

- Riedl, R. 1978. Order in living organisms: A systems analysis of evolution. John Wiley, Chichester.
- Rudall, P. J., and R. M. Bateman. 2002. Roles of synorganisation, zygomorphy and heterotopy in floral evolution: The gynostemium and labellum of orchids and other lilioid monocots. *Biol. Rev. (Camb.)* 77:403–441.
- Sanderson, M. J., and J. J. Doyle. 1992. Reconstruction of organismal phylogenies from multigene families: Paralogy, concerted evolution, and homoplasy. *Syst. Biol.* 41:4–17.
- Schuh, R. T. 2000. Biological systematics: Principles and applications. Cornell University Press, Ithaca, New York.
- Shubin, N. H., and C. R. Marshall. 2000. Fossils, genes, and the origin of novelty. Pages 324–340 in *Deep time: Paleobiology's perspective* (D. H. Erwin and S. L. Wing, eds.). *Paleobiology* 26 (4, supplement).
- Simmons, M. P., C. D. Bailey, and K. C. Nixon. 2000. Phylogeny reconstruction using duplicate genes. *Mol. Biol. Evol.* 17:469–473.
- Simmons, M. P. and J. V. Freudenstein. 2002a. Artifacts of coding amino acids and other composite characters for phylogenetic analysis. *Cladistics* 18:354–365.
- Simmons, M. P., and J. V. Freudenstein. 2002b. Uninode coding vs gene tree parsimony for phylogenetic reconstruction using duplicate genes. *Mol. Phylogenet. Evol.* 23:481–498.
- Sneath, P. A., and R. R. Sokal. 1962. Numerical taxonomy. *Nature* 193:855–860.
- Solignac, M., C. Periquet, D. Anxolabehere, and C. Petit. 1995. *Génétique et evolution*. Hermann, Paris.
- Sonnhammer, E. L. L., and E. V. Koonin. 2002. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* 18:619–620.
- Stadler, B. M. R., P. F. Stadler, G. P. Wagner, and W. Fontana. 2001. The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.* 213:241–274.
- Wagner, G. P. 1989a. The biological homology concept. *Annu. Rev. Ecol. Syst.* 20:51–69.
- Wagner, G. P. 1989b. The origin of morphological characters and the biological basis of homology. *Evolution* 43:1157–1171.
- Wagner, G. P., and P. F. Stadler. 2003. Quasi-independence, homology and the unity of type: A topological theory of characters. *J. Theor. Biol.* 220:505–527.
- Wheeler, W. C. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? *Cladistics* 12:1–10.
- Wheeler, W. C., and D. S. Gladstein. 1994. MALIGN: A multiple sequence alignment program. *J. Hered.* 85:417–418.
- Wheeler, W. C., and R. L. Honeycutt. 1988. Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Mol. Biol. Evol.* 5:90–96.
- Wiley, E. O. 1978. The evolutionary species concept reconsidered. *Syst. Zool.* 27:88–92.
- Wiley, E. O. 1980. Phylogenetic systematics and vicariance biogeography. *Syst. Bot.* 5:194–220.
- Wiley, E. O. 1981. *Phylogenetics*. John Wiley and Sons, New York.
- Wilkinson, 1995. A comparison of two methods of character construction. *Cladistics* 11:297–308.
- Won, H., and S. S. Renner. 2003. Horizontal gene transfer from flowering plants to *Gnetum*. *Proc. Natl. Acad. Sci. USA* 100:10824–10829.
- Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18:292–298.
- Zimmer, E. A., S. L. Martin, S. M. Beverely, Y. W. Kan, and A. C. Wilson. 1980. Rapid duplication and loss of genes coding for the  $\alpha$  chains of hemoglobin. *Proc. Natl. Acad. Sci. USA* 77:2158–2162.
- Zouine, M., Q. Sculo, and B. Labedan. 2002. Correct assignment of homology is crucial when genomics meets molecular evolution. *Comp. Funct. Genomics* 3:488–493.

First submitted 29 July 2004; reviews returned 2 February 2005;  
final acceptance 7 June 2005  
Associate Editor: Allan Baker

*Syst. Biol.* 54(6):973–983, 2005  
Copyright © Society of Systematic Biologists  
ISSN: 1063-5157 print / 1076-836X online  
DOI: 10.1080/10635150500354647

## Underparameterized Model of Sequence Evolution Leads to Bias in the Estimation of Diversification Rates from Molecular Phylogenies

LIAM J. REVELL,<sup>1</sup> LUKE J. HARMON,<sup>1,2</sup> AND RICHARD E. GLOR<sup>1,3</sup>

<sup>1</sup>Department of Biology, Campus Box 1229, Washington University, St. Louis, Missouri 63130, USA; E-mail: ljrevell@wustl.edu (L.J.R.)  
<sup>2</sup>Current address: Biodiversity Centre, University of British Columbia, 6270 University Blvd., Vancouver, British Columbia V6T 1Z4, Canada;  
E-mail: harmon@zoology.ubc.ca  
<sup>3</sup>Current address: Center for Population Biology, University of California, One Shields Avenue, Davis, California 95616, USA;  
E-mail: reglor@ucdavis.edu

Macroevolutionary inferences from molecular phylogenies are becoming increasingly common (see Harvey et al., 1996; Mooers and Heard, 1997; Pagel, 1999; Barraclough and Nee, 2001). Many methods in which phylogenies are invoked for historical inference assume that a molecular phylogeny is an errorless representation of the underlying phylogenetic history of the included taxa (but see Lutzoni et al., 2001; Huelsenbeck et al., 2000; Huelsenbeck and Rannala, 2003). However, molecular phylogenies are estimates of this history based on a particular model of evolution; thus, there is some

error associated with their estimation (Huelsenbeck and Kirkpatrick, 1996). Here we explore the effects of a particular type of error in phylogenetic branch-length estimation, that caused by assuming an underparameterized model of molecular evolution, on the  $\gamma$ -statistic of Pybus and Harvey (2000), a statistic that tests for changes in the rate of diversification through time. Although we restrict our attention to the estimation of diversification rates, our findings are germane to any macroevolutionary inferences relying on the accurate estimation of phylogenetic branch lengths such as molecular dating (e.g.,

Welch and Bromham, 2005) and probabilistic methods for ancestral state reconstruction (e.g., Ronquist, 2004).

In phylogenetic analyses, models of molecular evolution are often used to estimate branch lengths so that distances separating species on the reconstructed phylogeny are proportional to the time since the species shared a common ancestor multiplied by the substitution rate on the branches separating the taxa (Felsenstein, 2004). When recovering these branch lengths, models of nucleotide substitution are necessary to correct for the effect of superimposed substitutions as genetic differences between taxa accrue. Such models vary from simple to complex. The simplest model, the Jukes and Cantor (1969) model, assumes that all types of nucleotide substitutions are equiprobable, that all sites share a common substitution rate, and that base frequencies are equal. If these assumptions are violated, distances calculated using a Jukes-Cantor correction will likely be underestimates of the actual extent of evolutionary divergence, with misestimation particularly severe for branches connecting more divergent sequences (Yang et al., 1994; Lemmon and Moriarty, 2004). More sophisticated models allow rate heterogeneity across sites (Uzzell and Corbin, 1971; Jin and Nei, 1990), incorporate invariant sites (Hasegawa, 1987), or allow specific types of substitutions to occur at different rates (Kimura, 1980; Lanave et al., 1984; Felsenstein, 2004). The most heavily parameterized model commonly used in phylogenetic studies (the general time reversible model with invariant sites and  $\Gamma$ -distributed rate heterogeneity, GTR + I +  $\Gamma$ : Uzzell and Corbin, 1971; Jin and Nei, 1990; Rodríguez et al., 1990) requires the specification of 10 parameters (Felsenstein, 2004).

