INFERRING THE RATES OF BRANCHING AND EXTINCTION FROM MOLECULAR PHYLOGENIES

Takuya Kubo and Yoh Iwasa
Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-81, Japan
E-mail: tkuboscb@mbox.nc.kyushu-u.ac.jp; yiwasscb@mbox.nc.kyushu-u.ac.jp

Abstract.—Molecular techniques provide ancestral phylogenies of extant taxa with estimated branching times. Here we studied the pattern of ancestral phylogeny of extant taxa produced by branching (or cladogenesis) and extinction of taxa, assuming branching processes with time-dependent rates. (1) If the branching rate b and extinction rate c are constant, the semilog plot of the number of ancestral lineages over time is not a straight line but is curvilinear, with increasing slope toward the end, implying that ancestral phylogeny shows apparent increase in the branching rate near the present. The estimate of b and c based on nonlinear fitting is examined by computer simulation. The estimate of branching rate can be usable for a large phylogeny if b is greater than c, but the estimate of extinction rate c is unreliable because of large bias and variance. (2) Gradual decrease in the slope of the semilog plot of the number of ancestral lineages over time, as was observed in a phylogeny of bird families based on DNA hybridization data, can be explained equally well by either the decreasing branching rate or the increasing extinction rate. Infinitely many pairs of branching and extinction rates as functions of time can produce the same ancestral phylogeny. (3) An explosive branching event in the past would appear as a quick increase in the number of ancestral lineages. In contrast, mass extinction occurring in a brief period, if not accompanied by an increase in branching rate, does not produce any rapid change in the number of ancestral lineages at the time. (4) The condition in which the number of ancestral lineages of extant species changes in parallel with the actual number of species in the past is derived.

Key words.—Ancestral phylogeny, cladogenesis, explosive branching, extinction, mass extinction, speciation.

Molecular techniques allow us to estimate the ancestral phylogeny of extant taxa, together with the time of each branching point. The pattern of phylogeny becomes more accurate as more data are accumulated. The techniques to reconstruct the molecular phylogeny from a given data set of DNA sequences have been studied extensively (Nei 1987). The molecular evolution data can estimate not only the topological pattern of phylogeny but also the timing of branching points in the past, based on the "molecular clock" hypothesis (Zuckerkandl and Pauling 1962) that is not available for morphological characters.

Harvey et al. (1991) first suggested the possibility of extracting information about evolutionary processes (cladogenesis) from reconstructed ancestral phylogenies. They subsequently developed a method of analyzing such processes by using the number of ancestral lineages of extant taxa at different times (see fig. 1) and applied the idea to data from 122 bird families taken from Sibley and Ahlquist (1990) based on DNA-DNA hybridization technique (Nee et al. 1992). The number of ancestral lineages thus reconstructed must always become smaller as time goes back from the present, and hence, the number of lineages always appears to increase with time. The extinction of taxa in the past is not observable. Nee et al. (1992) plotted the logarithm of the number of ancestral lineages over time, and from the shape of the graph they attempted to infer past evolutionary events, such as the decrease in the "effective rate of cladogenesis," presumably owing to interspecific competition accompanied by an increase in the number of bird families. Although there are potential shortcomings inherent in the estimate (as discussed by Nee et al. 1992), the method deserves more careful examination, especially because more and more molecular phylogenies for various taxa will become available in coming years.

Here we first study the relationship between the shape of ancestral phylogenies and the branching and extinction rates of taxa given as functions of time. Our analysis is based on branching processes with time-dependent rates, in which both branching and extinction happen randomly, and all the existing taxa are independent of each other.

The pattern in which the number of ancestral lineages changes over time depends upon both branching rate and extinction rate as functions of time. When both branching and extinction rates are constant, the number of lineages in reconstructed ancestral phylogenies appears to increase at an accelerating rate in the final step of evolution, because the fraction of observable lineages among existing lineages increases near the present. (This result was also obtained by Harvey et al. [1994a]). We examine the usefulness and limitation of estimating branching and extinction rates by nonlinear fitting of the number of ancestral lineages over some period to a theoretical formula.

Second, we show that vastly different pairs of time-dependent branching and extinction rates can generate exactly the same pattern of the number of ancestral lineages over time. Third, we examine the effect of episodic events, such as a brief period of explosive branching, or a short period of high extinction rate. Mass extinction events turn out to be much more difficult to recognize from the ancestral phylogeny pattern than explosive branching events. We also discuss how to relate the time-dependent branching processes discussed here and the density-dependent time change of branching and extinction rates caused by the increase in the number of taxa discussed by Nee et al. (1992) and the relationship of our results to other studies in this area (Harvey et al. 1994a,b; Nee et al. 1994a,b).

MODEL

The model assumes that new lineages are randomly created by branching from existing lineages and that lineages go
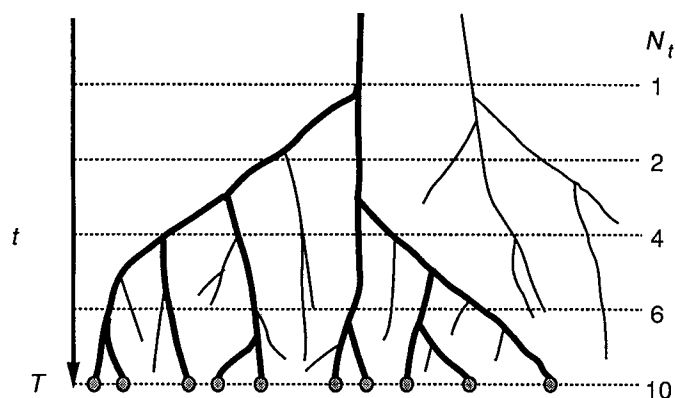
FIG. 1. Observable and unobservable lineages of a phylogeny. Thick branches are for lineages that are observable from the analysis of extant species (indicated by small circles). Thin branches are not observable, because they do not have any extant descendant. The number of ancestral lineages of extant taxa at various times is indicated to the right.

extinct randomly. These rates may vary with time t, where $-\infty < t \leq T$, with t indicating the present time. Let $b(t)$ be the rate of branching and $c(t)$ be the rate of extinction. After a short time interval of length Δt, any taxa existing at time t become two with probability $b(t)\Delta t$, become extinct with probability $c(t)\Delta t$, or remain a single taxon with probability $1 - b(t)\Delta t - c(t)\Delta t$. Branching and extinction events of a particular taxon are assumed independent of the behavior of other taxa. It is also assumed that there is no age effect, such that a newly born taxon has exactly the same probability of extinction and branching as taxa created a long time before. This mathematical model is a special case of time-dependent branching processes (Harris 1963, Karlin and Taylor 1975). Kendall (1948) refers to this as the "generalized birth-and-death processes" model.

