Syst. Biol. 60(4):410–419, 2011

DOI:10.1093/sysbio/syr007
Advance Access publication on March 4, 2011

Testing for Temporal Variation in Diversification Rates When Sampling is Incomplete and Nonrandom

CHAD D. BROCK[1,2,*], LUKE J. HARMON[3], AND MICHAEL E. ALFARO[4]

[1]*Department of Zoology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236;*
[2]*Department of Ecology, Evolution, and Behavior, Section of Integrative Biology, One University Station C0930, University of Texas at Austin, Austin, TX 78712;*
[3]*Department of Biological Sciences, University of Idaho, Moscow, ID 83844-3051; and*
[4]*Department of Ecology and Evolutionary Biology, University of California at Los Angeles, Los Angeles, CA 90095-1606;*
[*]*Correspondence to be sent to: Section of Integrative Biology, One University Station C0930, University of Texas at Austin, Austin, TX 78712; E-mail:maximaul@msn.com.*

Received 28 August 2009; reviews returned 28 October 2009; accepted 12 November 2010
Associate Editor: Todd Oakley

Abstract.—A common pattern found in phylogeny-based empirical studies of diversification is a decrease in the rate of lineage accumulation toward the present. This early-burst pattern of cladogenesis is often interpreted as a signal of adaptive radiation or density-dependent processes of diversification. However, incomplete taxonomic sampling is also known to artifactually produce patterns of rapid initial diversification. The Monte Carlo constant rates (MCCR) test, based upon Pybus and Harvey's gamma (γ)-statistic, is commonly used to accommodate incomplete sampling, but this test assumes that missing taxa have been randomly pruned from the phylogeny. Here we use simulations to show that preferentially sampling disparate lineages within a clade can produce severely inflated type-I error rates of the MCCR test, especially when taxon sampling drops below 75%. We first propose two corrections for the standard MCCR test, the proportionally deeper splits that assumes missing taxa are more likely to be recently diverged, and the deepest splits only MCCR that assumes that all missing taxa are the youngest lineages in the clade, and assess their statistical properties. We then extend these two tests into a generalized form that allows the degree of nonrandom sampling (NRS) to be controlled by a scaling parameter, α. This generalized test is then applied to two recent studies. This new test allows systematists to account for nonrandom taxonomic sampling when assessing temporal patterns of lineage diversification in empirical trees. Given the dramatic affect NRS can have on the behavior of the MCCR test, we argue that evaluating the sensitivity of this test to NRS should become the norm when investigating patterns of cladogenesis in incompletely sampled phylogenies. [γ-Statistic; adaptive radiations; comparative method; diversification rates; phylogenetics.]

The continued availability of molecular data coupled with more sophisticated methods for reconstructing time-calibrated molecular phylogenies has led to an increase in the number of studies investigating patterns of lineage diversification (e.g., Weir and Schluter 2004; Kozak et al. 2006; Rabosky 2006; Alfaro, Karns, et al. 2007; Alfaro, Santini, et al. 2007; Wiens 2007; Rabosky and Lovette 2008a; Phillimore and Price 2008; Alfaro et al. 2009). A pervasive pattern across many taxonomic groups and time scales is that rates of cladogenesis appear to be elevated in the early history of clades (Pybus and Harvey 2000; Ruber and Zardoya 2005; Weir 2006; Alfaro, Karns, et al. 2007; Price 2007; Mooers et al. 2007; Phillimore and Price 2008; Rabosky and Lovette 2008a, 2008b; see also McPeek 2008). This is often interpreted as evidence for ecologically dependent modes of diversification, including adaptive radiations (sensu Schluter 2000). Microevolutionary models of speciation have also shown that ecological diversification can lead to early-burst patterns of cladogenesis (Gavrilets and Vose 2005; see also McPeek 2008), providing further evidence for the linking of these patterns and ecologically dependent diversification.

Evidence for elevated initial diversification typically comes from the constant rates (CR) test of Pybus and Harvey (2000), which assesses the observed average distance of nodes to the midpoint of a tree to that expected under a model where clades have diversified under a CR using a test statistic, γ. Pybus and Harvey recognized that incompletely sampled phylogenies produced a bias in the CR test toward significantly negative γ-values (indicating rapid early diversification) and proposed a Monte Carlo procedure for generating a corrected null distribution for γ given the number of missing taxa in the tree. However, this widely employed procedure (the Monte Carlo constant rates [MCCR] test) assumes that the missing taxa have been randomly pruned from the tree.

The degree to which the assumption of random sampling is violated in phylogenetic studies is unknown, but it is potentially nontrivial. Systematists studying higher-level relationships will typically sample a small number of exemplar tips to capture major radiations, a practice known as overdispersed sampling. Furthermore, cryptic speciation (Coyne and Orr 2004; Bickford et al. 2006) may cause recently diverged tips to go unrecognized and thus unsampled. Both of these practices differentially sample older splits in the tree and should bias the distribution of node ages for undersampled trees toward the root. The effect of cryptic species has been addressed in at least two ways. The first involves truncating the tree by a given amount of time from the tip (e.g., 2 myr; Avise and Walker 1998; Weir 2006; Phillimore and Price 2008), thus leaving out any recent splits, realized or not, from the final analysis. The second, instead looks at the sensitivity of the MCCR results to the number of extant species, that is, evaluating how many extant species there would need to be for the empirical γ to no longer be significant (e.g., Kozak et al. 2006; Rabosky and Lovette 2008b). The practice of

sampling exemplars, however, has been largely ignored in empirical studies (but see Moyle et al. 2009), and the overall influence of incomplete overdispersed sampling on the performance of the MCCR test is currently unknown.

Here we take a simulation approach to assess the behavior of the MCCR test when the assumption of random sampling is violated. We consider two degrees of violation of this assumption: overdispersed sampling where more deeply diverging lineages are preferentially sampled and "worst-case" sampling, where all of the missing taxa are the youngest splits within a tree. We find that there is a dramatic increase in type-I error under both modes of nonrandom sampling (NRS) and first propose two modifications of the standard MCCR test to alleviate this artifact. We then extend these tests into a generalized NRS MCCR test that incorporates a scaling parameter, α, to control the degree to which sampling is overdispersed. This test is used to reanalyze two empirical studies and assess the sensitivity of the MCCR results to mode of sampling. We find that NRS can influence evolutionary inference and recommend assessing the sensitivity of the results of the MCCR test to mode of sampling in cases where taxonomic sampling falls below 75%.

