New Inferences from Tree Shape: Numbers of Missing Taxa and Population Growth Rates

O. G. Pybus, A. Rambaut, E. C. Holmes, and P. H. Harvey

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

Abstract.—The relative positions of branching events in a phylogeny contain information about evolutionary and population dynamic processes. We provide new summary statistics of branching event times and describe how these statistics can be used to infer rates of species diversification from interspecies trees or rates of population growth from intraspecies trees. We also introduce a phylogenetic method for estimating the level of taxon sampling in a clade. Different evolutionary models and different sampling regimes can produce similar patterns of branching events, so it is important to consider explicitly the model assumptions involved when making evolutionary inferences. Results of an analysis of the phylogeny of the mosquito-borne flaviviruses suggest that there could be several thousand currently unidentified viruses in this clade. [Coalescent; extinction; flavivirus; macroevolution; speciation; taxon sampling.]

It is no exaggeration to say that molecular phylogenetics has transformed the study of evolutionary biology. It has provided a new statistical framework for estimating historical relationships among organisms, and it has supplied the raw data to test models of evolutionary and population genetic processes. As a result, conceptual modeling has given way to hypothesis testing and parameter estimation. The range of evolutionary questions investigated using molecular phylogenies has grown rapidly (see Harvey et al., 1996). Examples include comparing speciation and extinction rates among clades (Sanderson and Donoghue, 1994; Purvis et al., 1995), searching for traits that correlate with rates of species diversification (Barraclough et al., 1998), correlating phylogeny with biogeographic events (Avise, 2000), and inferring the demographic history of human or microbial populations (Slatkin and Hudson, 1991; Holmes et al., 1995).

The information contained in molecular phylogenies can be crudely partitioned into two components: the topology or branching structure of the phylogeny and the node heights or branch lengths of the tree. Although these components are intimately linked, the distinction is nevertheless useful (Mooers and Heard, 1997). Some tests of evolutionary hypotheses require only a tree topology, such as methods that compare the diversification rates of sister clades (e.g., Slowinski and Guyer, 1989) or statistics that measure tree balance (e.g., Kirkpatrick and Slatkin, 1993). In general, such tests are used to detect variation in evolutionary processes among different parts of the tree, hence

the importance of the tree topology. However, other tests, particularly those based on the birth–death or coalescent processes (e.g., Slatkin and Hudson, 1991; Nee et al., 1992, 1994a), depend only on node height information. The common feature of these tests is that they can be used to estimate the rate (or change of rate through time) of evolutionary processes that are homogenous across the tree. Both the birth–death and coalescent models are independent of tree topology because they exhibit a statistical property known as exchangeability. In evolutionary terms, this means that at any point in time all lineages in the tree are assumed to behave according to the same stochastic rules, although these rules can change through time. Depending on the model, the lineages may represent species, individuals, or genes.

Here, we discuss how phylogenetic node height information can be summarized and used to test evolutionary hypotheses. We have developed a new statistical method that uses node heights to estimate the true number of taxa in a clade when only a proportion of the taxa has been sampled, and we have applied this method to estimate the number of viruses in the genus *Flavivirus*. In ultrametric phylogenies (Fig. 1), the branch lengths are proportional to time and all root-to-tip distances are equal, hence internal nodes represent relative divergence times. The position of an internal node in an ultrametric tree can be measured either as distance from the tip (node height) or as distance between successive nodes (internode interval). The choice is arbitrary, but here we used the latter measure. Until recently,
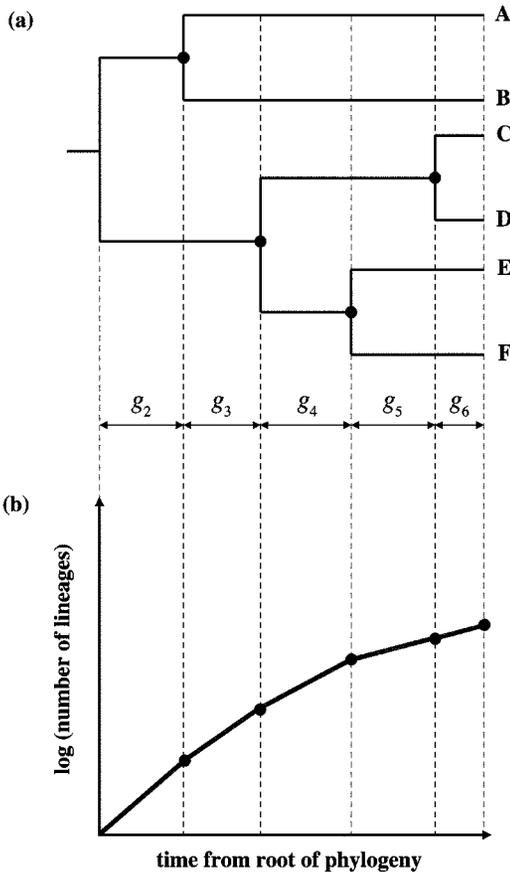
FIGURE 1. The use of ultrametric phylogenies. (a) An ultrametric phylogeny with six extant taxa, A–F. Branch lengths are proportional to time, and the solid circles show the position of the internal nodes. The values g_2, $g_3, \ldots, g_n$ are the internode interval sizes; the subscripts refer to the number of lineages present in the tree during each interval. (b) The lineage-through-time plot of the phylogeny. The plot displays the log of the number of lineages in the tree against time and is slightly convex in this example.