Let P_k be the probability that a single taxon randomly chosen at time t will have k descendent taxa at time T. According to the calculation in Appendix A, these probabilities are:

$$P_0 = \frac{1 + \alpha - \beta}{\alpha(1 + \alpha)},$$

$$P_k = \frac{\beta}{\alpha(1 + \alpha)}\left(\frac{\alpha}{1 + \alpha}\right)^k, \quad (k = 1, 2, 3, \ldots), \quad (1a)$$

where

$$\alpha = \int_t^T b(u)\exp\left\{\int_u^T [b(z) - c(z)] \, dz\right\} du, \quad (1b)$$

$$\beta = \exp\left\{\int_t^T [b(u) - c(u)] \, du\right\}, \quad (1c)$$

which indicates that the distribution of the number of descendant lineages at T has a geometric distribution, except for the zero term P_0 (for the robustness of this result, see Nee et al. [1994a]). The common ratio of the geometric series is $\alpha/(1 + \alpha)$, with α given by equation (1B).

Let N_t be the total number of ancestral lineages for the

extant taxa we consider. Because of the nature of observability, each of the N_t lineages has at least one extant descendant taxon. Each observable lineage either increases the number of descendants or remains one, and no extinction events are apparent for the reconstructed ancestral phylogeny. N_t must constantly increases with time. Now we consider the ratio of N_T, the number of extant taxa, to N_t, the number of their ancestral lineages in generation t. This ratio is the same as the average rate of lineage multiplication from time t to time T calculated over only observable lineages:

$$\frac{N_T}{N_t} = \frac{\sum_{k=1}^{\infty} kP_k}{\sum_{k=1}^{\infty} P_k} = 1 + \alpha, \quad (2)$$

where the average is taken over $k \geq 1$ (excluding the term of $k = 0$). Using equation (1), we have,

$$\frac{N_t}{N_T} = \frac{1}{1 + \int_t^T b(u)\exp\left\{\int_u^T [b(z) - c(z)] \, dz\right\} du}. \quad (3)$$

This is the basic equation upon which all the analyses presented here are based. Harvey et al. (1994a) derived the same result based on Kendall (1948). The number of ancestral lineages N_t in equation (3) always increases with time t, and the way it increases depends on the branching rate $b(t)$ and extinction rate $c(t)$ as functions of time.

We may ask how much information about branching and extinction rates can be obtained from the pattern of the number of lineages using equation (3). To determine this, we begin with several simplified cases.

Case 1: No Extinction of Taxa

Suppose that no extinction occurs [$c(t) = 0$]. Then by computing the integral in the denominator, equation (3) becomes simply

$$\frac{N_t}{N_T} = \exp\left[-\int_t^T b(z) \, dz\right].$$

The logarithm of this quantity plotted over time t is a straight line if the branching rate is constant; the slope of this line is equal to the branching rate. If instead the graph is curvilinear, then the slope of the graph at each point gives the instantaneous rate of cladogenesis, or the branching rate:

$$\frac{d}{dt} \ln N_t = b(t). \quad (4)$$

Case 2: Constant Rates of Branching and Extinction

In the presence of extinction, however, the observed pattern of the number of ancestral lineages does not have a simple relationship like equation (4), because only those lineages that have at least one extant descendent can be observable. Additionally, the fraction of observable lineages among those that existed at time T increases with t.

To see this, consider the simplest case of constant rates of branching b and extinction c. Equation (3) becomes

$$\frac{N_t}{N_T} = \frac{b - c}{be^{(b-c)(T-t)} - c}.$$ (5)

As illustrated in figures 2A and B, if N_t is plotted on a logarithmic scale, the curve is tangential to the line of the slope b when $T - t$ is very small, because

$$\ln N_t = \ln N_T - b(T - t) + O[(T - t)^2],$$

when $T - t$ is small. (6)

On the other hand, if $T - t$ is very large, we have an asymptotic line. The slope differs depending on whether $b > c$. If the branching rate is larger than the extinction rate ($b > c$):

$$\ln N_t = \ln\left[N_T\left(1 - \frac{c}{b}\right)\right] - (b - c)(T - t)$$

$$+ O[e^{-(b-c)(T-t)}], \quad \text{when } T - t \text{ is large.}$$ (7a)

The asymptotic line for large $T - t$ has a slope of $b - c$ (see fig. 2A). On the other hand, if the branching rate is smaller than the extinction rate ($b < c$), we have:

$$\ln N_t = \ln\left[N_T\left(1 - \frac{b}{c}\right)\right] + O[e^{-(c-b)(T-t)}],$$

when $T - t$ is large, (7b)

the asymptote of which has a slope of zero, irrespective of the magnitude of difference $c - b$ (fig. 2B). This implies that the coalescence of ancestral lineages as time goes back will stop with a finite time. The above result for $b > c$ was discussed by Harvey et al. (1994a). The border between (7a) and (7b) is the case in which $b = c$:

$$\ln N_t = \ln N_T - \ln[1 + b(T - t)],$$ (7c)

which has no asymptote.

Estimating Branching and Extinction Rates

Equations (6) and (7) suggest the possibility of estimating the rates of branching and extinction from the pattern of ancestral phylogeny if constant rates are assumed. However, the data available for analysis include only a finite number of lineages. Equation (5) is exact only when there are an infinitely large number of lineages, which is unlikely to be the case. Hence, there is inherent stochasticity caused by the finiteness of the number of lineages in applying the method, even if the branching and extinction events exactly follow the branching process model that leads to equation (5). To examine the potentiality and the limitation of the estimate based on equation (5), we conducted computer simulation studies.