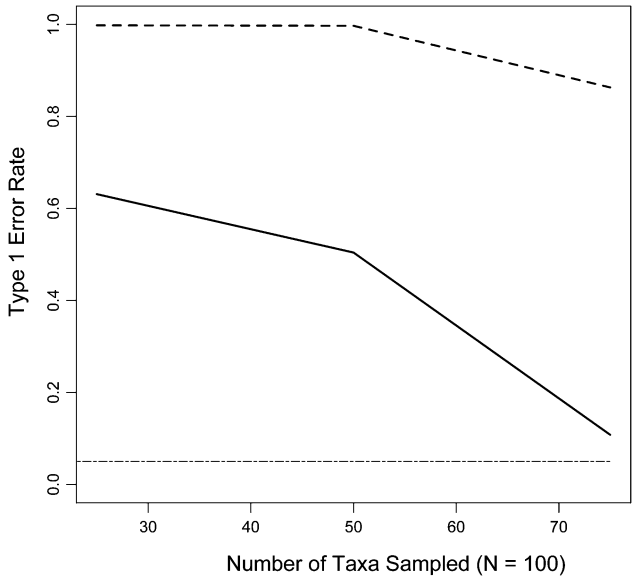

FIGURE 1. Type-I error of the standard MCCR test under DSO (dashed line) and PDS (solid line) modes of sampling for 25%, 50%, and 75% of extant taxa ($n = 100$) sampled. The hatched line in the lower portion of the graph represents the standard cut-off for acceptable type-I error (0.05).

METHODS

The CR and MCCR Tests

The most common method employed when investigating temporal patterns of cladogenesis has been the CR test of Pybus and Harvey (2000). This test evaluates the fit of a CR pure birth (PB) process to the branching events in a tree, using a statistic, γ [1] (Pybus and Harvey 2000; see also Zink and Slowinski 1995):

$$\gamma = \frac{\left(\frac{1}{n-2}\sum_{i=2}^{n-1}\left(\sum_{k=2}^{i} kg_k\right)\right) - \left(\frac{T}{2}\right)}{T\sqrt{\frac{1}{12(n-2)}}}, T = \left(\sum_{j=2}^{n} jg_j\right) \quad (1)$$

where n is the number of lineages in the tree, and g_2, g_3,…g_n are the internode distances. This statistic measures the standardized difference between the average sum of branch lengths between each internal node and the root and the midpoint of the tree (Pybus and Harvey 2000; see also Cox and Lewis 1966; Mooers et al. 2007) (Fig. 1). Trees with a disproportionate number of nodes toward the base of the tree (i.e., more rootward from the tree midpoint) will have a low average distance to the root, and thus result in negative γ-values. Alternatively, trees with more tipwardly distributed nodes yield positive γ-values. For a completely sampled PB process, gamma has a standard normal distribution with a mean of zero. Thus, for completely sampled phylogenies, a gamma-statistic less than -1.645 is considered significant evidence for rejecting a CR PB process of diversification in favor of declining rates of cladogenesis through time (Pybus and Harvey 2000). Positive values of γ can result from either extinction or increases in diversification rate through time and are generally ignored given the problems associated with estimating extinction rates from contemporary data (Pybus and Harvey 2000; but see Rabosky and Lovette 2008a).

As recognized by Pybus and Harvey, empirical phylogenies often suffer from incomplete sampling, which could lead to an apparent slowdown in cladogenesis through time (and a negative γ-value). Simulations confirm that as the proportion of sampled taxa in a tree decreases, the type-I error of the CR test increases (see figure 4 in Pybus and Harvey 2000). To account for incomplete sampling, Pybus and Harvey (2000) suggested a parametric test known as the MCCR test. This test constructs the null distribution of γ as follows:

1. Simulate N PB trees to a clade size equal to that of all known extant species in the group of interest.
2. Randomly prune taxa from each simulated tree until the number of sampled taxa is equal to that sampled for the empirical tree.
3. Calculate the γ-statistic for each pruned tree.
4. Calculate the MCCR corrected P value as the proportion of the N γ-values that are smaller than the test statistic calculated from the empirical tree.

When its assumptions are met, the MCCR test, by definition, has an acceptable type-I error. Furthermore, the test is robust to extinction, which increases γ by pruning proportionally more ancient lineages, leaving an excess of younger branching events (the so-called "pull-of-the-present"; Pybus and Harvey 2000; see also Weir 2006 and Rabosky and Lovette 2008a). However, the CR and MCCR tests make assumptions that may be violated in particular data sets, including that 1) the phylogeny is known without error; 2) node heights accurately represent relative branching times; 3) rates of cladogenesis are

equal at any one time slice across all lineages (the equal rates Markov model; ERM); and 4) taxa are sampled randomly for incompletely sampled trees.

The first three of these assumptions have been considered in the literature. Phylogenetic uncertainty is unlikely to have a strong impact on the CR test, though the relative branching times may (Pybus and Harvey 2000). Uncertainty in both topology and branching times can be accommodated by repeating analyses over a sample of credible phylogenies (such as those within 2 log-likelihood units of a maximum-likelihood estimated phylogeny or a sample from the posterior of a Bayesian analysis; see Rabosky and Lovette 2008b; see also Ruber and Zardoya 2005 and Revell et al. 2005 for potential biases stemming from incorrect models of molecular evolution and algorithms for reconstructing chronograms, respectively). The fit of an ERM model can be assessed using a number of different topological (Kirkpatrick and Slatkin 1993; Moore et al. 2004) or temporal-based approaches (Nee et al. 1994; Rabosky et al. 2007 and a tree can be "linearized" by pruning out the groups that show significant rate heterogeneity (e.g., Ruber and Zardoya 2005). We used simulations to investigate violations of the fourth assumption.

MCCR Test and NRS

To investigate the type-I error under conditions of NRS, we simulated thousand 100-taxon PB trees (birth rate = 1, extinction rate = 0) using the CR birth-death algorithm from the Geiger package for R (Harmon et al. 2008). Each tree was then sampled for 25, 50, and 75 taxa under two biased sampling methods:

1. Proportionally deeper splits (PDS): We prune the number of missing taxa, *m*, selecting taxa with probability inversely proportional to their tip branch length. This method will prune a higher proportion of taxa with shorter tip branches leaving a disproportionate amount of tips that attach to the rest of the tree at relatively deeper nodes. This sampling method is expected to elevate type-I error of the MCCR test though to a lesser extent than the deepest splits only (DSO) approach (see below).
2. DSO: For a given level of sampling, *s*, we pruned all but the *s*-1 bottom-most nodes (including the root) of the tree. This should dramatically inflate type-1 error of the MCCR test as one is sampling only the MOST rootward nodes of the tree.