ultrametric phylogenies could only be obtained by assuming a constant rate of molecular evolution, i.e., a molecular clock. Fortunately, the strict molecular clock is no longer compulsory because there are now several methods that can estimate divergence times when the rate of evolution varies across the tree (Sanderson, 1997; Thorne et al., 1998; Huelsenbeck et al., 2000).

When inferring evolutionary processes from molecular phylogenies, the sampled gene sequences constitute the data for analysis, whereas the reconstructed phylogeny is an estimate. Unless the true phylogeny is known, our evolutionary inference will always carry an error due to the uncertainty

in tree reconstruction. We ignore this error here and focus instead on developing statistical tests under various evolutionary models. Our methods could be extended to incorporate phylogenetic error by integrating over a weighted sample of all possible trees (Griffiths and Tavare, 1994; Kuhner et al., 1998; Huelsenbeck et al., 2001), a very interesting area of current and future research.

First, we review how node height information can be used to investigate hypotheses of species diversification under the birth–death model (Raup et al., 1973). This model represents speciation (birth) and extinction (death) as instantaneous stochastic events. During a short time interval Δt, each extant lineage speciates with probability $\Delta b(t)$ and goes extinct with probability $\Delta d(t)$; thus, $b(t)$ and $d(t)$ represent per-lineage rates. Special cases of the general birth–death model can also be defined: (1) the constant-rates model, which arises when b and d are constant both among lineages and through time, and (2) the pure-birth model, which arises when b is constant and d is zero. Nee et al. (1994b) suggested that the constant-rates model is an appropriate null model for species diversification, and they developed a maximum likelihood framework to infer b and d from molecular phylogenies of extant taxa. Estimation of these rates informs us of the tempo of macroevolution; furthermore, rejection of the constant-rates model may indicate evolutionary events such as density-dependent speciation, adaptive radiation, mass extinction, or the appearance of key innovations (see Schluter, 2000).

The relationship between node height and speciation and extinction rate is well known. As the extinction rate d increases, the internal nodes of a phylogeny occur increasingly closer to the tips because lineages that arose in the distant past are more likely to be removed by extinction than are lineages that arose recently (Nee et al., 1994a). A phylogeny's nodes also occur closer to the present if the speciation rate $b(t)$ increases through time. Similarly, nodes occur closer to the root if $b(t)$ decreases through time. These observations were initially obtained through the use of lineage-through-time plots, which display the number of lineages in a reconstructed phylogeny against time (Fig. 1; Nee et al., 1994a, 1994b; Rambaut et al., 1997). To quantify the relative positions of internal nodes in a phylogeny we previously

presented the following statistic (Pybus and Harvey, 2000).