In phylogenetic analyses, models of molecular evolution are often used to estimate branch lengths so that distances separating species on the reconstructed phylogeny are proportional to the time since the species shared a common ancestor multiplied by the substitution rate on the branches separating the taxa (Felsenstein, 2004). When recovering these branch lengths, models of nucleotide substitution are necessary to correct for the effect of superimposed substitutions as genetic differences between taxa accrue. Such models vary from simple to complex. The simplest model, the Jukes and Cantor (1969) model, assumes that all types of nucleotide substitutions are equiprobable, that all sites share a common substitution rate, and that base frequencies are equal. If these assumptions are violated, distances calculated using a Jukes-Cantor correction will likely be underestimates of the actual extent of evolutionary divergence, with misestimation particularly severe for branches connecting more divergent sequences (Yang et al., 1994; Lemmon and Moriarty, 2004). More sophisticated models allow rate heterogeneity across sites (Uzzell and Corbin, 1971; Jin and Nei, 1990), incorporate invariant sites (Hasegawa, 1987), or allow specific types of substitutions to occur at different rates (Kimura, 1980; Lanave et al., 1984; Felsenstein, 2004). The most heavily parameterized model commonly used in phylogenetic stud-

ies (the general time reversible model with invariant sites and  $\Gamma$ -distributed rate heterogeneity, GTR + I +  $\Gamma$ : Uzzell and Corbin, 1971; Jin and Nei, 1990; Rodríguez et al., 1990) requires the specification of 10 parameters (Felsenstein, 2004).

Numerous studies have characterized the sensitivity of phylogenetic reconstruction to model misspecification (e.g., Fukami-Kobayashi and Tateno, 1991; Olsen, 1991; Ruvolo et al., 1993; Yang et al., 1994, 1995; Gaut and Lewis, 1995; Adachi and Hasegawa, 1995). Often, topological inference has been found to be robust to model misspecification even when branch lengths are severely misestimated (e.g., Fukami-Kobayashi and Tateno, 1991; Gaut and Lewis, 1995; Håstad and Björklund, 1998; Lemmon and Moriarty, 2004). Ignoring heterogeneity in substitution rate among sites has the most severe effect on branch lengths, causing them to be consistently underestimated (Yang et al., 1994; Gaut and Lewis, 1995).

The  $\gamma$ -statistic of Pybus and Harvey (2000) measures the relative positions of internal nodes in a tree (Pybus and Harvey, 2000) and is defined as:

$$\gamma = \frac{\left[ \frac{1}{n-2} \sum_{i=2}^{n-1} \left( \sum_{k=2}^i k g_k \right) \right] - \frac{\sum_{j=2}^n j g_j}{2}}{\left( \sum_{j=2}^n j g_j \right) \sqrt{\frac{1}{12(n-2)}}} \quad (15)$$

where  $n$  is the number of taxa in the phylogeny, and  $g_2, g_3, g_4, \dots, g_n$  are the ordered internode intervals from the root (Pybus et al., 2002). Under constant-rate pure-birth cladogenesis, the  $\gamma$ -statistic has a standard normal distribution (Pybus and Harvey, 2000). If  $\gamma$  is significantly less than zero, internal nodes are concentrated closer to the base of the tree than expected under a constant-rate model of diversification, suggesting that the net diversification rate of the group has slowed over time; conversely, positive  $\gamma$  indicates that more internal nodes are concentrated near the tips of the tree and suggests that the net rate of diversification has increased through time (Pybus and Harvey, 2000; Pybus et al., 2002).

In several empirical studies, the  $\gamma$ -statistic shows highly significant departures from constant-rate pure-birth cladogenesis (Harmon et al., 2003; Linder et al., 2003; Shaw et al., 2003; Kadereit et al., 2004; Machordom and Macpherson, 2004; Williams and Reid, 2004; Zhang et al., 2004). Several of these studies reveal slow-downs in the net rate of diversification through time (Harmon et al., 2003; Kadereit et al., 2004; Machordom and Macpherson, 2004; Williams and Reid, 2004; Zhang et al., 2004). For example, Harmon et al. (2003) found highly significantly negative  $\gamma$  in three of four lizard groups, with the fourth also negative but nonsignificant, and Kadereit et al. (2004) found similar results amongst several genera of alpine plants. In each case the authors hypothesized that geographical or ecological processes contributed to the observed patterns. In contrast, Linder et al. (2003) found significantly positive  $\gamma$  among African Restionaceae.

Considerable attention has been paid to the effect of incomplete taxonomic sampling (Pybus and Harvey, 2000; Pybus et al., 2002) and the method of phylogenetic ultrametricization (Barracough and Vogler, 2002; Martin et al., 2004; Rüber and Zardoya, 2005) on the estimation of  $\gamma$ . However, the effect of the assumptions of phylogenetic inference, and in particular the model of sequence evolution, has been given relatively short shrift. Many phenomena of interest to evolutionary biologists, such as the estimation of diversification rates, rely on phylogenetic branch lengths as estimates of the temporal extent of evolutionary divergence, and branch-length estimation is highly sensitive to proper model specification (e.g., Fukami-Kobayashi and Tateno, 1991; Ruvolo et al., 1993; Gaut and Lewis, 1995; Håstad and Björklund, 1998; Lemmon and Moriarty, 2004). As such, assessment of substitution model adequacy, a criterion that may be rarely satisfied by empirical data if the process of molecular evolution is considerably more complicated than our models describe (see Goldman, 1993), is crucial when evolutionary parameters are to be estimated from a molecular phylogeny with branch lengths.

Because the  $\gamma$ -statistic depends on branch length estimates, it may be adversely affected by estimating branches in the molecular phylogeny using an under-specified or inadequate model. In fact, Pybus and Harvey (2000) note that the statistic is liable to be misestimated if factors affecting error in branch-length estimation act unevenly in the tree. Here we analyze the extent of the bias in  $\gamma$  resulting from nucleotide substitution model underparameterization and focus on characterizing its magnitude and direction, as well as its interaction with other factors such as tree balance, tree size, total tree depth, and sequence length. We also describe the particular circumstances under which these potential biases are likely to be a concern for empirical studies.

#### SIMULATION AND ANALYSES

As we are primarily interested in testing the effect of model under parameterization on the estimation of  $\gamma$ , most of our analyses focus on relatively simple models of molecular evolution that vary with respect to a limited number of critical parameters (e.g., number of substitution parameters, rate heterogeneity among sites, invariant sites), while ignoring variation in tree balance, number of taxa, tree length, and size of the nucleotide data set. However, because these factors invariably differ among empirical studies and likely influence the estimation and hypothesis testing of  $\gamma$ , we also explore the impact of these additional factors using a somewhat more restricted set of simulations. Finally, because empirical studies usually use more heavily parameterized models of nucleotide substitution than those explored in the aforementioned analyses, we investigate the effect of more complicated models of sequence evolution on the estimation of  $\gamma$ .

##### *Simulation Test for Effect of Underparameterization*

We used simulations of phylogenies and associated nucleotide data sets to test the effect of model parame-

terization on the estimation of  $\gamma$  under a range of conditions. We first used the program Phyl-O-Gen (Rambaut, 2002) to simulate 100 phylogenies containing 100 taxa under a constant-rate pure birth model—the null model assumed by Pybus and Harvey (2000) for the  $\gamma$ -statistic. This set of 100 trees was used in the majority of the analyses described below.