Type-I error under both the PDS and DSO sampling approaches for all levels of incomplete sampling were calculated as the proportion of trees that show a significantly negative γ, indicative of a slowdown in net cladogenesis toward the tips.

Accounting for NRS in the MCCR Test I: the PDS and DSO Tests

Due to the drastic elevation in the type-I error of the MCCR test under both modes of NRS (see Results section), we developed a correction for the MCCR test designed to help alleviate bias due to the overdispersed sampling of species in a phylogeny. To do this, we incorporated the two methods of NRS described above into the Monte Carlo simulation of the null distribution of γ. The procedures for the corrected MCCR tests are as follows:

1. simulate *N* PB trees,
2. for each tree, prune the *m* missing taxa nonrandomly according to either the DSO or PDS approach,
3. calculate γ for each pruned tree
4. calculate the MCCR corrected *P* value as the proportion of the *N* γ-values that are smaller than the test statistic calculated from the empirical tree.

We investigated the type-I error of these methods over three different levels of sampling (25%, 50%, and 75%) and assessed power for three different magnitudes of rate decrease through time (early rate = 2, 3, or 4*x* the later rate).

To calculate type-I error rates for each method we simulated one thousand 100-taxon trees using the same CR birth–death algorithm as above. Each tree was subsequently pruned to the sampled number of species using the PDS and DSO NRS schemes. We calculated γ for each tree and assessed statistical significance using both PDS and DSO modified MCCR tests. As described above, type-I error was calculated as the proportion of simulated trees found to have a significantly negative γ-value.

To assess power under each combination of sampling level and degree of rate shift, we used a birth–death tree simulation algorithm written by C.D.B. for R that allows the birth rate to shift at some defined point in the tree. For each combination of parameter values, we simulated one thousand 100-taxon trees under a specific elevated early rate of diversification until clade size was 50, at which point the rate was decreased to 1 until a clade size of 100 was reached. Each tree was then pruned to the respective level of sampling using both the DSO and PDS approach and γ was calculated. Each of the corrected MCCR tests were used to produce the null distribution of the test statistic and assess the significance of γ for each simulated tree.

Accounting for NRS in the MCCR Test II: αNRS MCCR Test

Sampling mode is often unknown in empirical studies. In these situations, it may be difficult to know which method to employ and rather than employing these two alternative tests (or arbitrarily choosing between them), we developed another test that generalizes the approach outlined above. This test, the αNRS test, employs a scaling parameter that controls the degree to which sampling is biased toward deeper nodes. Briefly, the branch length for each tip taxon is raised to a parameter, α, and the reciprocal of this value is used as the probability that the tip taxa in question is pruned out of the tree. When $\alpha = 0$, this probability becomes one for all taxa and sampling is completely random, as in the

standard MCCR test. When $\alpha = 1$, sampling is proportional to the reciprocal of the tip branch length, which is identical to the PDS sampling mode described above. As α increases, sampling becomes more biased toward the root of the tree until essentially only the oldest nodes are retained, consistent (though not identical due to the probabilistic nature) with the DSO mode of sampling.

Empirical Analyses

Most empirical phylogenies investigated to date have been found to have negative γ-values (McPeek 2008; Mooers et al. 2008). To investigate the potential effect that NRS has on this trend, we reanalyzed data sets from two previous studies, one on homalopsine snakes (Alfaro, Karns, et al. 2007) and a recent meta-analysis on diversification in birds (Phillimore and Price 2008). We employed the αNRS test to assess the sensitivity of each study's MCCR results to sampling mode.

Snakes.—Alfaro, Karns, et al. (2007) constructed a chronogram for 21 of the 34 extant species (61.7%) of the Southeast Asian colubroid family homalopsidae. Alfaro et al. employed the standard MCCR test to evaluate the fit of a CR PB process to the branching events corrected for the number of missing taxa (MCCR-corrected $P = 0.003$). This significant negative γ provided evidence for an early burst of cladogenesis during the Early Miocene (~22 Ma). However, sampling was nonrandom, as the intergeneric relationships were of primary interest, and thus the significant negative γ might have been an artifact of overdispersed sampling.

Birds.—Phillimore and Price (2008) collected phylogenies for 45 avian taxa, with sampling ranging from 67.6% to 100% (mean = 90.8%). Table 1 from their paper was used to compile data for the number of extant species, the number of missing species and γ-values

TABLE 1. Data set from Phillimore and Price (2008) with the original and PDS-corrected γ- and P-values

Data set	Extant spp.	Missing spp.	UC γ-value	PDS γ-value	UC P value	PDS P value
Wrens	74	24	−3.628	−1.534	<0.001	>0.05
Phylloscopus and *Seicercus*	53	11	−2.991	−1.76	<0.01	<0.05
Anthus	44	9	−2.855	−1.72	<0.01	<0.05
Catharus	11	0	−2.83	−2.83	<0.01	<0.01
Grackles and allies	40	4	−2.828	−2.407	<0.01	<0.01
Estrildidae	140	1	−2.743	−2.796	<0.01	<0.01
Parus	54	14	−2.622	−1.038	<0.01	>0.05
Tangara	49	7	−2.465	−1.626	<0.01	>0.05
Turdus & allies	69	10	−2.278	−1.12	<0.05	>0.05
Dendroica, Parula, et al.	45	5	−2.224	−1.665	<0.05	<0.05
Amazona	31	3	−1.856	−1.511	<0.05	>0.05
Tringa	13	1	−1.85	−2.03	<0.05	<0.05
Swallows	29	0	−1.776	−1.776	<0.05	<0.05
Caciques and oropendolas	19	2	−1.765	−2.035	<0.05	<0.05
Ficedula	25	0	−1.673	−1.673	<0.05	<0.05
Hemispingus	12	2	−1.635	−1.535	>0.05	>0.05
Acanthiza	12	0	−1.543	−1.543	>0.05	>0.05
Anas	47	6	−1.377	−1.49	>0.05	>0.05
Toxostoma	10	0	−1.358	−1.358	>0.05	>0.05
Thamnophilus	25	1	−1.282	−1.332	>0.05	>0.05
Geositta	11	0	−1.271	−1.271	>0.05	>0.05
Storks	19	3	−1.254	−1.201	>0.05	>0.05
Meliphaga	13	1	−1.179	−0.876	>0.05	>0.05
Trogons	39	10	−0.91	−1.39	>0.05	>0.05
Sylvia	24	2	−0.741	0.38	>0.05	>0.05
Alcinae	22	0	−0.705	−0.705	>0.05	>0.05
Picoides and *Venillornis*	23	2	−0.645	−0.53	>0.05	>0.05
Empidonax	15	0	−0.642	−0.642	>0.05	>0.05
Icterus	25	0	−0.543	−0.543	>0.05	>0.05
Crax	14	0	−0.54	−0.54	>0.05	>0.05
Ramphastos	11	3	−0.483	−0.457	>0.05	>0.05
Aegotheles	8	1	−0.434	−0.229	>0.05	>0.05
Penguins	17	0	−0.16	−0.16	>0.05	>0.05
Pteroglossus	13	1	−0.125	−0.011	>0.05	>0.05
Larus	49	1	0.073	−0.33	>0.05	>0.05
Grouse, turkeys, et al.	57	4	0.126	0.051	>0.05	>0.05
Myioborus	12	0	0.165	0.598	>0.05	>0.05
Alectoris	7	0	0.287	0.316	>0.05	>0.05
Cinclodes	12	0	0.465	0.32	>0.05	>0.05
Cranes	15	0	0.671	0.671	>0.05	>0.05
Albatross	14	0	0.866	0.866	>0.05	>0.05
Sterna	44	10	1.365	0.994	>0.05	>0.05
Puffinus	21	3	1.49	2.77	>0.05	>0.05
Tauraco	14	1	1.657	1.916	>0.05	>0.05
Myiarchus	22	3	1.854	1.509	>0.05	>0.05