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

$$T = \left(\sum_{j=2}^{n}jg_j\right), \tag{1}$$

where n is the number of taxa and the values $g_2, g_3, \ldots, g_n$ are the internode intervals of the tree, ordered from the root (see Fig. 1). There are k lineages during interval g_k. Equation 1 is derived from similar statistics of Cox and Lewis (1966) and Zink and Slowinski (1995) and can be considered a numerical summary of the shape of a lineage-through-time plot. Under the pure-birth process, the γ values of reconstructed phylogenies tend rapidly to a standard normal distribution as n increases (Cox and Lewis, 1966). Trees with many nodes close to the tips tend to generate concave lineage-through-time plots and have $\gamma > 0$, whereas trees with many nodes close to the root tend to generate convex plots and have $\gamma < 0$. The pure-birth model can be rejected at the 95% level if $\gamma < -1.96$ or $\gamma > 1.96$, and the constant-rates model can be similarly rejected if $\gamma < -1.645$ (one-tailed test). In practice, γ can only reject the constant-rates model if the speciation rate has decreased through time (Pybus and Harvey, 2000). The γ statistic is a measure of relative node heights and is therefore independent of the nucleotide substitution rate of the sequences used to estimate the phylogeny.

In addition to testing models of speciation and extinction, the γ statistic can also be used to investigate the level of taxon sampling in a clade. Suppose that a clade has diversified according to a pure-birth process, but our reconstructed phylogeny only contains a fraction of the extant taxa. Sampling fewer taxa leads to lower γ values, so long as the sampled taxa are chosen randomly (Fig. 2). Nodes near the root of the tree leave more extant descendents than do nodes near the tips and are therefore more likely to be included in a small random sample (Nee et al., 1994a). The γ-based tests can be corrected for random incomplete sampling using Monte Carlo simulation, provided that the true number of species in the sampled clade is known (Pybus and Harvey, 2000).
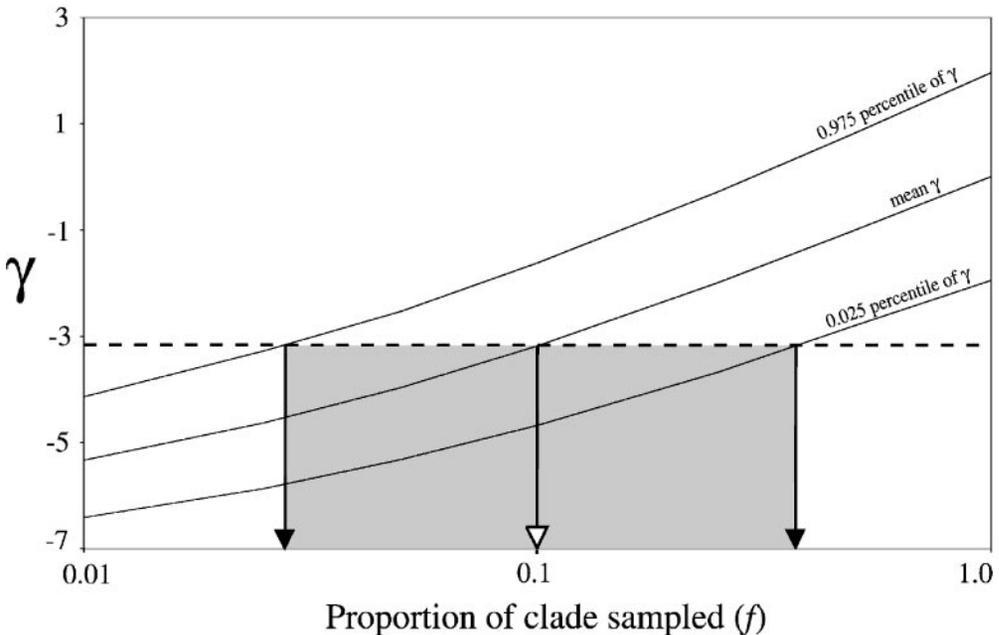

FIGURE 2. The relationship between γ and the proportion of extant taxa in a clade that have been sampled (f). There were 30 taxa in the incompletely sampled phylogeny. The three curves show the mean, 0.975 percentile, and 0.025 percentile of the γ distribution. The dotted line represents an observed γ value, and the shaded area represents the range of f values that are consistent with the observed γ at the 95% level. The open arrow points to the estimate of f, and the closed arrows point to the upper and lower confidence limits of this estimate.

