. 42:1–17.
Sanderson, M. J., A. Purvis, and C. Henze. 1998. Phylogenetic supertrees: Assembling the trees of life. Trends Ecol. Evol. 13:105–109.
Sattler, R. 1988. Homeosis in plants. Am. J. Bot. 75:1606–1617.
Scotland, R. W., and M. J. Sanderson. 2004. The significance of few versus many in the tree of life. Science 303:643.
Sillen-Tullberg, B. 1993. The effect of biased inclusion of taxa on the correlation between discrete characters in phylogenetic trees. Evolution 47:1182–1191.
Soltis, P. S., D. E. Soltis, and M. W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402:402–404.
Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of evolution. Mol. Biol. Evol. 15:1647–1657.
Vamosi, J. C., S. P. Otto, and S. C. H. Barrett. 2003. Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. J. Evol. Biol. 16:1006–1018.
Yule, G. U. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.P.R. Soc. Lond. B Biol. Sci. 213:21–87.
Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pages 97–166 *in* Evolving genes and proteins (V. Bryson and H. J. Vogel, eds.). Academic Press, New York.

First submitted 28 October 2005; reviews returned 22 February 2006; final acceptance 15 June 2006
Associate Editor: Mike Steel

APPENDIX 1

Model of character evolution.—Following the notation of Pagel (1994) we consider two characters, X and Y, that can take two states, 0 and 1 (see also Lewis, 2001, for a generalization of the model). A state equal to 1 represents the presence of the trait, but the coding is arbitrary. Under the continuous-time Markov process, the matrix of transition probabilities is given by $\mathbf{A} = e^{\mathbf{Q}t}$. The matrix, $\mathbf{A}$, contains the elements, $A_{ab}(t)$, that represent the probability that a character changes from state $a \in \{0, 1\}$ to the final state $b \in \{0, 1\}$ in time, t. $\mathbf{Q}$ is the matrix with the instantaneous rates of transition from one state to the next (see Pagel, 1994, for details).

When t is sufficiently long, the probability of switching from state a to state b is π_b (except in a set of cases with 0 measure), where π_b represents the equilibrium (i.e., stationary) frequency of the character state b. Let π_{X_1} and π_{X_0} represent the equilibrium frequencies of the presence and absence, respectively, of the dependent character. Let π_{ab} represent the stationary frequency of $X = a$ and $Y = b$ together.

When X and Y are independent, two 2×2 rate matrices, $\mathbf{Q}_x$ and $\mathbf{Q}_y$, describe their evolution (Pagel, 1994). However, when they are correlated, one 4×4 rate matrix is used to describe their joint evolution (Pagel, 1994). In the latter case, the two characters together can be represented by four states, 00, 10, 01, and 11, where the first digit represents the state of X and the second digit represents the state of Y. For these derivations, we consider the evolution of two characters to be correlated when the rate of gain of X is increased by the presence of Y, or the rate of loss of X is decreased in the presence of Y, but clustering of traits can occur through other processes.

Yule process of speciation.—Due to its mathematical tractability, we use the Yule process (Yule, 1924) to model phylogenesis independently from character evolution. The Yule process is a non-homogenous pure birth branching process in which the time, T, separating speciation events is exponentially distributed. At a given time, t_{total}, one of the n taxa is selected with uniform probability to speciate so that at time $t_{total} + T$ there are $n + 1$ taxa, resulting from the split of the selected taxon. We added time with density $n\lambda e^{-tn\lambda}$ to all terminal branches to represent time that passed since the last speciation event, where λ is the initial rate of speciation.