Bird data set from Phillimore and Price (2008). UC γ-value and P value are from the original study and do not account for NRS. The PDS values for γ- and P value are corrected for overdispersed sampling using the PDS mode and the approach described in Harmon et al. (2003).

estimated after correcting for incomplete (but random) sampling. These values were then used to calculate a corrected γ-value for each clade using the αNRS method of NRS over varying α-values (see below).

Correcting γ for NRS.—In order to account for NRS when calculating γ we adjusted the values taken from the two studies using the approach of Harmon et al. (2003). For each data set, 2500 PB trees were simulated using the CR birth–death algorithm in the GEIGER package for R (**?**) and then pruned to the appropriate level of sampling using the αNRS test described above for some value of α. For each pruned tree, γ was calculated and the median value from all simulations was subtracted from the empirical γ-values reported in the studies. This value was then divided by the standard deviation of the simulated γ-values to produce the corrected value for each data set. The sensitivity of the MCCR results to NRS was assessed by identifying the value of α that caused the empirical γ-value to lose significance (homalopsid data set), or the mean γ to no longer be significantly less than zero (bird data set).

RESULTS

MCCR Test: Type-I Error

The standard MCCR test performed poorly under all investigated modes and degrees of NRS. Type-I error was significantly elevated (>0.05) over all combinations of sampling mode and degree (Fig. 1). When sampling was only 25% complete, the MCCR test found that 100% of the simulated trees showed significant evidence for rate declines under the DSO mode of sampling (i.e., a type-I error rate of 1.0!). Even under the less extreme PDS mode of sampling, the type-I error of the MCCR test was 0.63. When sampling reached 75% complete, the MCCR test performed much better under the PDS mode of sampling (type-I error = 0.086) though under the DSO mode, type-I error was still substantially elevated (0.87).

DSO and PDS MCCR Tests

The PDS test performed poorly when the empirical tree was sampled using the extreme DSO method, with type-I error significantly inflated (Fig. 2a), though substantially less so than the standard MCCR test (e.g., 0.36 vs. 0.87 at 75% sampling). When the PDS sampling method was used, type-I error was acceptable (≤ 0.05) as was the power, which increased with sampling (Fig. 2b).

The DSO test was by far the most conservative of the two tests developed. When the empirical trees were nonrandomly sampled using the PDS method for 25% of taxa, only ~1% of the trees simulated under a model of 2-fold decline in rate were found significant. Similarly, only ~12% and 37% of the trees for the 3- and 4-fold decline simulations were found significant (Fig. 3b).

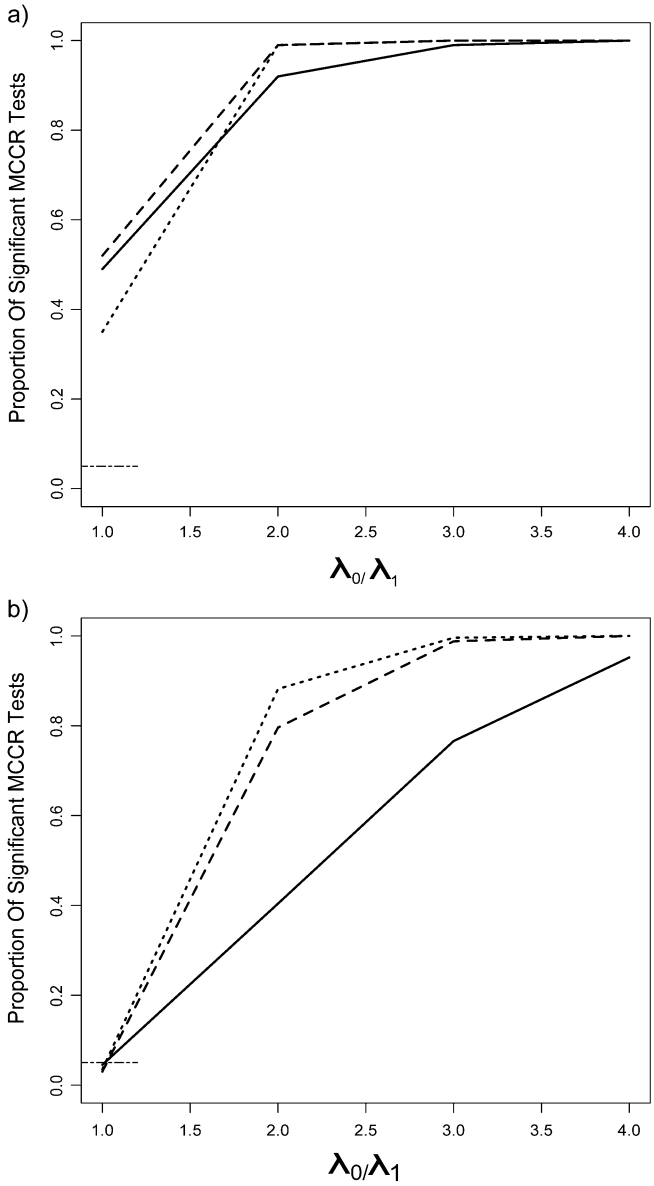

FIGURE 2. Type-I error ($\lambda_0/\lambda_1 = 1$) and power ($\lambda_0/\lambda_1 > 1$) of the PDS test under DSO (a) and PDS (b) modes of sampling for 25% (solid line), 50% (dashed line), and 75% (dotted line) of extant taxa ($n = 100$) sampled. The hatched line in the lower left corner represents the standard cutoff for acceptable type-I error (0.05).