We then used the program Seq-Gen (Rambaut and Grassly, 1997) to simulate 1000 base-pair data sets on each phylogeny under three of different models of molecular evolution and a range of parameter values for each model: (1) the Jukes-Cantor model (JC; Jukes and Cantor 1969) with heterogeneous rates among sites, under a four category discrete approximation of  $\Gamma$ -distributed rate heterogeneity (JC +  $\Gamma$ ; Uzzell and Corbin, 1971; shape parameter of the  $\Gamma$  distribution  $\alpha_{\Gamma} = 0.1, 0.5, 1.0, 5.0$ , and  $10.0$ ; Jin and Nei, 1990); (2) the Jukes-Cantor model with invariant sites (JC + I; proportion of invariant sites,  $p_{inv} = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ , and  $0.9$ ; Hasegawa et al., 1987); and (3) the K80 model (also known as the Kimura 2-parameter model; Kimura, 1980), in which transitions and transversions are allowed different rates (transition : transversion ratio  $\kappa = 0.5, 1, 2, 4, 8$ , and  $16$ ). For each simulated data set, we scaled total tree length to 0.5 substitutions per site.

For each simulated data set, we estimated maximum likelihood branch lengths using PAUP\*4b10 (Swofford, 1998) while constraining the correct topology and assuming a molecular clock under two models: (1) the generating model (JC+ $\Gamma$ , JC+I, or K80) with its known parameter value, and (2) the simpler JC model. We then calculated  $\gamma$  for each of the trees with estimated branch lengths. All  $\gamma$  calculations were performed using a C program, ltt.c, available from the authors upon request.

Because all phylogenies were simulated under a pure-birth model with a constant speciation rate over time, we expect that  $\gamma$  will have a standard normal distribution. Thus, we calculated one-tailed type I error rates as the proportion of data sets for which the value of  $\gamma$  for the estimated phylogeny was significant (i.e., fell below 95% of the standard normal distribution). We focus on the left tail of the distribution because a one-tailed rejection threshold is often applied in empirical studies (e.g., Pybus and Harvey, 2000; Harmon et al., 2003; Kadereit et al., 2004). We also calculated bias as the average difference between the estimated value of  $\gamma$  and its true value. We assessed the significance of the type I error rate by comparing it to the one-tailed 95% binomial confidence limit assuming that the true type I error rate of the statistic was 0.05. We considered the elevation in type I error significant if it exhibited a value in excess of this limit.

##### *Tree Balance*

Because tree balance can influence phylogenetic reconstruction (Heard, 1992; Huelsenbeck and Kirkpatrick, 1996; Mooers et al., 1995; Mooers and Heard, 1997), we calculated mean corrected  $I'$  (a metric for assessing tree balance: Fusco and Cronk, 1995; Purvis et al., 2002; Agapow and Purvis, 2002) for each of the 100 simulated trees using a C program (balance.c) available from

TABLE 1. Empirical parameters used to test the effect of model underparameterization when data were simulated using realistic models and parameter values.

Model	Gene	Rate matrix						Base frequencies				$\alpha_\Gamma$	$p\text{-inv}$	Source
		AG	AC	AT	GC	GT	CT	%A	%C	%G	%T			
a	RAG-1	1.33	4.70	0.92	0.86	5.72	1.0	0.30	0.22	0.23	0.25	1.71	0.34	Townsend et al. (2004)
b	ND-2	0.36	2.59	0.50	0.32	2.48	1.0	0.41	0.34	0.06	0.19	0.57	0.10	Townsend et al. (2004)
c	ND-2	1.13	16.7	0.84	1.42	12.8	1.0	0.40	0.20	0.07	0.33	0.67	0.28	Kozak et al. (2005)

the authors. For a subset of our simulated data sets, we tested for a significant relationship between the deviation of  $\gamma$  from its true value and the mean corrected imbalance of the simulated phylogeny. For this analysis, we focused on data sets simulated under JC+ $\Gamma$  with  $\alpha_\Gamma = 0.5$ , JC+I with  $p\text{-inv} = 0.5$ , and K80 with  $\kappa = 16$ , with associated branch-length estimates made assuming an underparameterized substitution model (JC), because each of these cases showed substantial bias in  $\gamma$  (see Results).

#### *Number of Taxa, Total Tree Length, and Sequence Length*

To investigate the effect of number of taxa, total tree length, and sequence length on the estimation and hypothesis testing of  $\gamma$  we restricted our attention to a single substitution model, JC+ $\Gamma$  with  $\alpha_\Gamma = 0.5$ . We simulated sequences on pure-birth phylogenies containing various numbers of taxa, while varying total tree length. To do this, we simulated 100 stochastic phylogenies that included 10 to 100 taxa at intervals of 10 taxa. We then simulated sequences with 1000 characters on each of these topologies while scaling total tree length to between 0.1 and 1 at intervals of 0.1 substitutions per site. We also simulated sequence data sets with varying numbers of characters, again while varying total tree length. We simulated data sets on 100 taxon phylogenies with nucleotide sequence lengths ranging from 100 to 1000 at 100-nucleotide intervals and used the same range of total tree lengths as in the previous analysis. We then estimated branch lengths, again constraining to the correct topology, for each resulting data set assuming both the generating model (JC+ $\Gamma$ ) and the underparameterized model (JC). For each phylogeny with estimated branch lengths we calculated  $\gamma$ , and for each set of phylogenies (corresponding to an estimation model and a set of tree length and number of taxa or sequence length) we calculated mean deviation from the true  $\gamma$ , and type I error rate in  $\gamma$  (with its associated binomial probability assuming no bias) for both the full and underparameterized models.

#### *More Complex Models*

One limitation of the above simulations is that empirical studies usually use more complex models to estimate branch lengths than the JC and K80 models discussed above. Furthermore, parameters are often estimated from the data rather than known a priori. To investigate the behavior of  $\gamma$  under more realistic conditions, we simulated sequence data under the most complex model that is commonly employed in molecular phylogenetic studies, the general time reversible model with invariant sites and  $\Gamma$ -distributed rate heterogeneity

(GTR+I+ $\Gamma$ ; Uzzell and Corbin, 1971; Jin and Nei, 1990; Rodríguez et al., 1990). Because this model has 10 parameters, we could not reasonably explore the entire parameter space in our simulations; instead, we used three sets of substitution rates, base frequencies, invariant sites, and rate heterogeneity drawn from empirical studies of mitochondrial DNA and a nuclear intron (Table 1; Townsend et al., 2004; Kozak et al., 2005). For each set of parameters, we simulated a single 1000-nucleotide data set on each of five pure-birth phylogenies generated using Phyl-O-Gen, with 100 taxa per phylogeny and total tree length scaled to 0.5 substitutions per site. We then used maximum likelihood in PAUP\*4b10 to estimate parameter values and branch lengths on these topologies under the full model (GTR+I+ $\Gamma$ ) and seven models representing special cases of the generating model: JC, F81 (Felsenstein, 1981), HKY (Hasegawa et al., 1985), TrN (Tamura and Nei, 1993), TIM (Rodríguez et al., 1990), GTR (Rodríguez et al., 1990), and GTR+I (Hasegawa et al., 1987; Rodríguez et al., 1990). We then calculated  $\gamma$  for these estimated phylogenies and compared these estimates to their known values.