Derivation of distribution of ADs.—We derive the equation for the distribution of ADs based on the Yule model

and character evolution model. Following Figure 2, the taxon labeled "Sampled Taxon" has been sampled, and we want to find the probability that the time back to the MRA that first gave rise to one or more descendants having Y in the "Sister Lineage" falls within a tiny range of times, dt, at time t. We use the dt notation to deal with mixed random variables such as ancestral distances, A, whose distributions have both discrete and continuous portions. For a purely continuous random variable, R, $E[g(R)] = \int g(t)\mathbb{P}(R \in dt) = \int g(t) f_R(t)dt$, where f_R is the density of R and g is any nonnegative measurable function. When R is a discrete random variable, integration turns into a sum. Let A be the time from the Sampled Taxon back to the MRA. Starting from the Sampled Taxon and moving back in time, we encounter a first speciation event, a second, and so on. Let s_i be the subclade formed by the ith speciation event on the branch that does not lead to the Sampled Taxon.

Let the event, S_i, be the event that no taxa in subclade, s_i, have states 01 or 11 (i.e., no $Y = 1$). Let p_{t_i} be the probability that after time t_{total_i} event S_i occurs, where $t_{total_i} = \sum_{l=1}^{l=i} t_l$ is the total time from the ith ancestral node to the present (Fig. 2).

$$p_{t_i} := \mathbb{P}(S_i) = \sum_{k=1}^{\infty} \sum_{s=1}^{(k-1)!} \int_{b \in \Omega_B} \mathbb{P}[S_i|\tau(s,k), b]\mathbb{P}[b, \tau(s,k)]\, db \tag{1}$$

This equation is derived by conditioning the probability, $\mathbb{P}(S_i)$, on the "histories," $\tau(s,k)$, and "feasible" times, b, between nodes (Fig. 2). ("Feasible" branch times are branch times such that the total time along any path from the ancestor of the subclade to any of the k terminal taxa is t_{total_i}. The set of sets of all feasible branch times is Ω_B.) The outer summation allows for subclades of all size, whereas the second sum is across all the $(k-1)!$ evolutionary "histories" of a clade of size k (Edwards, 1970). $\mathbb{P}[S_i|\tau(s,k), b]$ is the probability of the event S_i given a particular clade topology (s), size (k), and branch lengths (b). The calculation of this probability, referred to as the likelihood, is described in detail elsewhere (e.g., Pagel, 1994). The probability of a particular set of feasible branch times and history, $\mathbb{P}[b, \tau(s,k)]$, is the product of exponentials that represent waiting times between speciation events, with attention paid to whether the waiting times separate two internal nodes or an internal node and a terminal node. Edwards (1970) and Sanderson and Bharathan (1993) provide details of this calculation.

When character evolution is very fast relative to speciation (fast characters, slow speciation; FCSS), character states may reach stationary frequencies along the branch separating a daughter node from its ancestral node. When character states reach stationarity rapidly relative to speciation rates, neither the topology of the phylogeny nor the states of the ancestral nodes influence the states of the terminal taxa, so only the number of taxa in a subclade is required to calculate the p_{t_i} of Equation (1). Under FCSS, when X and Y are independent, the p_{t_i} reduce to

$$p_{t_i} := \mathbb{P}(S_i) = \sum_{k=1}^{\infty} \mathbb{P}(S_i|N_i = k)\mathbb{P}(N_i = k)$$

where N_i is the number of extant taxa of subclade s_i. $\mathbb{P}(N_i = k)$ is the probability that k taxa descended from a single common ancestor in time t_{total_i}: $\mathbb{P}(N_i = k) = P_{1k}(t_{total_i})$. Under a Yule process, $P_{1k}(t_{total_i})$ is geometrically distributed with parameter $e^{\lambda t_{total_i}}$ (Hoel et al., 1987: 99), where λ is the rate of speciation. The probability of S_i in a subclade with k taxa equals $(\pi_{00} + \pi_{10})^k$ under FCSS conditions. When X and Y are independent, and character states evolved under FCSS contitions, $\mathbb{P}(S_i|N_i = k) = (\pi_{00} + \pi_{10})^k = (\pi_{X_0}\pi_{Y_0} + \pi_{X_1}\pi_{Y_0})^k = \pi_{Y_0}^k(\pi_{X_1} + \pi_{X_0})^k = \pi_{Y_0}^k$ because $\pi_{X_1} + \pi_{X_0} = 1$. Under FCSS, Equation (1) simplifies to

$$p_{t_i} = \sum_{k=1}^{\infty} e^{\lambda t_{total_i}} \left(1 - e^{\lambda t_{total_i}}\right)^k \pi_{Y_0}^k \tag{1a}$$