We first generated phylogenies by branching processes with a constant rate of branching and extinction. Then at a time, say T, we stopped the process and constructed the ancestral phylogeny for the lineages existing at T. We then counted the number of their ancestral lineages N_t over some period before T, fitted these counts to the "theoretical curve" equation (5), and estimated two parameters b and c in the
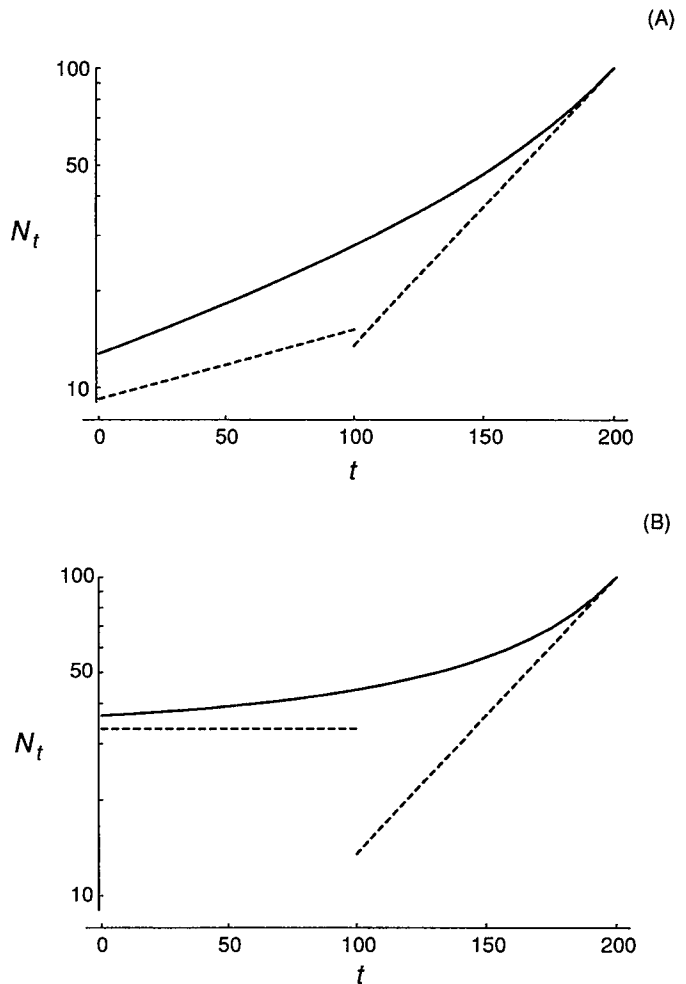

FIG. 2. The number of ancestral lineages given by equation (5). The vertical axis represents logarithmic scale, and the horizontal axis represents time in normal scale. Branching and extinction rates are constant. Other parameters are $T = 200$, and $N_T = 100$. The branching rate is larger than the extinction rate in (A): ($b = 0.02$, $c = 0.015$). The curve shows a tangential line of slope b at $t = T$. The asymptotic line when t is far back from the present ($T - t$ is large) has a slope of $b - c$. In (B), the branching rate is smaller than the extinction rate ($b = 0.02$, $c = 0.03$), and the asymptote, when t is far back from the present, ($T - t$ is large) has a slope of zero.

formula. For curve fitting, we used a nonlinear fitting algorithm, called the polytope method (Gill et al. 1981). The $\hat{f}b$ and $\hat{c}$ thus obtained are then compared with "real" parameters that are used to generate the phylogeny. For each set of parameters (b and c), we generated 1000 independent replicates and calculated the bias and variance of estimates $\hat{f}b$ and $\hat{c}$.

The length of the time period used for fitting is chosen so that the initial number of lineages is ≥ 50, and the final number is between 100 and 150. Samples from five regularly spaced times were used for nonlinear fitting. This size range of phylogenies was chosen by considering that the phylogeny of bird families analyzed by Nee et al. (1992) included 122 taxa, as illustrated in figure 3A. Also, to go back too far
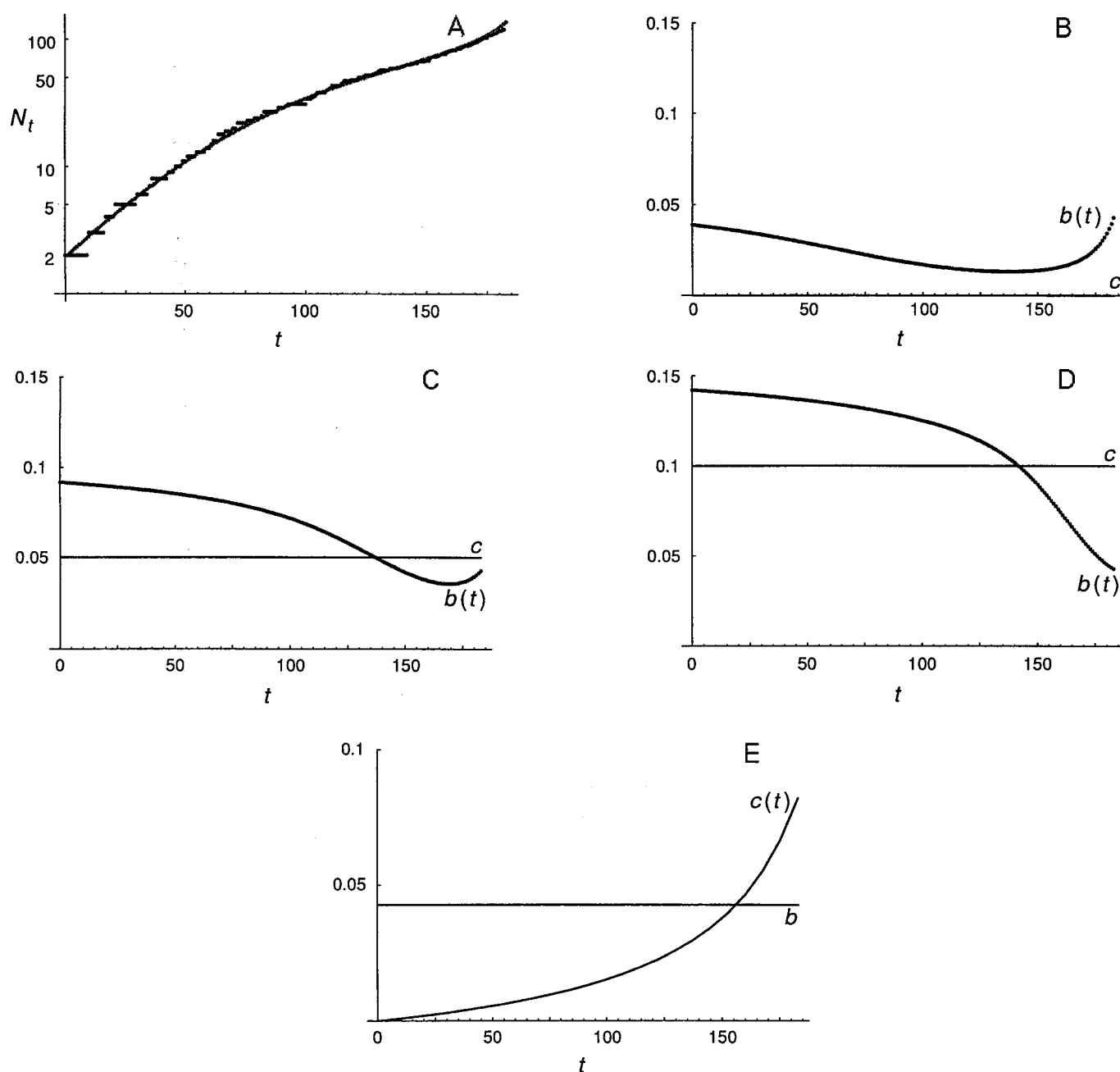
FIG. 3. (A) Black dots represent the number of ancestral lineages versus time, based on the phylogeny in Sibley and Ahlquist (1990). The horizontal axis shows the time in arbitrary units used in their paper. The first 122 taxa are displayed for analysis (as used by Nee et al. 1992). The gray solid curve shows the pattern generated by the branching-process model, using the time-dependent parameter shown in (B)–(E). (B) The branching rate decreases with time, and the extinction rate is zero. (C–D) The branching rate decreases with time, and the extinction rate stays a positive constant ([C] $c = 0.05$, [D] $c = 0.1$). (E) The branching rate is a constant $b = 0.042816$, but the extinction rate increases with time $c(t) = 0.000084t/(1 - 0.004442\,t)$, which was obtained by nonlinear fitting (polytope method). (B)–(E) generate the same curve of the number of ancestral lineages shown in (A).