## RESULTS

### *Simulation Test for Effect of Underparameterization*

For data generated under a range of rate heterogeneities,  $\gamma$  is unbiased and exhibits appropriate type I error rates when estimated under the correct model (JC+ $\Gamma$ ) (Fig. 1a, b, open dots). In contrast,  $\gamma$  is strongly negatively biased, with elevated type I error rates, when branch lengths for molecular sequences simulated with moderate to high degrees of rate heterogeneity ( $\Gamma$  shape parameter,  $\alpha_\Gamma = 0.1\text{--}1$ ) are estimated with the underparameterized model (JC), which does not incorporate rate heterogeneity (Fig. 1a, b, filled dots). For extreme rate heterogeneity, bias resulting from model underparameterization corresponds with conspicuously short early branches and relatively long late branches in the reconstructed phylogeny (compare Fig. 2a to Fig. 2c), an effect absent from the phylogeny in which branch lengths are estimated using the generating model (Fig. 2b). Not surprisingly, therefore, type I error rates with the underparameterized model are highest (up to 0.82) for the data sets simulated under the highest degree of rate heterogeneity (Fig. 1b). This bias persists, but becomes slight as rates become more uniform across sites ( $\alpha_\Gamma \geq 5$ ; Fig. 1a, b). At  $\alpha_\Gamma \geq 5$  the underparameterized model of sequence evolution no longer exhibits significantly elevated type I error (Fig. 1b).

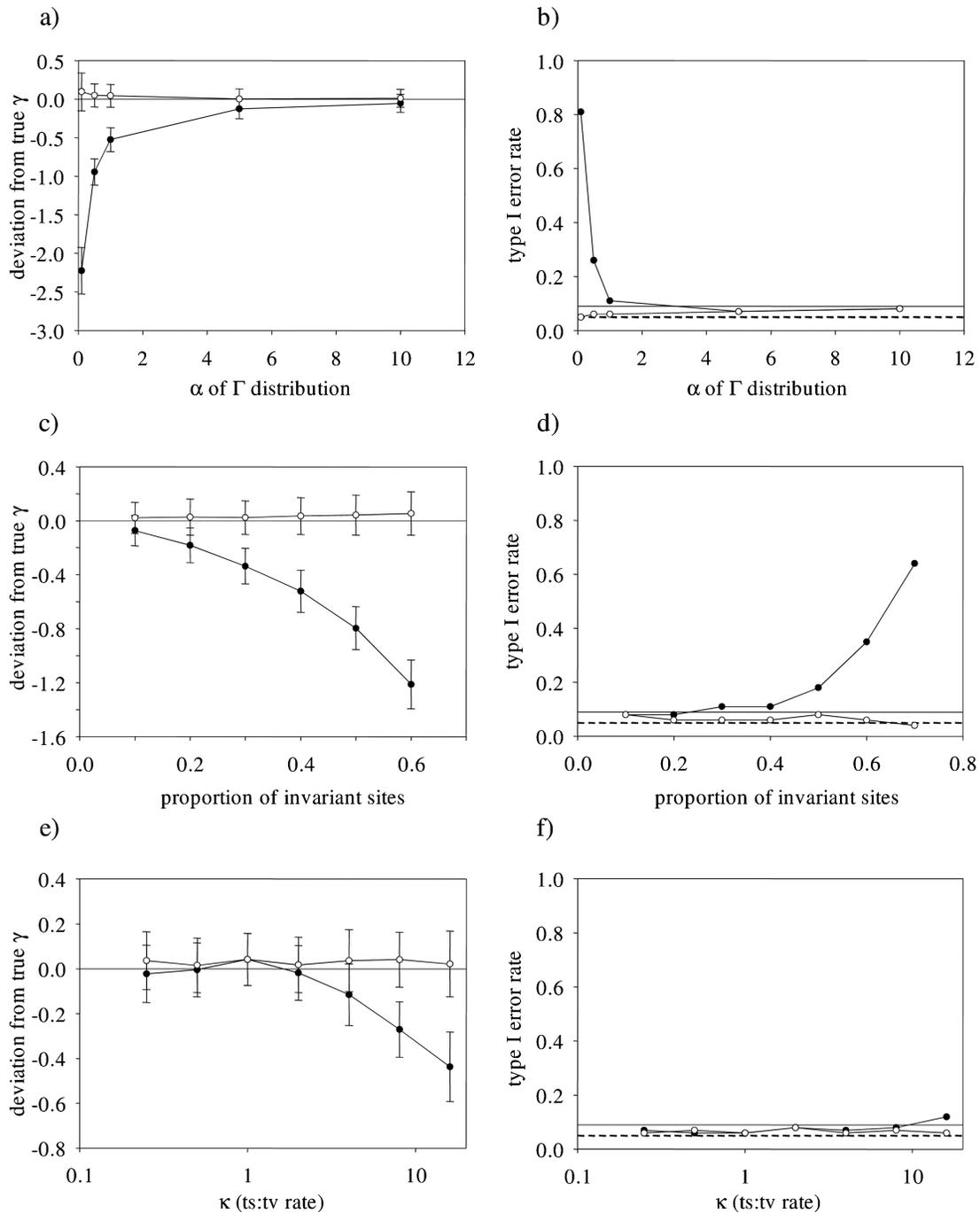


FIGURE 1. Mean and SD of the deviation from true  $\gamma$  and one-tailed type I error rates for  $\gamma$  from simulated molecular phylogenies with branches estimated under fully parameterized (open dots) and under parameterized models of sequence evolution (filled dots). Full models are (a) and (b) Jukes-Cantor with  $\Gamma$ -distributed rate variation among sites for various values of the  $\Gamma$  shape parameter,  $\alpha_\Gamma$ ; (c) and (d) Jukes-Cantor with invariant sites, for various proportions of invariant sites; and (e) and (f) Kimura two-parameter for various values of  $\kappa$ , the ratio of transition to transversion rate. One-tailed type I error is calculated as the frequency of estimated  $\gamma < -1.645$ . The rate (0.05) and 95% confidence threshold for the rate of type I error expected by chance are indicated by dashed and solid horizontal lines, respectively.

When sequences are simulated under a model that incorporates invariant sites (JC+I) and branch lengths are estimated using the fully parameterized model,  $\gamma$  is unbiased with appropriate type I error rates (Fig. 1c, d). However,  $\gamma$  is negatively biased when estimated using the under-parameterized model that lacks the invariant

sites parameter (JC) for even moderate proportions of invariant sites ( $p\text{-inv} = 0.2$ ), with the magnitude of the bias increasing with the proportion of invariant sites (Fig. 1c). Despite this bias, underparameterization results in only insignificantly or marginally significantly elevated type I error when the proportion of invariant sites is relatively