The relationship between the γ statistic and the fraction (f) of sampled taxa raises an interesting question: can we use γ to estimate the true size of a clade? Although the probability density function of γ is not known, we can still estimate clade size using Monte Carlo simulation. Figure 2 was obtained by simulating pure-birth trees under different values of f and plotting the distribution of the resulting γ values. Thus, we can estimate and provide confidence intervals for f by measuring where our observed γ value intersects the simulated mean and 95% percentile curves (see Fig. 2). This is an approximate method of moments estimate. The curves shown in Figure 2 are only valid for a particular sample size but can be re-simulated for any given sample size. Alternatively, we could use an approximate likelihood method. For each value of f, we count how many simulated trees have γ values that are arbitrarily close to our observed γ. Plotting these counts against f provides an approximate likelihood curve (see Weiss and von Haeseler, 1998, for a mathematical definition of this approach). Both of these estimation methods are illustrated below in different contexts. We also discuss the consequences of assuming that the investigated clade has diversified according to a pure-birth process.

To illustrate the estimation of clade size, we analyzed viral gene sequences from the genus *Flavivirus* (family Flaviviridae). This group of single-stranded positive-sense RNA viruses contains many species that cause disease in humans, including yellow fever virus, West Nile virus, and dengue virus. An estimate of the true number of flaviviruses may help to quantify the risk of future zoonotic transmissions and subsequent emerging epidemics in human populations. To obtain a flavivirus phylogeny we used the data published by Kuno et al. (1998), comprising NS5 gene sequences from almost all known flavivirus species. The genus is split into three reciprocally monophyletic clades that are named after (and associated with) different transmission vectors: mosquito borne, tick borne, and no known vector. We chose the mosquito-borne clade for analysis because it is the largest data set and has the most robust classification. The sequences were obtained from GenBank and aligned by hand using the program SeAl (http://evolve.zoo.ox.ac.uk). A maximum-

likelihood phylogeny was estimated using PAUP* (Swofford, 2000) under the Hasegawa–Kishino–Yano (Hasegawa et al., 1985) substitution model and a codon-position model of rate heterogeneity among sites. There was no evidence for variation in diversification rates among lineages in this tree using the B_1 tree balance statistic ($B_1 = 22.26$, $P > 0.05$; Kirkpatrick and Slatkin, 1993). The molecular clock was rejected using a likelihood ratio test ($P > 0.05$; Felsenstein, 1981), so a second phylogeny was obtained using Sanderson's (1997) non-parametric rate smoothing (NPRS) method, which takes among-lineages rate variation into account when estimating node heights. The two trees were similar, and their corresponding γ values were used to estimate the true number of flaviviruses, using the method of moments approach (Fig. 3). However, the statistical properties of the NRPS method are not well known, so we cannot exclude the possibility that it may bias γ values.

Both phylogenies suggest that the number of unsampled taxa is approximately 2,000. This large number is wholly feasible in the context of the viral-vector biology involved. Horizontal transmission requires viral infection and replication in the mosquito midgut epithelium and salivary gland, and the ability of a mosquito to act as a vector for a specific flavivirus depends primarily on the genetic susceptibility of its midgut lining. Thus, flaviviruses exhibit a high degree of vector specificity and in some cases are only capable of infecting particular strains of vector species (Monath and Heinz, 1996). There are >3,000 mosquito species worldwide (Walter Reed Biosystematics Unit, 2001), many of which harbor multiple flaviviruses, so there may be several thousand unsampled mosquito-borne flaviviruses. Unfortunately, we have no way of predicting how many of these viruses have the potential to infect humans.

The analysis above assumes a pure-birth process. What if this assumption is invalid? Both background extinction and increasing speciation rates through time produce an increase in γ and will result in an underestimation of clade size. Therefore, even if these factors are present we can still interpret our estimate as a lower bound on clade size. However, a decreasing speciation rate through time reduces γ and could provide an alternative explanation to incomplete