would break the assumption of constant rates of branching and extinction. Figure 3A shows that after the time when there were about 50 ancestral lineages the slope of $\ln N_t$ is approximately constant (except for the increase in slope toward the end of the period, which is expected for a constant-rates case).

Figure 4 illustrates a distribution of $\hat{b}$ and $\hat{c}$ for a given

(b,c). Using the Kolmogolov-Smirnov test, we confirmed that the pair of estimated parameters have a distribution close to normal when the given branching rate is larger than the extinction rate $(b > c)$.

Figure 5 illustrates the bias of estimates of branching and extinction rates. The bias of $\hat{b}$ becomes smaller with b, whereas that of $\hat{c}$ becomes larger. Figure 6 shows that both variances
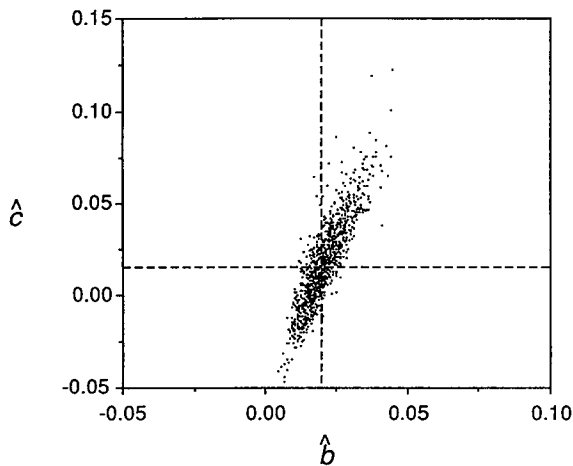
FIG. 4. An example of estimates of branching rate fb and extinction rate $\hat{c}$ using nonlinear fitting (polytope method) to equation (5). True parameters are $b = 0.02$ and $c = 0.015$, indicated by broken lines. Procedures of simulations are described in the text. The estimate of branching rate fb has a smaller variance than the estimate of extinction rate $\hat{c}$.

and covariances of these parameters decrease with b. Lines on figure 6 are results of multiple regression. If we use data only for $b > c$, the multiple regression for the variance of $\hat{b}$ is $\log_{10} \text{Var}(\hat{b}) = -4.22 - 10.5b + 23.9$, and that of $\hat{c}$ is $\log_{10} \text{Var}(\hat{c}) = -3.09 - 17.4b + 22.6c$. The multiple regression for covariance between $\hat{b}$ and $\hat{c}$ is $\log_{10} \text{Cov}(\hat{b},\hat{c}) = -3.74 - 13.8b + 23.8c$.

To summarize, if the branching rate is larger than the extinction rate ($b > c$), the estimate of b can be relatively accurate for a large phylogeny and may be applicable for the practical purpose of estimating the branching rate. In contrast, if b is similar to or smaller than c, both the bias and variance of the branching rate estimate would become extremely large. The reason for this large deviation may be related to the fact that, if $b < c$, the asymptotic slope of the curve for the logarithmic number of ancestral lineages is zero (see eq. 7B), no longer reflecting the magnitude of b and c.

The variance and the bias of the estimate of extinction rate ($b < c$) are always very large, and hence, this method is not reliable for estimating the extinction rate.

To illustrate the variability of these estimates, we examine the same phylogeny, including 122 bird families shown in figure 3A. We used only the last half of these data (during the 130–180 time unit of figure 1 of Nee et al. 1992, corresponding to the last 2.25 million yr). The estimate of branching rate is 0.022 families per 0.45 million yr, (C.I. [0.005, 0.039]). The estimate of extinction rate is 0.015 per 0.45 million yr (C.I. [−0.048, 0.068]), which is not very informative, because the interval includes zero.

Nee et al. (1994b) calculated a contour map of the likelihood function, which was derived from Nee et al. (1994a). It also shows a large variance of simultaneous estimates of b and c.