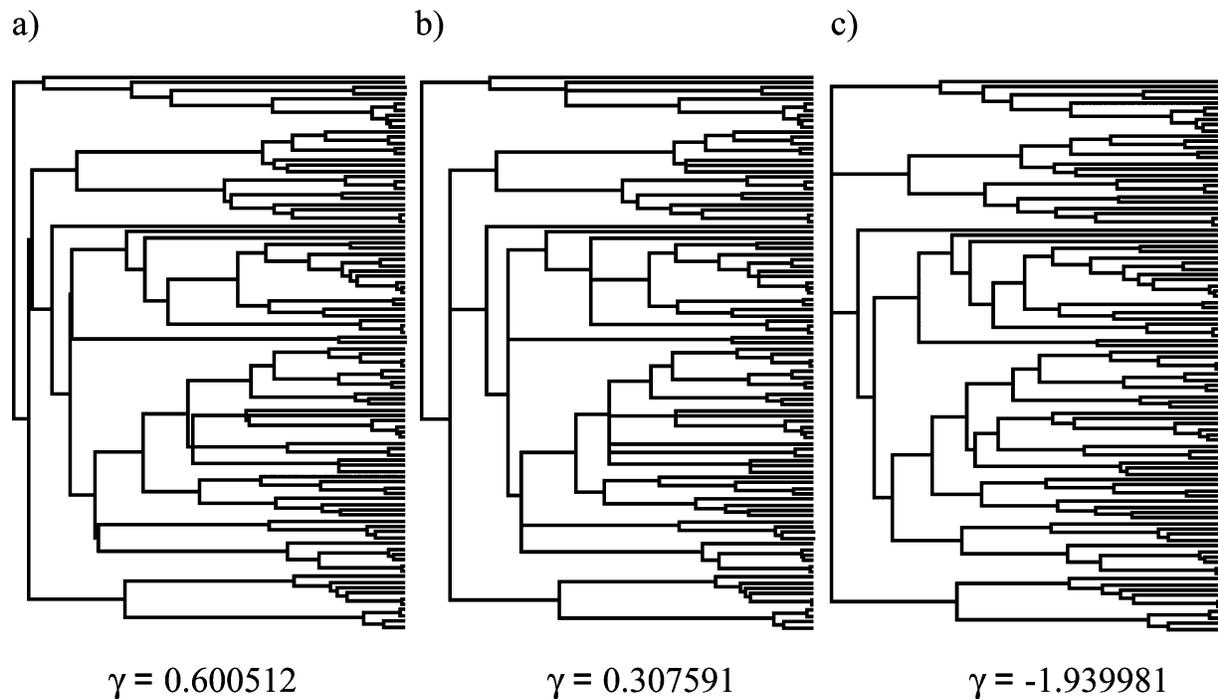


FIGURE 2. Example stochastic pure-birth phylogeny with branch lengths (a) on which sequences were simulated using JC+ $\Gamma$  with shape parameter  $\alpha_{\Gamma} = 0.1$ . Phylogenies with estimated branch lengths assuming (b) the generating model, JC+ $\Gamma$ , and (c) the underspecified model, JC. To highlight relative changes in branch lengths, total tree lengths have been rescaled to be the same in all three phylogenies.

low (Fig. 1d). For proportions of invariant sites between 0.1 and 0.4, one-tailed type I error rates are  $\leq 0.11$ . However, as invariant sites are increased above  $p\text{-inv} = 0.5$ , type I error increases (Fig. 1d) such that all  $\gamma$  values are significant at  $p\text{-inv} = 0.9$ .

When simulated sequence data is evolved under the K80 model, which incorporates different rates of transitions and transversions, and branch lengths are estimated with the fully parameterized model of sequence evolution,  $\gamma$  is unbiased (Fig. 1e) and type I error is low (Fig. 1f). Although  $\gamma$  is slightly biased when estimated under JC (Fig. 1e), the observed bias is of lower magnitude over all tested values of the ratio of transitions to transversions ( $0.25 \leq \kappa \leq 16$ ) than that created by rate heterogeneity or invariant sites (Fig. 1) and does not lead to significantly elevated type I error rates except at the most extreme value of  $\kappa$  tested in this study, at which point type I error is only marginally elevated ( $\kappa = 16$ : type I error = 0.12; Fig. 1f).

#### Tree Balance

There was no significant correlation between tree imbalance and the deviation from true  $\gamma$  when an under-specified model was used in branch-length estimation for any of the models tested (JC+ $\Gamma$ ,  $\alpha_{\Gamma} = 0.5$ :  $r^2 = 0.02$ ,  $P = 0.17$ ; JC+I,  $p\text{-inv} = 0.5$ :  $r^2 = 0.005$ ,  $P = 0.49$ ; K80,  $\kappa = 16$ :  $r^2 = 0.006$ ,  $P = 0.46$ ).

#### Number of Taxa, Total Tree Length, and Sequence Length

When sequences were evolved and branch lengths estimated under the generating model (JC+ $\Gamma$ ), mean de-

viation from true  $\gamma$  was low across all tree depths and numbers of taxa (Fig. 3a). Type I error rates were also low (Fig. 3c), and in no instance was type I error significantly inflated (Fig. 3e). In contrast, when branch lengths were estimated using the underparameterized model (JC), deviation from true  $\gamma$  was negative under all combinations of tree size and length, but most severe for long trees including many taxa (Fig. 3b). One-tailed type I error rates are significantly elevated for trees of length longer than 0.5 substitutions per site, regardless of the number of taxa (Fig. 3f).

When branches were estimated using the full model (JC+ $\Gamma$ ), deviation from true  $\gamma$  was low except in data sets with very short trees and small sequence length, for which substantial positive bias in  $\gamma$  was observed (Fig. 4a). One-tailed type I error rates were not significantly different from 0.05 for all combinations of sequence length and tree depth (Fig. 4c, e). When branch lengths were estimated ignoring rate heterogeneity, deviation from true  $\gamma$  is severe and negative (Fig. 4b), leading to significantly elevated type I error rates for all but the shortest tree depths regardless of sequence length (Fig. 4d, f). For very short sequence length and tree length, there is a slight positive bias in  $\gamma$  (Fig. 4b).

#### Complex Models

In all cases when sequences were simulated under empirically derived parameter values for the GTR+I + $\Gamma$  model,  $\gamma$  is significantly negatively biased when the simplest models (JC, F81) are used to estimate branch