Increased taxon sampling significantly improved the power of the DSO test, though overall it remained conservative when rate changes were small (i.e., $\lambda_0/\lambda_1 = 2$). When the DSO sampling method was used, the DSO test had acceptable type-I error (≤ 0.05) and reasonable power, especially with more complete sampling (Fig. 3a).

Empirical Examples

The PDS corrected P value (i.e., $\alpha = 1$) for the homalopsine γ approached significance (adjusted $\gamma = -1.35$,

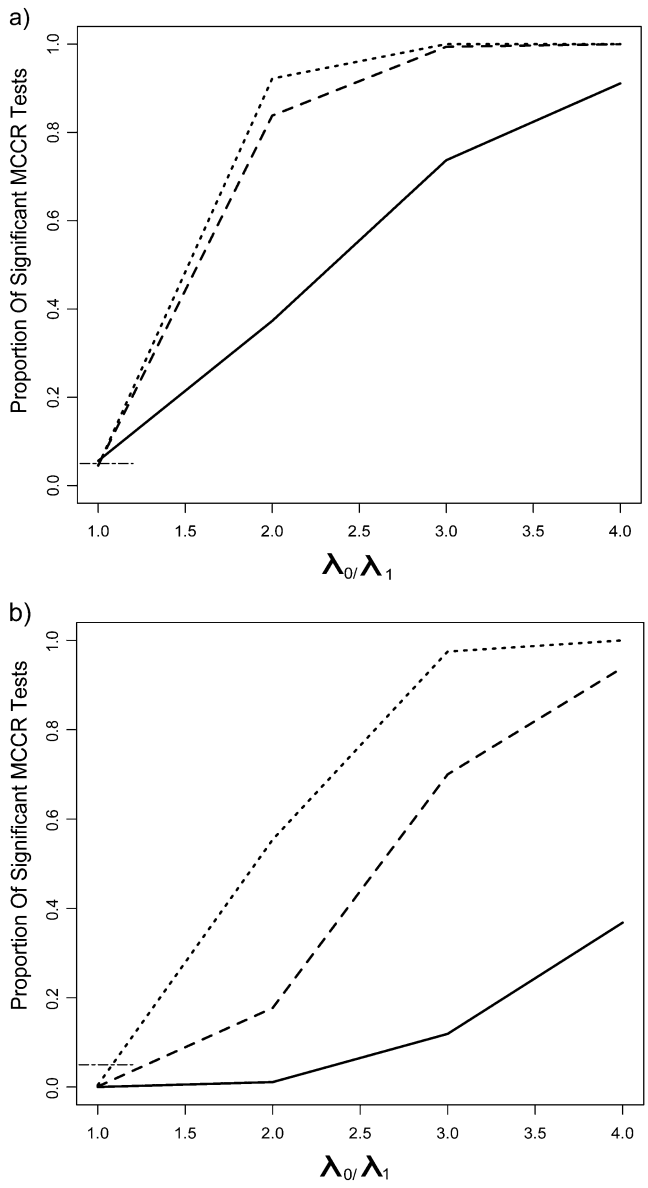

FIGURE 3. Type-I error ($\lambda_0/\lambda_1 = 1$) and power ($\lambda_0/\lambda_1 > 1$) of the DSO test under DSO (a) and PDS (b) modes of sampling for 25% (solid line), 50% (dashed line), and 75% (dotted line) of extant taxa ($n = 100$) sampled. The hatched line in the lower left corner represents the standard cut-off for acceptable type-I error (0.05).

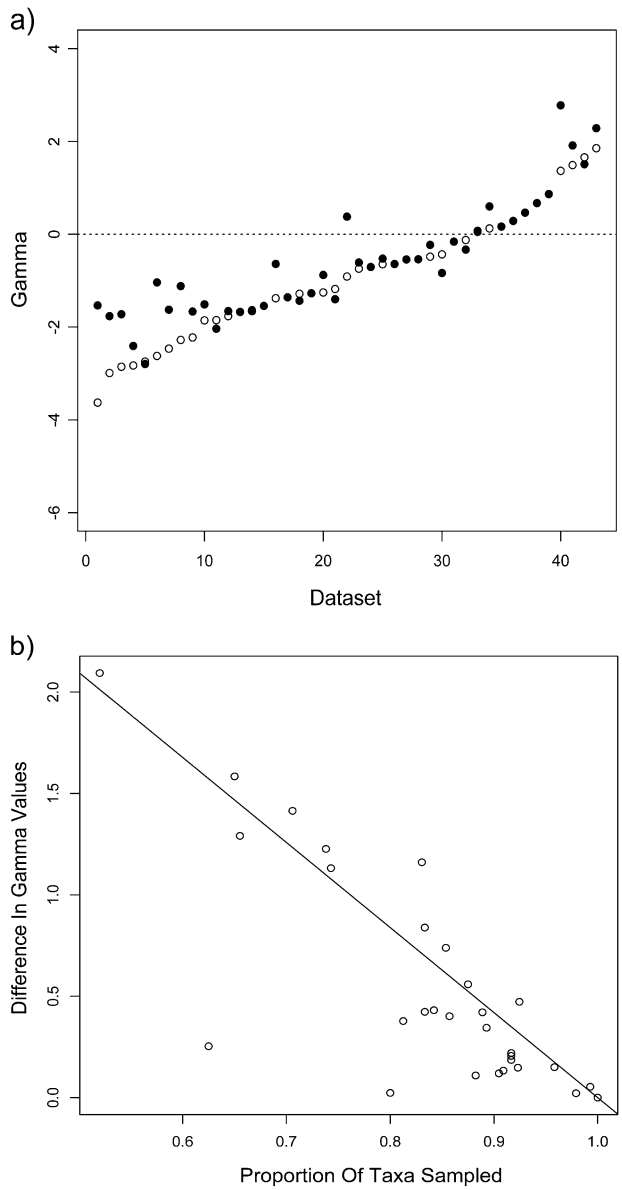

FIGURE 4. a) γ-Values before (open circle) and after (closed circles) PDS correction for each data set given in Phillimore and Price (2008). The order of the data sets (from left to right) matches that of table 1 in Phillimore and Price (2008) and Table 1 in this article (see below). b) Regression of the difference between the PDS corrected and uncorrected γ-values and the proportion of taxa sampled in the respective study ($r = 0.83$; $P < 0.001$).