When states of taxa depend on the structure of the phylogeny (i.e., not FCSS) and X and Y are independent, the probability of observing all taxa with $Y = 0$ and all taxa with either $X = 1$ or $X = 0$ is the probability that all taxa have $Y = 0$ times the probability that all taxa have either $X = 1$ or $X = 0$. Let the event that all taxa have $Y = 0$ be Y_a^0 and the event that all taxa have $X = 1$ or $X = 0$ be X_a^{01}. When X and Y are independent, Equation (1) is therefore:

$$p_{t_i} = \sum_{k=1}^{\infty} \sum_{s=1}^{(k-1)!} \int_{b \in B} \mathbb{P}\left[X_a^{01}|\tau(s,k), b\right]\mathbb{P}\left[Y_a^0|\tau(s,k), b\right] \times \mathbb{P}[b, \tau(s,k)]\, db$$

$\mathbb{P}[X_a^{01}|\tau(s,k), b]$ equals 1 because this calculation exhausts all combinations of states of character X across all terminal and ancestral taxa. Equation (1) simplifies to

$$p_{t_i} = \sum_{k=1}^{\infty} \sum_{s=1}^{(k-1)!} \int_{b \in B} \mathbb{P}\left[Y_a^0|\tau(s,k), b\right]\mathbb{P}[b, \tau(s,k)]\, db \tag{1b}$$

We now consider the distribution of ADs. This distribution has both a continuous portion at $t > 0$ and a discrete portion (atom) at $t = 0$. The p_{t_i} terms do not take into account ADs equal to 0, i.e., $\mathbb{P}(A = 0)$. We describe that calculation later. From independence of lineage evolution under the Yule process, the calculation for ADs greater than 0 is the product of the probability that the AD does not equal 0 (i.e., $1 - \mathbb{P}(A = 0)$), times the product of all the p_{t_i} terms, times the probability that the Sister Lineage (see Fig. 2) has one or more taxa with $Y = 1$ (i.e.,

$1 - p_{t_a}$). All possible numbers of ancestors, a, must be considered, as well as all times t_i such that the t_i sum to t

$$\mathbb{P}(A \in dt) = [1 - \mathbb{P}(A = 0)] \sum_{a=1}^{\infty} \int_{\substack{t_1,\ldots,t_{a-1} \geq 0, \\ t_a = t-(t_1+\cdots+t_{a-1}) \geq 0}} \times \left[\prod_{i=1}^{a-1} \lambda \exp(-\lambda t_i) p_{t_i} dt_i \right] \lambda \exp(-\lambda t_a) \times (1 - p_{t_a}) dt. \quad (2)$$

To show our objective that the distribution of ADs is the same for the target and nontarget samples when characters are independent, first consider the probability that $A = 0$ given that the sampled taxon is in the target sample (substitute $X = 0$ for the nontarget sample for an analogous result), $\mathbb{P}(A = 0|X = 1)$: $\mathbb{P}(A = 0|X = 1) = \mathbb{P}(A = 0, X = 1)/\mathbb{P}(X = 1) = \mathbb{P}(X = 1, Y = 1)/\mathbb{P}(X = 1) = \mathbb{P}(X = 1)\mathbb{P}(Y = 1)/\mathbb{P}(X = 1) = \mathbb{P}(Y = 1)$. The second to last step follows from the independence of X and Y. A similar result follows for positive ADs. Because the event $A \in dt$ is completely determined by the values of Y for the taxa in the phylogeny (by substitution of 1a or 1b into 2) when evolution of Y is independent of X, $A \in dt$ is independent of X. As a corollary, under FCSS conditions with independent characters, the frequency of sampled taxa with AD equal to 0 when characters are independent is expected to be π_{Y_1}.