Gradual Changes of Rates

In analyzing the phylogeny of bird families estimated by Sibley and Ahlquist (1990), Nee et al. (1992) recognized that
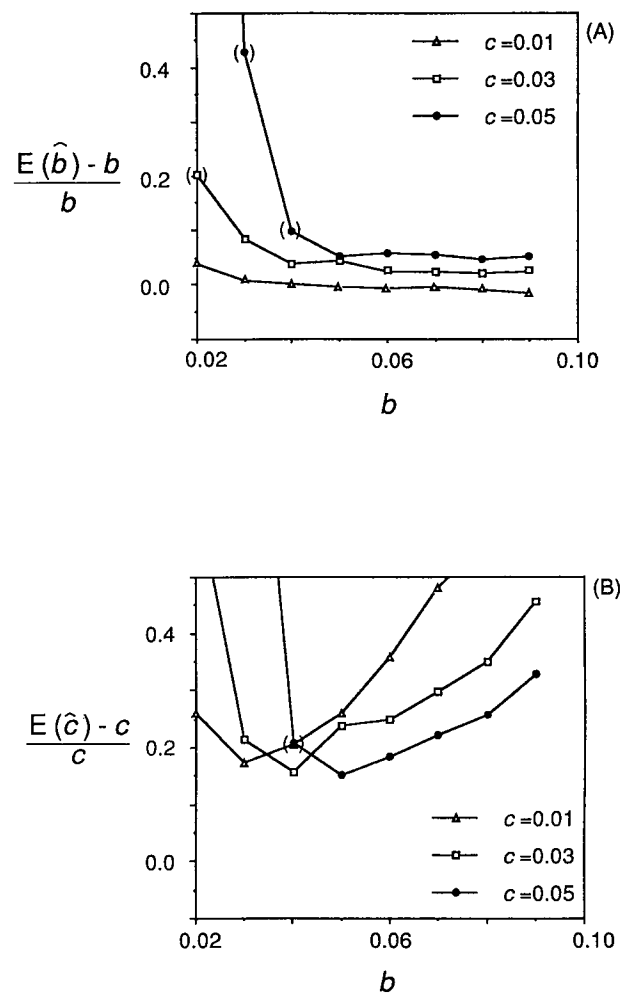

FIG. 5. The relative magnitude of bias of the estimates (A) branching rate $\hat{b}$ and (B) extinction rate $\hat{c}$. The horizontal axis designates the true branching rate b and different lines correspond to different extinction rates c. The point for $b < c$ is in brackets. We can see that the bias of $\hat{b} < 10\%$ if $b > c$, but would be much larger if $b < c$. The bias of $\hat{c}$ is always large.

the slope of the graph of $\ln N_t$ over t (see fig. 3A) is larger for the early half of the period than for the later half. This was discussed as the result of a decrease in the effective rate of cladogenesis, possibly caused by the increase in the number of competing taxa.

In figure 3A, a solid curve illustrates the number of lineages $\ln N_t$ generated by the model, using various pairs of time-dependent rates of branching and extinction. In fact, $b(t)$ and $c(t)$ given by figures 3B–E are able to generate the same curve indicated in figure 3A. Figure 3B shows a gradually decreasing branching rate, assuming no extinction occurs. Figures 3C and D are examples with a time-dependent branching rate and a constant extinction rate. In contrast, figure 3E shows the case of a constant branching rate and an increasing extinction rate.

This result can be generalized: suppose in the first case that the branching rate is a time dependent $b_1(t)$, and the extinction rate is a constant c_1. In the second case, the branching rate is a fixed constant b_2, and the extinction rate is $c_2(t)$
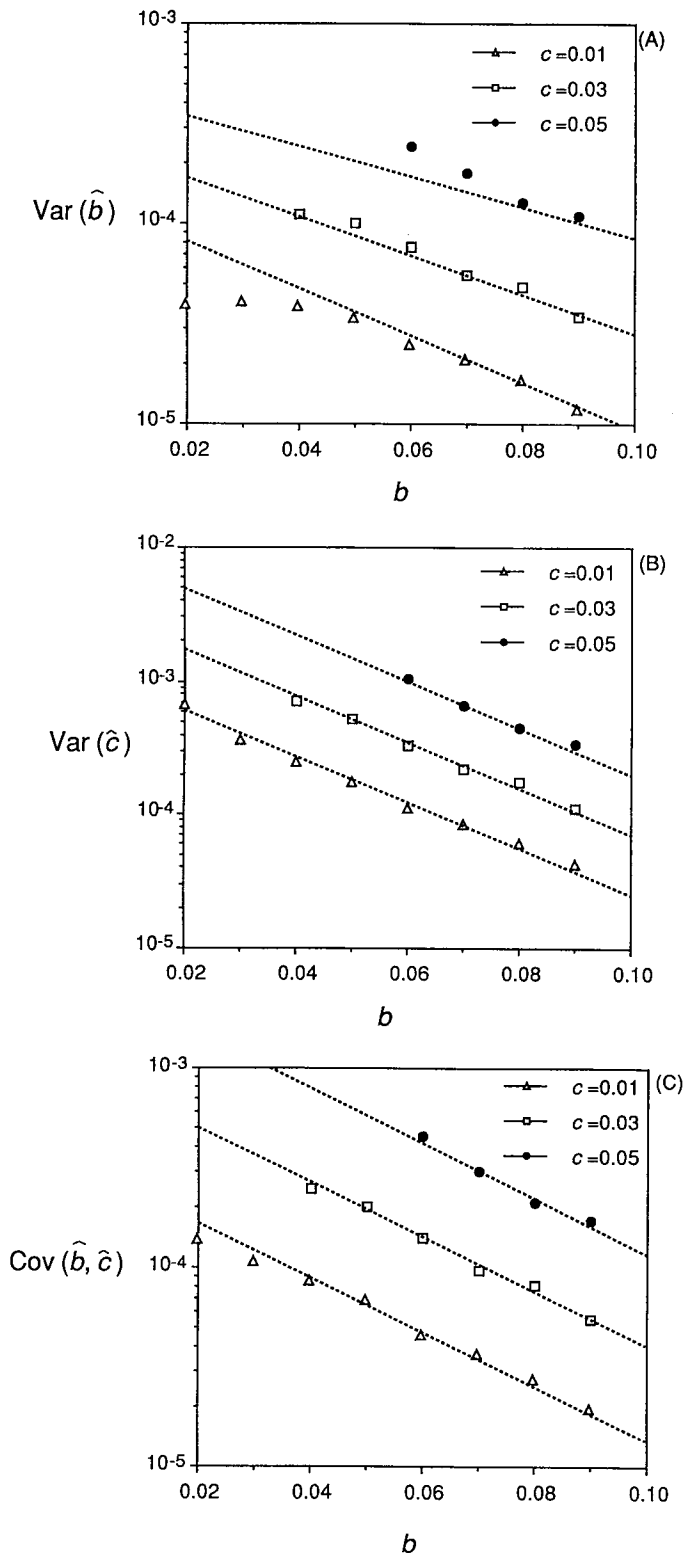
Fig. 6. Variances and covariance of estimated parameters. (A) shows the variance of the branching-rate estimate; (B) shows the variance of the extinction-rate estimate; and (C) shows their co-variance. Only the data for $b > c$ are indicated. Regression formulas are in the text.

is a function of time t. These two cases give the same pattern of the ancestral lineage number N_t, provided that an integral in equation (3) is the same between them; that is, if

$$\int_t^T b_1(u)\exp\left[\int_u^T [b_1(z) - c_1]\,dz\right]du$$

$$= \int_t^T b_2\exp\left\{\int_u^T [b_2 - c_2(z)]\,dz\right\}du$$

holds for all t. After some computation including derivatives, we can confirm that this condition is mathematically equivalent to the following two conditions:

$$b_2 = b_1(T),\tag{8a}$$

and

$$c_2(t) = \frac{d}{dt}\ln b_1(t) - b_1(t) + b_2 + c_1.\tag{8b}$$

Therefore, if equation (8B) is positive for all t, then equations (8A,B) specify the way to produce exactly the same pattern. There are infinitely many cases intermediate between these two that also generate the same $\ln N_t$ pattern, in which both the branching rate and the extinction rate change with time. This shows a limitation of this method for use in estimating past events.