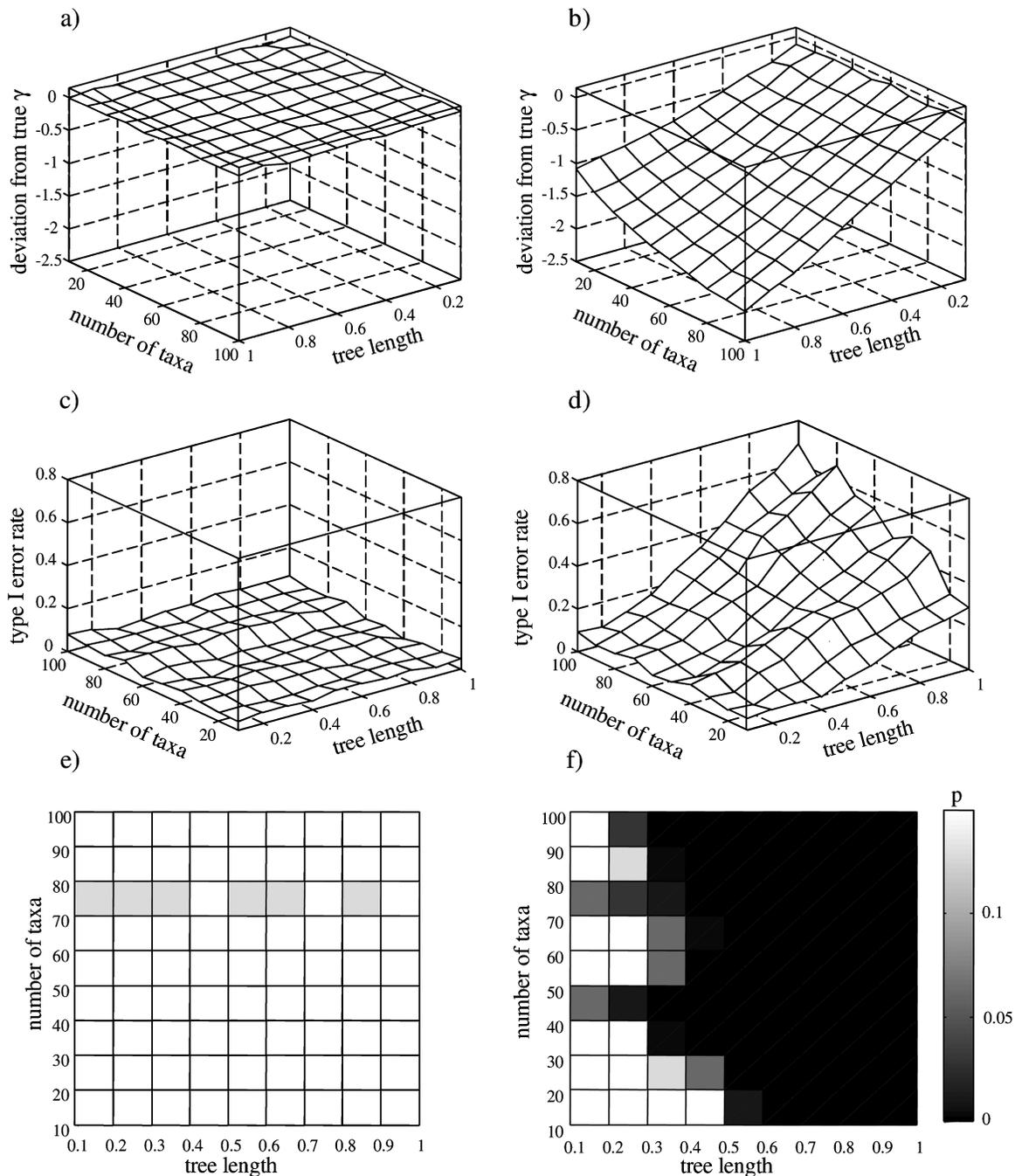


FIGURE 3. Mean deviation from true  $\gamma$  for simulated molecular phylogenies of various size and total length with branches estimated under the fully parameterized (a) and underparameterized models of sequence evolution (b). Full model is the generating model: JC+ $\Gamma$  with shape parameter  $\alpha_{\Gamma} = 0.5$ , whereas the underparameterized model is JC with homogeneous rates among sites. One-tailed type I error rates, calculated as in Figure 1, are also shown for the full (c) and underparameterized (d) models, with binomial significances of the one-tailed type I error in (e) and (f). Each tree size and length combination was explored with 100 simulated trees and data sets. Note that the axes are reversed in (a) and (b) for clarity.

lengths. Minor improvement in  $\gamma$  is observed when more substitution types are included in the model (models HKY and GTR), but  $\gamma$  is still significantly negatively biased. There is also some improvement when invariant sites are included (model GTR+I), but  $\gamma$  values are not unbiased until the correct model (GTR+I+ $\Gamma$ ) is used (Fig. 5). Although the magnitude of the bias depends on

the particular parameter values used, the general patterns are the same in each case.

#### DISCUSSION

Model underparameterization can lead to significant negative bias and inflated type I error in the  $\gamma$ -statistic

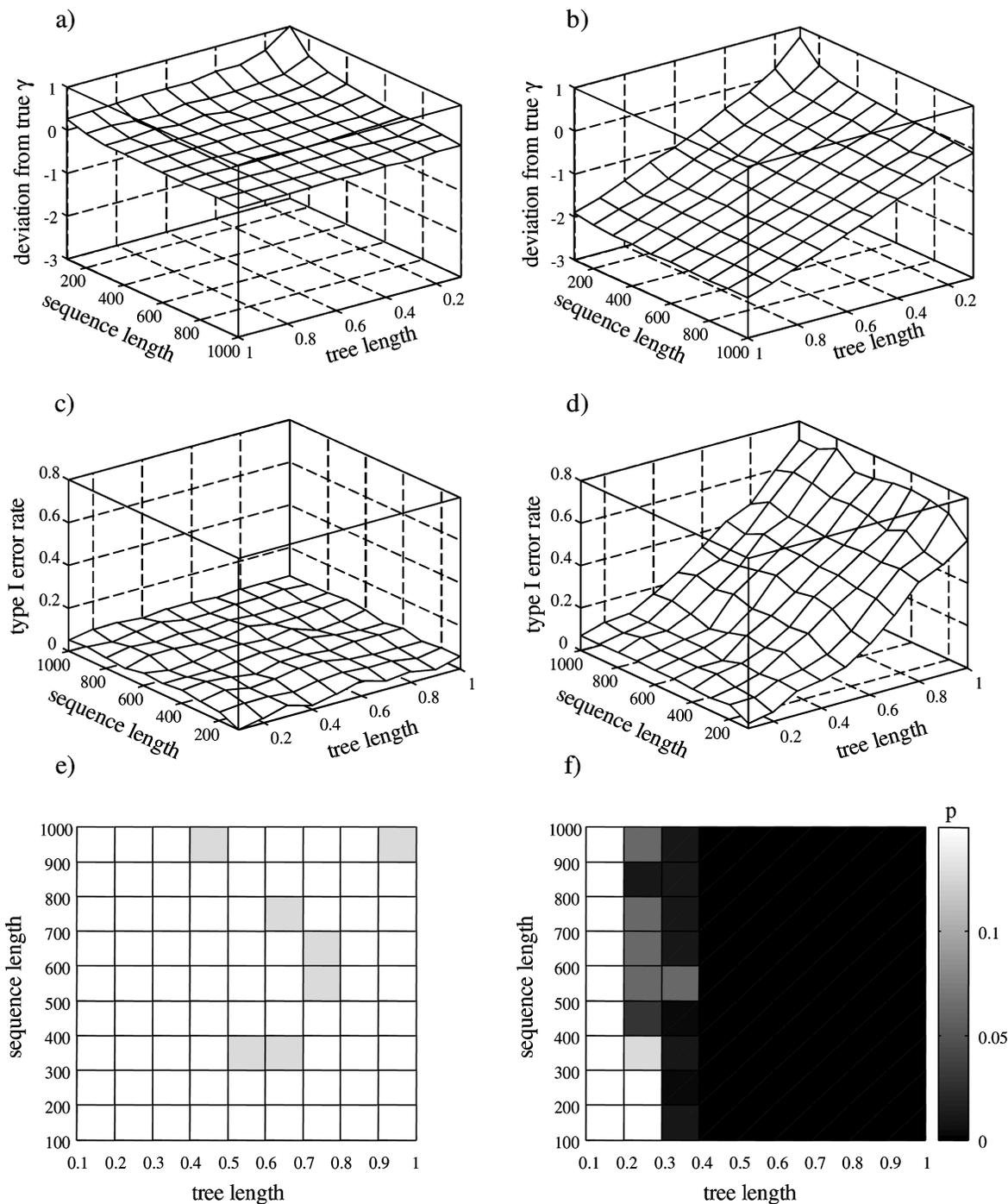


FIGURE 4. Mean deviation from true  $\gamma$  for simulated molecular phylogenies of various total tree depths whose branch lengths were estimated from simulated data sets of varied sequence length using the fully parameterized (a) and underparameterized models of sequence evolution (b). Full model is the generating model: JC +  $\Gamma$  with shape parameter  $\alpha_{\Gamma} = 0.5$ ; underparameterized model is JC with homogeneous rates among sites. One-tailed type I error rates, calculated as in Figures 1 and 2, are also shown for the full (c) and underparameterized (d) models, with binomial significances of the one-tailed type I error in (e) and (f). In (a) and (b) the axes are reversed for clarity.