$\gamma = -5.772$

$$\text{Estimated number of taxa in clade} = 2,500 \ (825, 7900)$$

$\gamma = -5.435$

$$\text{Estimated number of taxa in clade} = 1,790 \ (625, 5500)$$
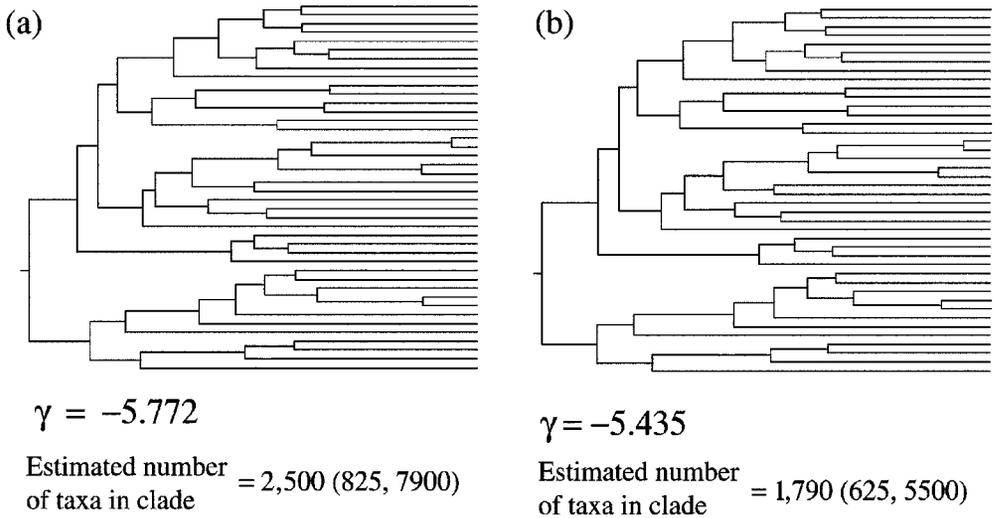
FIGURE 3. The estimated phylogenies for mosquito-borne flaviviruses. The γ values and corresponding estimates of clade size (with 95% confidence limits) are also shown. (a) Branch lengths estimated using the molecular clock. (b) Branch lengths estimated using nonparametric rate smoothing.

sampling for the observed tree shape. Although we cannot rule out this hypothesis, decreasing speciation rates are often ascribed to density-dependent saturation of available niches (e.g., Purvis et al., 1995; Zink and Slowinski, 1995), and the above discussion of vector biology suggests that this scenario is unlikely for the mosquito-borne flaviviruses.

In the absence of mechanisms that generate reproductive isolation, the concept of a viral "species" is difficult to define. In some clades there is little distinction between within- and among- "species" genetic distances (e.g., the flavivirus tick-borne encephalitis complex) and a corresponding tendency to use arbitrary values of sequence divergence to distinguish between viral subtypes, strains, and species. If this value is increased, then nodes near the tips are removed and the corresponding estimate of clade size will be larger. Thus, the clade size estimated represents the number of "species"-level viral lineages in the clade, however such lineages are defined.

To study trees that represent the ancestry of individuals belonging to a single population, a different stochastic model, the coalescent process, is more appropriate (Kingman, 1982). The nodes heights of such genealogies depend on the dynamics of population size through time (Griffiths and Tavaré, 1994), and incomplete sampling is no longer an issue because the coalescent model implic-

itly assumes that the genealogy represents a small random sample from a large population. The coalescent model has proven particularly useful in investigating the demographic history of human populations (Slatkin and Hudson, 1991; Weiss and von Haeseler, 1998) and the epidemic history of viral pathogens (Holmes et al., 1995; Pybus et al., 2001).

Because both coalescent and reconstructed birth–death trees can be generated by nonhomogenous Markov processes, we developed a new statistic equivalent to γ for the coalescent process:

$$\delta = \frac{\left(\frac{T}{2}\right) - \left[\frac{1}{n-2} \sum_{i=n}^{3} \left(\sum_{k=n}^{i} \frac{k(k-1)g_k}{2} \right)\right]}{T\sqrt{\frac{1}{12(n-2)}}},$$

$$T = \left(\sum_{j=2}^{n} \frac{j(j-1)g_j}{2} \right). \tag{2}$$

As before, n is the number of tips in the genealogy, and $g_2, g_3, \ldots, g_n$ are the internode intervals, ordered from the root (see Fig. 1). Equation 2 was derived by applying the same rationale used to derive Equation 1: The observed internode intervals g_k are transformed such that they follow an exponential distribution under the appropriate null hypothesis. In Equation 1, the intervals are multiplied by k, whereas in Equation 2 the

factor is $[k(k-1)]/2$. To avoid bias, the external sum in the numerator of Equation 2 stops at $i = 3$.