However, we should not conclude that all the temporal patterns of a branching rate can be expressed by a time-dependent extinction rate with a constant branching rate. Equation (8B) is positive only when the change of $b(t)$ is gradual. If there is a brief period of explosive speciation (as in fig. 7A) during the period of rapid decline of $b(t)$ back to a "normal" level, the term $d[\ln b_1(t)]/dt$ in equation (8B) would have a large negative value, making it impossible for a positive value of $c_2(t)$ to satisfy the equation. Then the resulting pattern of N_t cannot be explained by a combination of a fixed branching rate and a time-dependent extinction rate. Examples that illustrate the asymmetry of the role of branching and extinction will be shown below

Inferring Past Episodes

Here we examine the possibility of learning episodic events, such as explosive branching or mass extinction. These are likely to occur together, because vacant niches created by mass extinction of previous taxa would provide the opportunity for new species to fill the niches, causing subsequent high cladogenesis. To examine the separate effects of these two factors on the pattern of ancestral phylogeny, we next consider the case where the branching rate has a brief period of high value but the extinction rate stays constant, and the case where only the extinction rate has a sharp peak.

Case 3: Explosive Branching

Figure 7 illustrates the case in which explosive branching occurs over a short period. Specifically $b(t)$ is enhanced to be 10 times higher than the normal level for four generations. In this brief period, the number of lineages increased about 1.5 times, as before. The $\ln N_t$ curve in figure 7B shows a
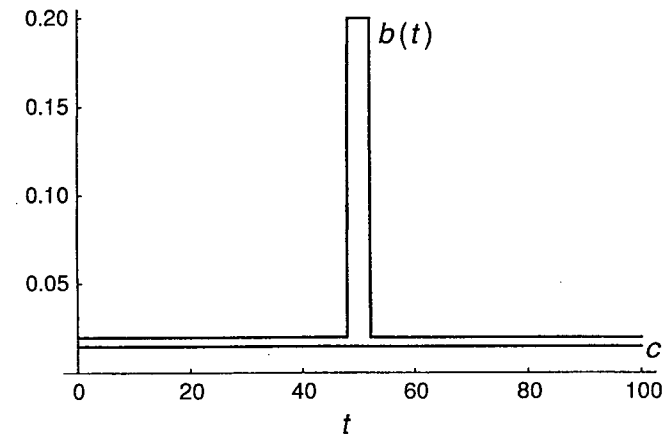
FIG. 7. Explosive branching. The branching rate is usually 0.020 but is 0.200 for a brief period (48 ≤ t ≤ 52). The extinction rate is 0.015 throughout the time.

FIG. 8. Mass extinction. The extinction rate is usually 0.015 but is 0.150 for a brief period (48 ≤ t ≤ 52). The branching rate is 0.020 throughout the time.

clear pattern of quick increase. It is also possible to determine the magnitude of the explosion from the pattern.

To express a very sharp peak in branching rate at a time $t = \tau$, we may use Dirac's delta function $\delta(t - \tau)$:

$$b(t) = b_0 + b_1\delta(t - \tau), \tag{9a}$$

$$c(t) = c_0, \tag{9b}$$

which indicates that during a very brief period, including $t = \tau$, the (actual) number of taxa should increase to e^{b_1} times, as before. We can calculate that α in equation (B3) is

for $t > \tau$,

$$\alpha = \frac{b_0}{b_0 - c_0}[e^{(b_0-c_0)(T-t)} - 1], \tag{10a}$$

for $t < \tau$,

$$\alpha = \frac{b_0}{b_0 - c_0}[e^{(b_0-c_0)(T-t)} - 1] + e^{(b_0-c_0)(T-\tau)}(e^{b_1} - 1)$$

$$+ \frac{b_0}{b_0 - c_0}e^{(b_0-c_0)(T-\tau)}e^{b_1}[e^{(b_0-c_0)(\tau-t)} - 1], \tag{10b}$$

which has a discontinuity at $t = \tau$. Hence, the number of ancestral lineages N_t also has a jump at a time of explosive branching.

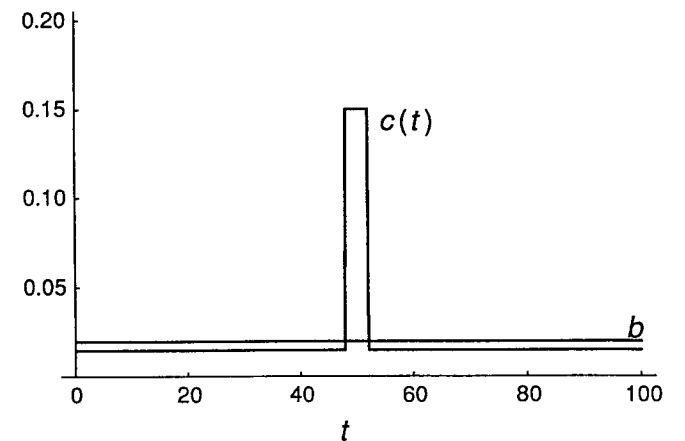## Case 4: Mass Extinction

Figure 8 shows the result of mass extinction, in which $c(t)$ is enhanced to be 10 times higher than the normal level for four generations, but the branching rate $b(t)$ stays at a constant level. As a result, the number of lineages changes little, as before, in this brief period. Figure 8B shows that there is no clear jump around the time of mass extinction, although the slope of $\ln N_t$ seems to change. The effect of the mass extinction is not as visible as explosive branching.