of Pybus and Harvey (2000) calculated from molecular phylogenies. In other words, model underparameterization may lead to a spurious pattern of rapid cladogenesis early in a group's history. This problem is particularly severe when rate heterogeneity ( $\Gamma$ ) is ignored (see also Lemmon and Moriarty, 2004) but is also significant if

the invariant sites parameter (I) is excluded and when highly unequal transition and transversion rates ( $\kappa$ ) are ignored (Fig. 1). The finding that underparameterization leads to negatively biased values of  $\gamma$  is concordant with the observation that branch lengths are more severely underestimated early rather than late in the tree when

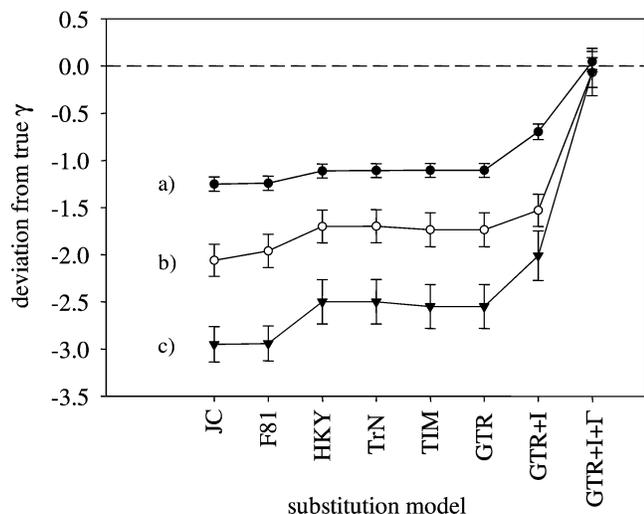


FIGURE 5. Mean deviation from true  $\gamma$  for data simulated using empirical parameters. Simulation models were GTR+I+ $\Gamma$  with various parameter values as listed in Table 1. Shown is the mean and SD of the deviation from true  $\gamma$  from trees with branch lengths estimated under a range of models from simple to complex.

they are estimated using an underspecified substitution model (Gojobori et al., 1982; Kuhner and Felsenstein, 1994). However, the large magnitude of the effect (type I error = 1.0 in some cases) highlights the importance of assessing model adequacy before carrying out analyses of diversification, such as the  $\gamma$  test, on molecular phylogenies with branch lengths.

In considering the consequences of this finding, a distinction must be drawn between model selection, such as the likelihood ratio-based method implemented in ModelTest (Posada and Crandall, 1998), and tests of model adequacy (see Goldman, 1993; Bollback, 2002). Tests of model adequacy consider the absolute, rather than the relative, fit of the data to the prescribed model: such tests may indicate that even the most heavily parameterized model available for phylogenetic analyses (e.g., GTR+I+ $\Gamma$ ) is inadequate despite the fact that it has been selected by a relative criterion such as the likelihood ratio test. Although absolute model adequacy can now be assessed using several methods (e.g., Goldman, 1993; Bollback, 2002; also described in Lemmon and Moriarty, 2004), such tests are stringent and may reject the adequacy of all available models when applied to empirical data sets (see Goldman, 1993).

For the time being, then, we recommend that significantly negative  $\gamma$  be considered cautiously in empirical studies in which model adequacy has not been assessed or in which model adequacy has been rejected. It may be possible to further increase model adequacy—and the resulting accuracy of  $\gamma$  estimation—for empirical data sets by modeling additional aspects of rate heterogeneity that can now be incorporated into modern phylogenetic analyses. For example, Bayesian analyses as implemented in MrBayes (Ronquist and Huelsenbeck, 2003) may be conducted with data sets that are divided into several partitions for which parameters are independently estimated.

It may also be useful to consider the length of the tree, the number of taxa, and the number of characters used in the analysis when considering the potential for bias in the  $\gamma$ -statistic as a consequence of model inadequacy. Trees of very short length were not particularly susceptible to type I error in the  $\gamma$ -test, nor were trees containing few taxa. For long trees, in which branches contain many superimposed substitutions, the consequences on  $\gamma$  due to model misspecification are much more severe. For such trees, adding more taxa actually increases the power of the  $\gamma$ -statistic to detect spurious results such that even mild apparent deviations from constant-rate speciation, whether real or an artifact of model underparameterization, are statistically significant. Thus, our results suggest that trees of long length, trees containing many taxa, or trees featuring both of these properties are particularly susceptible to bias in the estimation and hypothesis testing of diversification rates using the  $\gamma$ -statistic.

Short sequence length has an opposite effect on the  $\gamma$ -statistic to that inflicted by model underparameterization. Short sequence length results in some internodes lacking substitutions entirely. Their length is estimated to be very short regardless of the model of nucleotide substitution used in the analysis. Because there are more branches towards the tips of any phylogenetic tree, this phenomenon results in positively inflated  $\gamma$ , particularly when sequence and tree lengths are very short. However, for the range of sequence lengths used in empirical studies, the magnitude of this positive bias is small relative to the effect of model underparameterization.

Although we restrict our attention to the  $\gamma$ -statistic of Pybus and Harvey (2000), the observation that model underparameterization can lead to severe bias in tree shape extends to other evolutionary inferences that rely on unbiased estimates of phylogenetic branch lengths. For example, comparative methods often assume that the probability of character change on a given branch is proportional to the branch length (e.g., Harvey and Pagel, 1991; Pagel, 1994). Under conditions such as those described above, where oversimplified models are used, such methods will tend to concentrate more change on less severely underestimated, later branches, and less on earlier branches, than if their true lengths were known without error.

## CONCLUSIONS

Our analyses suggest that model underparameterization leads to strong negative bias in the estimation of Pybus and Harvey's  $\gamma$  statistic, which may result in the incorrect inference that the rate of cladogenesis has slowed in the course of a group's history. The observation of a decreasing rate of cladogenesis over time is not unexpected, having both theoretical justification (Walker and Valentine, 1984; Hubbell, 2001) and empirical support from paleontological studies (e.g., Sepkoski, 1978, 1979). However, in some cases, molecular phylogenetic studies that recover this pattern via estimation of the  $\gamma$ -statistic must be viewed with caution. Tests of absolute model

adequacy, such as that described in Bollback (2002), have the potential to alleviate this problem, but are often likely to reject all available models (e.g., Goldman, 1993). Failure in a test of absolute model adequacy is of considerable concern particularly if tree length and sequence length are very long, and if the tree contains many taxa. If the preferred model is found to be inadequate by absolute criteria, any observation of negative  $\gamma$  should be interpreted with caution.

#### ACKNOWLEDGEMENTS

This work was supported in part by a grant from the National Science Foundation (DEB-9982736). We thank J. Losos, K. Kozak, R. B. Langerhans, and S. Heard for comments on the manuscript and members of the Losos lab for much useful discussion. H. P. Linder, R. D. M. Page, D. Posada, and an anonymous reviewer provided insightful criticisms of a previous version of this manuscript.