$P = 0.09$), but was much larger than that of the standard MCCR test ($\gamma = -2.75$, $P = 0.003$). Tests were statistically significant for all $\alpha \leq 0.61$.

The PDS-adjusted γ-values (mean $= -0.58$) for the bird data set did not differ significantly from that of the original values (mean $= -0.98$) from Phillimore and Price (paired t-test, $t = -9.23$; $P << 0.001$; Fig. 4a). As with the original values, the PDS corrected γ-values also differed significantly from zero ($t = -3.137$; $P = 0.003$), consistent with trend toward negative γ across the bird clades investigated. This trend was consistent for $\alpha \leq 10$, at which sampling began to mimic the conservative DSO mode outlined above. The absolute value of the difference between the empirical and corrected γ-values was negatively correlated with the proportion of taxa sampled ($r = 0.83$; $P < 0.001$), and for 75% sampling or better the difference seldom exceeded 1.0 (Fig. 4b).

DISCUSSION

Our study demonstrates that overdispersed sampling has a significant impact on the type-I error of

the standard MCCR test (Fig. 1), and this suggests that some of the negative γ-values reported in empirical studies could potentially reflect this bias. Our corrected MCCR test provides a means of assessing the sensitivity of the results of an MCCR test to differing degrees of overdispersed sampling.

The Standard MCCR Test

Pybus and Harvey (2000) introduced the MCCR test as a way of accommodating incomplete taxon sampling when investigating temporal patterns of diversification using their γ-statistic. As noted by the authors, this test made a number of assumptions, including that the taxa included in the phylogeny were randomly sampled from extant members of the clade. Although intuitively one would expect overdispersed sampling to bias γ toward negative values, the degree to which this artifact inflates type-I error in the standard MCCR test is striking (Fig. 1). When sampling is 25% complete, the standard MCCR test has a type-I error >0.6 even under the less severe PDS mode of sampling. Even when sampling is 75% complete (just above the cutoff level for inclusion in the Phillimore and Price metanalysis), type-I error is elevated in the standard MCCR test under both modes of NRS (Fig. 1). Given this degree of type-I error inflation and the high proportion of empirical data sets that are incompletely sampled, we feel the standard MCCR test should be used with caution. Assessing the sensitivity of the results to overdispersed sampling is advisable.

Accommodating Overdispersed Sampling: the αNRS Test

The poor performance of the standard MCCR test when sampling is overdispersed (see above) led us to develop corrections aimed at accommodating overdispersed sampling when evaluating the significance of γ. We initially developed two modified MCCR tests, the DSO and PDS-MCCR test and then extended these to a more generalized model (the αNRS test), where sampling is controlled by a scaling parameter, α. When the sampling schemes matched that of the employed test (e.g., DSO and the PDS-MCCR test), type-I error was appropriate ($\sim$0.05) and power increased with the proportion of sampled taxa, as seen in the standard MCCR test (Pybus and Harvey 2000). When the more severe method of overdispersed sampling was employed in the empirical trees (DSO) and then evaluated using the PDS test, type-I error was inflated (Fig. 2a), though to a much smaller degree than when the standard MCCR test is used (Fig. 1). When the PDS sampling scheme was investigated using the more conservative DSO test, type-I error was appropriate (<0.05), but power was extremely low (Fig. 3b). Given the difference in power between these two tests, it is important to consider which is most appropriate to implement given that the degree of sampling overdispersion is unknown in most studies. For studies designed to investigate higher-level relationships (e.g., interordinal relationships within Mammalia), there can be a tendency to capture mostly basal splits in the tree and thus the DSO test may be a more appropriate (and conservative) choice than either the PDS or the standard MCCR test. Although it is certainly possible (and even probable) that these extreme sampling schemes may not capture only the basal most splits in the tree (especially if there is heterogeneity in cladogenetic rate between subclades), they will tend to capture mostly basal splits, making the DSO test the conservative choice when assessing the significance of γ in these scenarios. It is worth noting that a recent study (Moyle et al. 2009) independently developed the DSO test to assess the sensitivity of their results to overdispersed sampling, lending support to the necessity of this extreme test in some circumstances.

In most studies, sampling mode probably falls somewhere in between the DSO and random sampling schemes, rather than being predominantly one extreme or the other. For instance, the homalopsine data set was designed to capture the main splits between genera within the family. However, it also included some closely related lineages (*Cerberus "rhynchops"* and *C. microlepis*) as well as a number of recently described "cryptic" species with relatively shallow nodes. For a study such as this it may be difficult to assess the degree of NRS and a better approach might be a sensitivity analysis.

Our generalized αNRS test allows researchers to evaluate the sensitivity of evidence for significant rate slowdown to sampling mode, which can be altered incrementally rather than being constrained to the three general approaches outlined above (e.g., random, PDS, and DSO). The parameter value for which significance disappears can then be used in simulations to assess how biased this mode of sampling is given the clade size and the percent sampled. For instance, the probability of capturing deep versus shallow nodes or the probability of capturing a specific node giving its relative placement in the tree could be assessed, as in Figures 5a and 5b. In the case of a clade with 100 extant taxa for which 50 are sampled, a value of γ that remains significant even at $\alpha = 6$ provide strong support for an early burst of cladogenesis, as this degree of overdispersion is essentially maximally biased (i.e., only the basal most nodes are usually captured under this sampling scheme).

Ideally, researchers would benefit from a method that explicitly assesses the degree of NRS, perhaps by estimating the value of the scaling parameter α. Unfortunately, such a method has yet to be developed and the feasibility of assessing the degree of sampling overdispersion based solely on an incompletely sampled molecular tree remains uncertain. Although using node depth as a metric to evaluate lineage inclusion does not precisely mimic the method by which taxa are sampled in most (if any) studies, it seems a reasonable and easily implemented approximation to help account for the obvious bias toward negative γ-values due to NRS.