This can be shown by considering the case in which the branching rate has a very sharp peak at a time $t = \tau$, expressed in terms of Dirac's delta function $\delta(t - \tau)$:

$$b(t) = b_0, \tag{11a}$$

$$c(t) = c_0 + c_1\delta(t - \tau), \tag{11b}$$

which indicates that during a very short period around $t = \tau$, the number of taxa decrease to $-e^{c_1}$ times, as before. In Appendix B, we can calculate that α in equation (3) is

for $t > \tau$,

$$\alpha = \frac{b_0}{b_0 - c_0} [e^{(b_0-c_0)(T-t)} - 1], \tag{12a}$$

for $t < \tau$

$$\alpha = \frac{b_0}{b_0 - c_0} [e^{(b_0-c_0)(T-\tau)} - 1]$$

$$+ \frac{b_0}{b_0 - c_0} e^{-c_1} [e^{(b_0-c_0)(T-t)} - e^{(b_0-c_0)(T-\tau)}]. \tag{12b}$$

Unlike in equation (10), α has no discontinuity. However it has a "kink," that is, discontinuity of its derivative at $t = \tau$. Harvey et al. (1994a) derived equation (12) (they called it "random extinction"), and compared it with the mass extinction of two other types (clade and uniform extinctions). Harvey et al. (1994b) pointed out that equation (12B) with $\tau = T$ predicts the case in which only a fraction (e^{-c_1} of extant lineages has been analyzed, and they discussed potential effects of incomplete sampling on the shape of phylogenies of viral DNA sequences.

In short, an explosive branching event is easy to observe, but mass extinction, if not accompanied by subsequent explosive branching, does not appear as a clear event in the pattern of the number of ancestral lineages of extant taxa.

Density-Dependent Branching and Extinction

Nee et al. (1992) analyzed the pattern of the number of ancestral lineages in figure 3A as the result of an "effective branching rate" decreasing with an increase of the number of taxa. To perform this analysis, they used the instantaneous rate of increase in N_t over the number of ancestral lineages N_t, and they found that the pattern is compatible with $dN/dt = pN/N^\alpha$ but not with $dN/dt = rN(1 - N/K)$. This analysis is very interesting and can be interpreted as evidence for the branching rate decreasing and for the extinction rate increasing with the number of bird families existing. However, this argument is valid only when N_t can be regarded as a good measure of the real number of existing families. Unfortunately, in the previous sections, we have seen that the temporal pattern in the number of ancestral lineages of extant taxa N_t can be very different from that of the actual number of taxa. The number of existing species are denoted here by $S(t)$, which affect the branching and extinction rates of taxa in a density-dependent manner. As shown in the previous example, the number of taxa $S(t)$ may be greatly reduced in a short time by mass extinction, but the number of ancestral lineages $N(t)$ shows no clear change during that time (see fig. 8). In addition, the number of ancestral lineages always increases with time, even though the actual number of lineages existing, $S(t)$, does not change, or greatly fluctuates, or even decreases.

Fortunately, under certain conditions, we can show that $S(t)$ in the past can, in fact, change in parallel to $N(t)$, justifying the analysis by Nee et al. (1992) who use N_t as a surrogate of $S(t)$. Here we assume that the branching rate $b[S(t)]$ and extinction rate $c[S(t)]$ are functions of the actual number of species $S(t)$. Then the number of species changes as

$$\frac{d}{dt} \ln S = b[S(t)] - c[S(t)]. \tag{13}$$

The processes we consider here are not branching processes but birth-and-death processes (more general than Kendall [1948]), because the species interact with each other by affecting the number of species $S(t)$. If the number of taxa is sufficiently large, then the number of taxa increase with equation (13), although there is stochasticity caused by the finiteness of the number of taxa, which is correctly dealt with by birth-and-death processes. Instead, we simply regard the results as time-dependent branching processes and assume that the time-dependent branching rate and extinction rate are $b[S(t)]$ and $c[S(t)]$, which are functions of the number of taxa $S(t)$.

Here we assume that: (1) The branching rate is significantly higher than the extinction rate; (2) The extinction rate does not become very large; and (3) The time we observe is sufficiently far from the present. According to the computations in Appendix C, we can prove that the following equation should hold approximately:

$$\frac{d}{dt} \ln N_t = b(t) - c(t)$$

$$= b[S(t)] - c[S(t)]. \tag{14}$$

Comparing this with equation (13) shows that the number of ancestral lineages $N(t)$ changes in a way parallel to the actual number of lineages $S(t)$, including both observable and non-observable lineages.

The above argument still holds if b and c depend explicitly on t in addition to indirectly via $S(t)$.

DISCUSSION

Suppose that we have an ancestral phylogeny of a set of species and are asked to compute the rate of speciation. The simplest way to do so is to calculate the factor of multiplication in the number of species between some time ago, say t, and the present, denoted by T, by computing the ratio of the number of extant species N_T to the number of ancestral lineages at t, N_t. This ratio should then be set equal to $\exp[\bar{b}(T - t)]$, where $\bar{b}$ is the average rate of cladogenesis. This, in effect, calculates $\bar{b}$ by fitting it to $N_T = N_t e\bar{b}(T - t)$. However, the interpretation of $\bar{b}$ is not as simple as it looks. The rate thus computed depends on the rate of branching $b(t)$ and the rate of extinction $c(t)$, as:

$$\bar{b} = \frac{1}{T - t} \ln \frac{N_T}{N_t}$$

$$= \frac{1}{T - t} \ln \left[1 + \int_t^T b(u) \exp \left\{ \int_u^T [b(z) - c(z)] \, dz \right\} du \right], \tag{15}$$

from equation (3). If both the branching rate and the extinction rate are constant, $\bar{b}$ takes the value between true branching rate b and the difference between branching rate and extinction rate c ($b - c < \bar{b} < b$). It is close to b if the time horizon is short ($T - t$ is shown) but becomes closer to $b - c$ if the time horizon is long. Rate $\bar{b}$ should underestimate the true rate of branching b, and overestimate the rate of increase in the number of species, which is $b - c$.

Here we examined the pattern of ancestral phylogeny reconstructed by the extant taxa when the branching rate and

extinction rate are given. We ask specifically how much we can learn about these processes of cladogenesis from the pattern of ancestral phylogeny.

If viewed as the method to estimate branching rate and extinction rate, the method of fitting to the results obtained by the branching-process model assuming constant rates, such as equation (5), is not very effective, because the estimate includes a large variance, even for a fairly large phylogeny including more than 100 extant taxa. The estimate of extinction rate is especially poor and practically useless because of the large bias and variance, although the branching rate estimate can be useful. However, the mathematical study here and elsewhere (e.g., Harvey et al. 1994a,b; Nee et al. 1994a,b) would still be important in demonstrating that the relationship between the shape of the ancestral phylogeny and the basic processes, such as branching and extinction, is more complicated than we might think.