#### REFERENCES

- Adachi, J., and M. Hasegawa. 1995. Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: Heterogeneity among amino acid sites. *J. Mol. Evol.* 40:622–628.
- Agapow, P.-M., and A. Purvis. 2002. Power of eight tree shape statistics to detect nonrandom diversification: A comparison by simulation of two models of cladogenesis. *Syst. Biol.* 51:866–872.
- Barracough, T. G., and S. Nee. 2001. Phylogenetics and speciation. *Trends Ecol. Evol.* 16:391–399.
- Barracough, T. G., and A. P. Vogler. 2002. Recent diversification rates in North American tiger beetles estimated from a dated mtDNA phylogenetic tree. *Mol. Biol. Evol.* 19:1706–1716.
- Bollback, J. P. 2002. Bayesian model adequacy and choice in phylogenetics. *Mol. Biol. Evol.* 19:1171–1180.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17:368–376.
- Felsenstein, J. 2004. *Inferring phylogenies*. Sinauer Associates, Sunderland, Massachusetts.
- Fukami-Kobayashi, K., and Y. Tatenno. 1991. Robustness of maximum likelihood tree estimation against different patterns of base substitution. *J. Mol. Evol.* 32:79–91.
- Fusco, G., and Q. C. B. Cronk. 1995. A new method for evaluating the shape of large phylogenies. *J. Theor. Biol.* 175:235–243.
- Gaut, B. S., and P. O. Lewis. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.* 12:152–162.
- Gojobori, T., K. Ishii, and M. Nei. 1982. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *J. Mol. Evol.* 18:414–423.
- Goldman, N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* 36:182–198.
- Harmon, L. J., J. A. Schulte II, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in Iguanian lizards. *Science* 301:961–964.
- Harvey, P. H., A. J. Leigh Brown, J. Maynard Smith, and S. Nee eds. 1996. *New uses for new phylogenies*. Oxford University Press, Oxford, UK.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- Hasegawa, M., H. Kishino, and T. Yano. 1987. Man's place in Hominoidea as inferred from molecular clocks of DNA. *J. Mol. Evol.* 26:132–147.
- Håstad, O., and M. Björklund. 1998. Nucleotide substitution models and estimation of phylogeny. *Mol. Biol. Evol.* 15:1381–1389.
- Heard, S. B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution* 46:1818–1826.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, New Jersey.
- Huelsenbeck, J. P., and D. M. Hillis. 1993. Success of phylogenetic methods in the four taxon case. *Syst. Biol.* 42:247–264.
- Huelsenbeck, J. P., and M. Kirkpatrick. 1996. Do phylogenetic methods produce trees with biased shapes? *Evolution* 50:1418–1424.
- Huelsenbeck, J. P., and B. Rannala. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* 57:1237–1247.
- Huelsenbeck, J. P., B. Rannala, and J. P. Masly. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* 288:2349–2350.
- Jin, L., and M. Nei. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* 7:82–102.
- Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pages 21–132 in *Mammalian protein metabolism*, Volume III, (M.N. Munro, ed.). Academic Press, New York.
- Kadereit, J. W., E. M. Griebler, and H. P. Comes. 2004. Quaternary diversification in European alpine plants: Pattern and process. *Philos. T. Roy. Soc. B* 359:265–274.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–120.
- Kozak, K. H., A. Larson, R. M. Bonett, and L. J. Harmon. 2005. Phylogenetic analysis of ecomorphological divergence, community structure, and diversification rates in dusky salamanders (Plethodontidae: *Desmognathus*). *Evolution* 59:2000–2016.
- Kuhner, M. K., and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11:459–468.
- Lemmon, A. R., and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol.* 53:265–277.
- Linder, H. P., P. Eldenäs, and B. G. Briggs. 2003. Contrasting patterns of radiation in African and Australian Restionaceae. *Evolution* 57:2688–2702.
- Lutzoni, F., M. Pagel, and V. Reeb. 2001. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* 411:937–940.
- Machordom, A., and E. Macpherson. 2004. Rapid radiation and cryptic speciation in squat lobsters of the genus *Munida* (Crustacea, Decapoda) and related genera in the South West Pacific: molecular and morphological evidence. *Mol. Phylogenet. Evol.* 33:259–279.
- Martin, A. P., E. K. Costello, A. F. Meyer, D. R. Nemergut, and S. K. Schmidt. 2004. The rate and pattern of cladogenesis in microbes. *Evolution* 58:946–955.
- Mooers, A. Ø., and S. B. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* 72:31–54.
- Mooers, A. Ø., R. D. M. Page, A. Purvis, and P. H. Harvey. 1995. Phylogenetic noise leads to unbalanced cladistic tree reconstructions. *Syst. Biol.* 44:332–342.
- Olsen, G. J. 1991. Systematic underestimation of tree branch lengths by Lake's operator metrics: An effect of position dependent substitution rates. *Mol. Biol. Evol.* 8:592–608.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Posada, D., and K. A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Purvis, A., A. Katzourakis, and P.-M. Agapow. 2002. Evaluating phylogenetic tree shape: Two modifications to Fusco & Cronk's method. *J. Theor. Biol.* 214:99–103.
- Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. Lond. B* 267:2267–2272.
- Pybus, O. G., A. Rambaut, E. C. Holmes, and P. H. Harvey. 2002. New inferences from tree shape: Numbers of missing taxa and population growth rates. *Syst. Biol.* 51:881–888.
- Rambaut, A. 2002. Phyl-O-Gen: phylogenetic tree simulator package, v1.1. See <http://evolve.zoo.ox.ac.uk/software.html?id=phylogen>.
- Rambaut, A., and N. C. Grassly. 1997. Seq-gen: An application for Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13:235–238.
- Rodríguez, F., J.L. Oliver, A. Marín, and J.R. Medina. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142: 485–501.
- Ronquist, F. 2004. Bayesian inference of character evolution. *Trends Ecol. Evol.* 19:475–481.

- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Rüber, L., and R. Zardoya. 2005. Rapid cladogenesis in marine fishes revisited. *Evolution* 59:1119–1127.
- Ruvolo, M., S. Zehr, M. von Dornum, D. Pan, B. Chang, and J. Lin. 1993. Mitochondrial COII sequences and modern human origins. *Mol. Biol. Evol.* 10:1115–1135.
- Sepkoski, J. J., Jr. 1978. A kinetic model of Phanerozoic taxonomic diversity I. Analysis of marine orders. *Paleobiology* 2:223–251.
- Sepkoski, J. J., Jr. 1979. A kinetic model of Phanerozoic taxonomic diversity II. Early Phanerozoic families and multiple equilibria. *Paleobiology* 3:222–251.
- Shaw, A. J., C. J. Cox, B. Goffinet, W. R. Buck, and S. B. Boles. 2003. Phylogenetic evidence of a rapid radiation of pleurocarpous mosses (Bryophyta). *Evolution* 57:2226–2241.
- Swofford, D. L. 1998. PAUP\*: Phylogenetic analysis using parsimony (\*and other methods). Sinauer Associates, Sunderland, Massachusetts.
- Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10:512–526.
- Townsend, T. M., A. Larson, E. Louis, and J. R. Macey. 2004. Molecular phylogenetics of Squamata: The position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. *Syst. Biol.* 53:735–757.
- Uzzell, T., and K. W. Corbin. 1971. Fitting discrete probability distributions to evolutionary events. *Science* 172:1089–1096.
- Walker, T. D., and J. W. Valentine. 1984. Equilibrium models of the evolutionary species diversity and the number of empty niches. *Am. Nat.* 124:887–899.
- Welch, J. J., and L. Bromham. 2005. Molecular dating when rates vary. *Trends Ecol. Evol.* 20:320–327.
- Williams, S. T., and D. G. Reid. 2004. Speciation and diversity on tropical rocky shores: A global phylogeny of snails of the genus *Echinolittorina*. *Evolution* 58:2227–2251.
- Yang, Z., N. Goldman, and A. Friday. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* 11:316–324.
- Yang, Z., N. Goldman, and A. Friday. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Syst. Biol.* 44:384–399.
- Zhang, L. B., H. P. Comes, and J. W. Kadereit. 2004. The temporal course of quaternary diversification in the European high mountain endemic *Primula* sect. *Auricula* (Primulaceae). *Int. J. Plant Sci.* 165:191–207.

First submitted 5 November 2004; reviews returned 4 March 2005;

final acceptance 24 May 2005

Associate Editor: Peter Linder