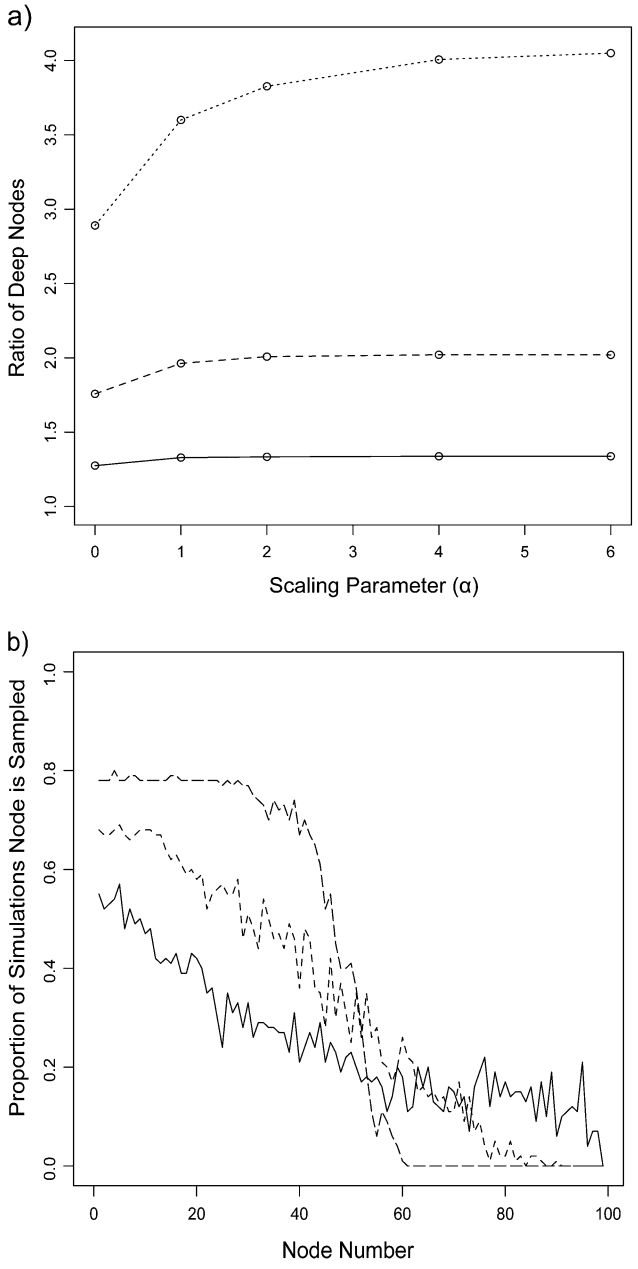

FIGURE 5. a) The ratio of deep nodes (e.g., those in the first half of the tree) captured in the sampled tree to those in the completely sampled tree for different levels (25, 50, and 75% complete) and modes ($\alpha = 0, 1, 2, 4, 6$) of sampling. b) The proportion of simulations for which a given node was captured after sampling for α values of 0 (solid line), 1 (dotted line), and 6 (dashed line). Nodes are ordered from the deepest (node 1) to the most shallow (node 99). For both a) and b) 1000 100-taxon trees were simulated and for b) trees were sampled at 50%.

Meta-analyses

The mean γ-value for the bird (-0.98, sensu Phillimore and Price 2008) was significantly less than zero, indicative of a trend toward negative γ-values for these clades (see Phillimore and Price 2008 for discussion of this trend). Using the PDS test to accommodate for overdispersed sampling had no effect on the patterns seen in the bird data set, as the corrected mean γ-value was still significantly less than zero ($P < 0.001$). This significance remained for all α values ≤ 10, a level of biased sampling that approximates the most conservative DSO method outlined above. Simulations using the mean clade size ($\sim$27) and proportion of sampling ($\sim$0.91) show that at this α-value, the probability of capturing only the deepest nodes is $\sim$0.98, consistent with significant overdispersion in sampling. Given this was a meta-analysis and we are not familiar with the taxonomy of the respective groups, it is difficult to assess the likelihood of this extreme a degree of sampling bias in each tree, nor whether the same degree of bias is found in each data set. However, the robustness of the results to even the most egregiously biased mode of sampling makes it unlikely that the bird results are an artifact of overdispersed sampling.

The relative robustness of Phillimore and Price (2008) results lend support to early-burst patterns of cladogenesis being the predominant mode of diversification in the bird clades investigated and is consistent with diversification being density dependent (as the authors argue). This robustness is not surprising in retrospect. As noted above, Phillimore and Price (2008) only included data sets for which sampling was >70% complete and for which the mean level of sampling was relatively high ($\mu = 90.8\%$).

Overdispersed Sampling and Temporal Patterns of Diversification

The predominance of negative γ-values in empirical data sets has been used to argue in support of lineage diversification being largely ecologically driven (e.g., McPeek 2008; Phillimore and Price 2008; Rabosky and Lovette 2008a; Rabosky 2009) and indeed this is one possibility. It has long been argued (e.g., Simpson 1953) that clades may radiate into unoccupied niche space, with diversification being maximal in the earliest stages (when most niches remain unfilled) and slowing as the lineage diversifies. Although these models have been largely verbal (Simpson 1953; Schluter 2000), two recent studies employing explicit mathematical models of ecological diversification (e.g., Gavrilets and Vose 2005; McPeek 2008; see also Rabosky 2009) have demonstrated that early-burst patterns of cladogenesis are a common outcome. Although these processes may be sufficient in explaining the temporal patterns of diversification seen in empirical trees, they are by no means necessary, and other explanations (including the pattern being partly an artifact of methodology) remain possible.

Overdispersed sampling has an important impact on the distribution of branching events in a phylogenetic tree, as seen in the elevated type-I error in the standard MCCR test. Given the popularity of this method (especially in meta-analyses of diversification; e.g., Ruber and Zardoya 2005; Mooers et al. 2007; McPeek 2008; Phillimore and Price 2008) and the uncertainty of the degree of overdispersed sampling in most studies, our results provide further evidence in support of caution

when interpreting trends in γ-values as indicative of an underlying biological process. Recent studies have found that the method of phylogenetic reconstruction (see also Revell et al. 2005; Ruber and Zardoya 2005; Rabosky and Lovette 2008b), the extant clade size (Price 2007; Phillimore and Price 2008) and sampling mode (this study) can all create more negative γ-values as artifacts. Thus, before biological processes are invoked to explain an apparent macroevolutionary pattern, these potential artifacts should be accommodated. The significant effect that overdispersed sampling has on the MCCR test, in our opinion, makes assessing the sensitivity to NRS the default approach when assessing temporal patterns of diversification, especially in poorly sampled phylogenies. This test can be combined with other sensitivity analyses to assess the impact of a number of factors simultaneously. For instance, cryptic speciation tends to leave out more tipwardly distributed taxa and one method of accounting for this is to evaluate the sensitivity of the MCCR results to the number of extant taxa (see above). The αNRS test presented here can be combined with this approach, allowing researchers to evaluate the affect of both overdispersed sampling and cryptic species (essentially a specific type of the former) on the MCCR results.