Genealogies with many nodes close to the tips tend to have $\delta > 0$, whereas trees with many nodes close to the root have $\delta < 0$. Under the null hypothesis of constant popula-

tion size, δ tends rapidly to a standard normal distribution, so this hypothesis can be rejected at the 95% level if $\delta < -1.96$ or $\delta > 1.96$. Exponential population growth results in starlike trees with long terminal branches that increase in length as the product of exponential growth rate and current population
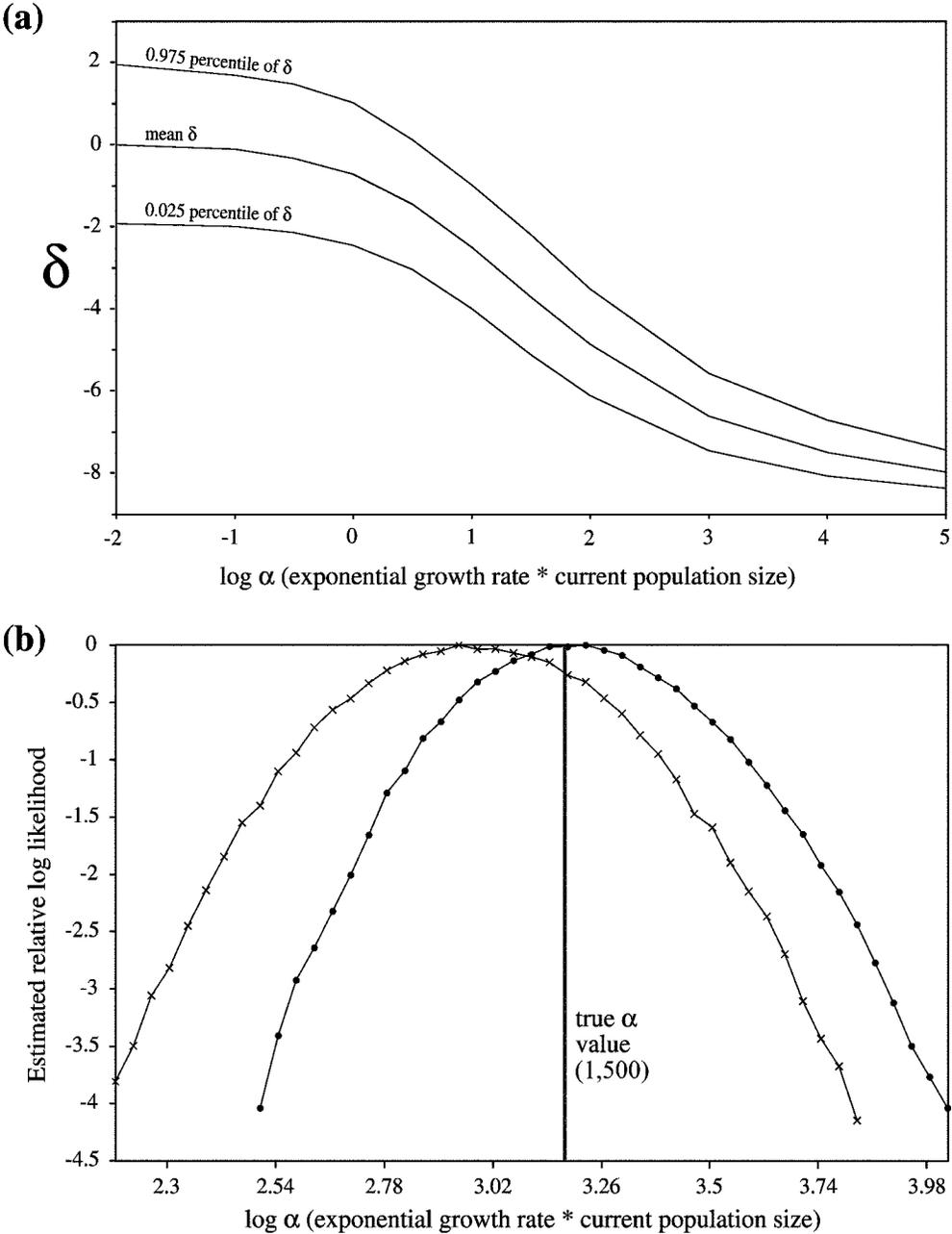

(a)

δ

0.975 percentile of δ

mean δ

0.025 percentile of δ

log α (exponential growth rate * current population size)

(b)

Estimated relative log likelihood

true α value (1,500)

log α (exponential growth rate * current population size)

FIGURE 4.   Performance of the δ statistic. (a) The relationship between δ and α for genealogies with 30 tips. The three curves show the mean, 0.975 percentile, and 0.025 percentile of the δ distribution. As α tends to zero (constant population size), the distribution of δ tends to the standard normal distribution. (b) Estimated relative log likelihood curves for the δ ($\bullet$) and σ ($\times$) statistics. The vertical line indicates the true value of α.