Lastly, it's worth emphasizing that any comparative method that is dependent on the distribution of nodes in a tree, including likelihood-based model fitting approaches to diversification (Rabosky 2006; Rabosky et al. 2007; Rabosky and Lovette 2008) will be affected by NRS. As here, it may be necessary to consider the sampling scheme used when analyzing incompletely sampled phylogenies using these methods (Rabosky 2006; Rabosky and Lovette 2008b).

Conclusion

Overdispersed sampling substantially increases type-I error in the standard MCCR test. Our corrected MCCR tests helps to alleviate this bias and behave well under the sampling schemes investigated. Uncertainty in the degree of overdispersion when sampling makes it difficult to assess which correction is most appropriate. As such, we present a generalized framework (the α NRS test) for assessing the sensitivity of the results of an MCCR test to varying degrees of sampling overdispersion. When assessing temporal patterns of diversification in poorly sampled phylogenies ($<$75% complete), it is advisable to evaluate how sensitive one's results are to sampling mode. Although early-burst patterns of cladogenesis is the dominate pattern seen in empirical phylogenies, it still remains to be seen how much this reflects underlying biological processes versus methodological artifact and care should be taken when interpreting trends in diversification analyses.

Funding

This work was supported in part by and National Science Foundation (NSF) training grant to C.D.B. (DGE 0549425) and NSF grant 0445453 to M.E.A.

Acknowledgments

Useful comments on this manuscript and related presentations were provided by Hugo Alamillo, Barb Banbury, Devin Drown, and Mike Webster. The authors would also like to thank Todd Oakley, Trevor Price, and one anonymous reviewer for their constructive comments and suggestions.

References

Alfaro M.E., Santini F., Brock C.D. 2007. Do Reefs drive diversification in marine teleosts? Evidence from the pufferfishes and their allies. Evolution. 61:2104–2126.

Alfaro M.E., Karns, D.R., Voris, H.K., Brock, C.D., Stuart B.L. 2007. Phylogeny and evolutionary history of Oriental-Australian rear-fanged water snakes (Colubridae:Homalopsidae) inferred from mitochondrial and nuclear DNA sequences. Mol. Phylo. Evol. 46: 576–593.

Alfaro M.E., Santini, F., Brock, C.D., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G. and Harmon. L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. PNAS 32: 13410–13414.

Avise J.C., Walker D. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proc. R. Soc. B: Biol. Sci. 265:457–463.

Bickford D., Lohman D.J., Sodhi N.S., Ng P.K.L., Meier R., Winker K., Ingram K.I., Das I. 2006. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22:3.

Cox D.R., Lewis P.A.W. 1966. Statistical analyses of series of events. London: Chapman & Hall.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland, (MA): Sinauer Association.

Gavrilets S., Vose A. 2005. Dynamic patterns of adaptive radiation. Proc. Natl. Acad. Sci. U.S.A. 102:18040–18045.

Harmon L.J., Schulte J.A., Losos J.B., Larson A. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. Science. 301: 961–964.

Harmon L.J., Wier J., Brock, C., Glor R.E., Challenger W. 2008. Geiger: a statistical package for investigating evolutionary radiations in a comparative context. Bioinformatics. 24:129–131.

Hubbell S.P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton (NJ): Princeton University Press.

Kirkpatrick M., Slatkin M. 1993. Searching for evolutionary pattern in the shape of a phylogenetic tree. Evolution. 47:1171–1181.

Kozak K.H., Weisrock D.W., Larson A. 2006. Rapid lineage accumulation in a non-adaptive radiation: phylognetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). Proc. R. Soc. Lond. B. 273: 539–546.

McPeek M.A. 2008. The ecological dynamics of clade diversification and community assembly. Am. Nat. 172:E270–E284.

Mooers A., Harmon L.J., Blum M.G.B., Wong D.H.J., Heard S.B. 2007. Some models of phylogenetic tree shape. In: O. Gascuel, M. Steel, editors. Reconstructing evolution: new mathematical and computational advances. New York: Oxford University Press, p. 149–170.

Moore B.R., Chan K.M.A., Donoghue M.J. 2004. Detecting diversification rate variation in supertrees. In: O.R.P. Bininda-Edmonds, editor. Phylogenetic supertrees: combining information to reveal the tree of life. Dordrecht (The Netherlands): Kluwer Academic.

Moyle R.G., Filardi C.E., Smith C.E., Diamond J. 2009. Explosive Pleistocene diversification and hemispheric expansion of a "great speciator." Proc. Natl. Acad. Sci. U.S.A. 106(4):1863–1868.

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B. 344:305–311.

Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. B. 267:2267–2272.

Phillimore A.P. and Price T. 2008. Density-dependent cladogenesis in birds. PLOS Biol. 6:e71.

Price T. 2007. Speciation in birds. Greenwood Village (CO): Roberts and Company.

Rabosky D.L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution. 60:1152–1164.

Rabosky D.L., Donellan S.C., Talaba A.L., Lovette I.J. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's largest vertebrate clade. Proc. R. Soc. Lond. B. 274:2915–2923.

Rabosky D.L., Lovette I.J. 2008a. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution. 62:1866–1875.

Rabosky D.L., Lovette I.J. 2008b. Density dependent diversification in North American wood-warblers. Proc. R. Soc. B. 275:2363–2371.

Rabosky, D. L. 2009. Ecological limits on clade diversification in higher taxa. Am. Nat. 173:662-674.

Revell L.J., Harmon L.J., Glor R.E. 2005. Underparameterized model of sequence evolution leads to a bias in the estimation of diversification rates from molecular phylogenies. Sys. Biol. 54: 973–983.

Ruber L., Zardoya R. 2005. Rapid cladogenesis in marine fish revisited. Evolution. 59:1119–1127.

Schluter D. 2000. The ecology of adaptive radiations. New York: Oxford University Press.

Simpson, G. G. 1953. The major features of evolution. New York: Columbia University Press.

Weir J.T. 2006. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. Evolution. 60: 842–855.

Weir J.T., Schluter D. 2004. Ice sheets promote speciation in boreal birds. Proc. R. Soc. Lond. B. 271:1881–1887.

Wiens J.J. 2007. Global patterns of species richness and diversification in amphibians. Am. Nat. 170:S86–S106.

Zink R.M., Slowinski J.B. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. Proc. Natl. Acad. Sci. U.S.A. 92:5832–5835.