size (denoted α) increases (Slatkin and Hudson, 1991). Thus, δ decreases as α increases, and δ could be used to estimate α (Fig. 4a). A similar measure of relative node heights, the mid-depth statistic, has also been used to estimate α (Pybus et al., 1999), but there are two reasons to believe that δ will be more powerful. First, the mid-depth statistic (denoted σ) is defined as the number of internal nodes in an ultrametric tree that are closer to the root than to the tips. Thus, σ can only take integer values in the range $[0, n - 2]$ and therefore will be insensitive to small changes in α. Furthermore, the distribution of σ is poorly defined when its mean is close to 0 or $n - 2$. Second, δ is expected to be an optimal statistic under a wide range of conditions (Cox and Lewis, 1966).

To investigate the relative performance of δ and σ, we applied both statistics to a simulated coalescent tree with $\alpha = 1,500$. Likelihood curves for these statistics were estimated using the Monte Carlo simulation approach described above. As shown in Figure 4b, δ appears unbiased, whereas σ does not, and δ has a slightly lower variance (Fisher's information is -14.5 for δ and -11.5 for σ). Both δ and the mid-depth method make the same assumptions about the sampled population: no selection, no population subdivision, and no recombination within the gene region used.

Full likelihood and Bayesian frameworks for inference under the coalescent model are already available (Griffiths and Tavare, 1994; Kuhner et al., 1998; Drummond et al., 2002). These methods can be used to estimate the growth rate and population size as separate parameters that are (typically) functions of the nucleotide substitution rate. Although δ is statistically weaker than a full likelihood approach, it is independent of the substitution rate and therefore may be useful when this rate is unknown. In particular, if multiple independent loci have been sequenced from a sample of individuals, then each locus could have a different and unknown evolutionary rate, making the loci difficult to compare. The distribution of δ among loci might provide a useful estimator of population size and growth rate in such cases.

Where should macroevolutionary models of phylogenetic tree shape go next? The obvious next step is to reintegrate node height and topology information into a single framework that can be used to simultaneously infer both time-dependent and lineage-specific diversification rates. To achieve this integration, we must abandon the property of statistical exchangeability; however, experience from coalescent theory suggests that this will not be an easy step. Incorporating positive selection into the coalescent model breaks the assumption of exchangeability, resulting in complex models and little generality (Kaplan et al., 1988; Krone and Neuhauser, 1997). Reconstructed phylogenies may never contain enough information to make useful inferences under very general models of species diversification, so we must investigate Bayesian frameworks that do not assume a single phylogeny and can incorporate prior information about node heights from palaeontological and biogeographic sources. On a more positive note, genomic-level sequences will provide a wealth of data and enable us to address new evolutionary problems. In particular, birth–death models could be used to investigate the evolutionary history of paralogous gene families that evolve by gene duplication and gene excision or inactivation (Sanderson, 1994).

REFERENCES

AVISE, J. C. 2000. Phylogeography. The history and formation of species. Harvard Univ. Press, Cambridge, Massachusetts.

BARRACLOUGH, T. G., A. P. VOGLER, AND P. H. HARVEY. 1998. Revealing the factors that promote speciation. Philos. Trans. R. Soc. Lond. B 353:241–249.

COX, D. R., AND P. A. W. LEWIS. 1966. The statistical analysis of series of events. Methuen, London.

DRUMMOND, A. J., G. K. NICHOLLS, A. G. RODRIGO, AND W. SOLOMON. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161:1307–1320.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

GRIFFITHS, R. C., AND S. TAVARÉ. 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. B 344:403–410.

HARVEY, P. H., A. J. LEIGH BROWN, J. MAYNARD SMITH, AND S. NEE (eds.). 1996. New uses for new phylogenies. Oxford Univ. Press, Oxford, U.K.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating the human–ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

HOLMES, E. C., S. NEE, A. RAMBAUT, G. P. GARNETT, AND P. H. HARVEY. 1995. Revealing the history of infectious

disease epidemics through phylogenetic trees. Philos. Trans. R. Soc. Lond. B. 349:33–40.

HUELSENBECK, J. P., B. LARGET, AND D. SWOFFORD. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154:1879–1892.

HUELSENBECK, J. P., F. RONQUIST, R. NIELSEN, AND J. P. BOLLBACK. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

KAPLAN, N. L., T. DARDEN, AND R. R. HUDSON. 1988. The coalescent process in models with selection. Genetics 120:819–829.

KINGMAN, J. F. C. 1982. On the genealogy of large populations. J. Appl. Probab. 19A:27–43.

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

KRONE, S. M., AND C. NEUHAUSER. 1997. Ancestral processes with selection. Theor. Popul. Biol. 51:210–237.

KUHNER, M. K., J. YAMATO, AND J. FELSENSTEIN. 1998. Maximum likelihood estimation of growth rates based on the coalescent. Genetics 149:429–434.

KUNO, G., J. GWONG-JEN, K. CHANG, R. TSUCHIYA, N. KARABATSOS, AND C. B. CROPP. 1998. Phylogeny of the genus Flavivirus. J. Virol. 72:73–83.

MONATH, T. P., AND F. X. HEINZ. 1996. Flaviviruses. Pages 961–1034 in Fields virology, 3rd edition (B. N. Fields, D. M. Knipe, and P. M. Howley, eds.). Lippincott-Raven, Philadelphia.

MOOERS, A. Ø., AND S. B. HEARD. 1997. Inferring evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72:31–54.

NEE, S., E. C. HOMES, R. M. MAY, AND P. H. HARVEY. 1994a. Extinction rates can be estimated from molecular phylogenies. Philos. Trans. R. Soc. Lond. B 344:77–82.

NEE, S., R. M. MAY, AND P. H. HARVEY. 1994b. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B 344:305–311.

NEE, S., A. Ø. MOOERS, AND P. H. HARVEY. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. USA 89:8322–8326.

PURVIS, A., S. NEE, AND P. H. HARVEY. 1995. Macroevolutionary inferences from primate phylogeny. Proc. R. Soc. Lond. B 260:329–333.

PYBUS, O. G., M. A. CHARLESTON, S. GUPTA, A. RAMBAUT, E. C. HOLMES, AND P. H. HARVEY. 2001. The epidemic behaviour of the hepatitis C virus. Science 292:2323–2325.

PYBUS, O. G., AND P. H. HARVEY. 2000. Testing macroevolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. B 267:2267–2272.

PYBUS, O. G., E. C. HOLMES, AND P. H. HARVEY. 1999. The mid-depth method and HIV-1: A practical approach for testing hypotheses of viral epidemic history. Mol. Biol. Evol. 16:953–959.

RAMBAUT, A., P. H. HARVEY, AND S. NEE. 1997. End-Epi: An application for inferring phylogenetic and population dynamical processes from molecular sequences. Comput. Appl. Biosci. 13:303–306.

RAUP, D. M., S. J. GOULD, T. J. M. SCHOPF, AND D. SIMBERLOFF. 1973. Stochastic models of phylogeny and the evolution of diversity. J. Geol. 81:525–542.

SANDERSON, M. J. 1994. Reconstructing the history of evolutionary processes using maximum likelihood. Pages 13–26 in Molecular evolution of physiological processes (D. M. Fambrough, ed.). Rockerfeller Univ. Press, New York.

SANDERSON, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1232.

SANDERSON, M. J., AND M. J. DONOGHUE. 1994. Shifts in diversification rate with the origin of the angiosperms. Science 264:1590–1593.

SCHLUTER, D. 2000. The ecology of adaptive radiation. Oxford Univ. Press, Oxford, U.K.

SLATKIN, M., AND R. R. HUDSON. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555–562.

SLOWINSKI, J., AND C. GUYER. 1989. Testing the stochasticity of patterns of organismal diversity: An improved null model. Am. Nat. 134:907–921.

SWOFFORD, D. L. 2000. PAUP,* version 4: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, Massachusetts.

THORNE, J. L., H. KISHINO, AND I. S. PAINTER. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Walter Reed Biosystematics Unit. 2001. 2001 Systematic Catalog of Culicidae. Available from http://wrbi.si.edu.

WEISS, G., AND A. VON HAESELER. 1998. Inference of population history using a likelihood approach. Genetics 149:1539–1546.

ZINK, R. M., AND J. B. SLOWINSKI. 1995 Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. Proc. Natl. Acad. Sci. USA 92:5832–5835.