Figure 3 clearly illustrates that, in general, different pairs of branching rates and extinction rates can give exactly the same pattern of ancestral phylogeny, expressed in terms of the number of ancestral lineages over various times. We have shown the equivalence relationship between the case with a time-dependent branching rate and a constant extinction rate and the second case with a constant branching rate and a time-dependent extinction rate. Generating the same ancestral phylogeny does not mean that these two processes are equivalent in evolution, because temporal change in the actual number of taxa, denoted by $S(t)$, may greatly differ between these processes. This shows a clear limitation of the method of inferring past events from detailed knowledge, such as of molecular phylogenies, of the extant species only.

On the other hand, there is strong asymmetry between the branching extinction rates. Explosive branching, a short period of very high branching rate, would create a jump in the number of ancestral lineages—many extant species must show synchronized branching at the time of such an explosive branching episode (fig. 7). In sharp contrast to this, mass extinction, a brief period of random extinction of existing taxa, would not appear clearly in the number of ancestral lineages as a function of time (fig. 8).

Asymmetry between the branching extinction rates is also apparent from the equivalence relationship. We have shown that the pattern of ancestral phylogeny generated by a time-dependent branching rate having a sharp peak cannot be generated by a constant branching rate with a time-dependent extinction rate, because a sudden increase in the number of ancestral lineages produced by explosive branching cannot be generated by any time-dependent extinction-rate function. But the reverse is not the case. In Appendix D, we show that any pattern of ancestral phylogeny generated by a time-dependent extinction rate can also be generated by a time-dependent branching rate, again showing the difficulty of estimating any special extinction event.

In general, it is likely that mass extinction would be accompanied by explosive branching, and if so, we can recognize the latter, not the former, from the pattern of the ancestral phylogeny of extant taxa.

Despite recent extensive and careful attempts to detect mass extinction events from the pattern of molecular phylogenies (e.g., Harvey et al. 1994a,b; Nee et al. 1994a,b), our study suggests that it would be rather difficult to recognize (pure) mass extinction events if we assume both the branching and extinction rates are parameters that can be freely chosen to fit the data. A more profitable way to extract information on past extinction processes from molecular phylogenies might be to place additional constraints on how extinction and branching (or speciation) rates are coupled.

Nee et al. (1992, p. 8322) used the phrase "effective rate of cladogenesis," indicating "the rate at which the lineages that are still extant gave rise to new lineages that are also extant." This correctly defines the rate of increase in Nt, and the time-derivative of $\ln Nt$ may be called "effective instantaneous rate of cladogenesis." However, an effective instantaneous rate thus defined has no simple relationship with the evolutionary processes generating the pattern. Instead, it has a complicated dependency on the $b(t)$ and $c(t)$. Things would be a lot easier if it had a simple dependency as equation (14) holds, which seems to be implicitly assumed in the analysis by Nee et al. (1992, p. 8322), who suggested a straight line for a semilog plot of the number of ancestral lineages over time if "nothing peculiar happens." Unfortunately, equation (14) holds only when the number of ancestral lineages of the extant taxa change in parallel to the actual number of taxa, which, in turn, can be justified only under restricted conditions in which (1) the extinction rate does not have a sharp peak, (2) the branching rate is sufficiently larger than the extinction rate c, and (3) the time since before the present $(T - t)$ is sufficiently long. Although these seem to be satisfied in the situation analyzed by Nee et al. (1992), we must be cautious in using the information from ancestral lineages that assumes competition between taxa.

All the analyses here are based on the assumption that we have a correctly reconstructed phylogeny. The validity of the conclusions we derive is limited by the validity of this assumption. As carefully discussed by Nee et al. (1992), it is possible that the method for estimating branching time using DNA hybridization, for example, may include error or bias. This is certainly the factor affecting the apparent increase of branching rates, such as the one in figure 3A. In addition to DNA hybridization, there are several techniques, including the banding pattern of restricted enzyme fragments, and direct sequencing of DNA molecules (Nei 1987). In using these techniques for reconstruction of phylogeny, we must assume some sort of molecular clock hypothesis (Zuckerkandl and Pauling 1962; Kimura 1983), which may not be, strictly speaking, exactly as exemplified by overdispersed molecular evolution (Ohta and Kimura 1971; Kimura 1983, Takahata 1987; Gillespie 1991; Iwasa 1993). However, we may separate the problem of estimating the correct phylogeny from that of inferring the processes from the obtained phylogeny. The latter is the theme here, and we believe it is also an important theoretical question to ask. Nee et al. (1992) attempted more sophisticated analysis of the pattern of molecular phylogeny, such as those using information on body size. Although this is a preliminary study, it points out the need for further theoretical studies on how much can be known from reconstructed molecular phylogenies.

Education. We are grateful to S. Nee, P. Harvey, and B. May for reprints of their stimulating works and for important advice on our paper. We also thank following people for their helpful advice and discussions: H. Ezoe, F. Koike, P. Haccou, Y. Harada, H. Matuda, H. Matsuda, T. Murakami, Y. Nishi, A. Pomiankowski, and A. Sasaki.

LITERATURE CITED

Gill, P. E., W. Murray, and M. H. Wright. 1981. Practical optimization. Academic Press, New York.

Gillespie, J. H. 1991. The causes of molecular evolution. Oxford University Press, Oxford.

Harris, T. E. 1963. The theory of branching processes. Springer, New York.

Harvey, P. H., S. Nee, A. Ø. Mooers, and L. Partridge. 1991. These hierarchical views of life: phylogenies and metapopulations. Pp. 123–137 in R. J. Berry, T. J. Crawford, and G. M. Hewitt, eds. Genes in ecology. Blackwell, Oxford.

Harvey, P. H., R. M. May, and S. Nee. 1994a. Phylogenies without fossils: inferring lineage, birth, and death rates. Evolution. In press.

Harvey, P. H., E. C. Holmes, A. Ø. Mooers, and S. Nee. 1994b. Inferring evolutionary processes from molecular phylogenies. Pp. 313–333 in R. W. Scotland, D. J. Siebert, and D. M. Williams, eds. Models in phylogeny reconstruction. Systematics Association Special Volume, Series 52. Oxford University Press, Oxford.

Iwasa, Y. 1993. Overdispersed molecular evolution in constant environments. Journal of Theoretical Biology 164:373–393.

Karlin, S., and H. M. Taylor. 1975. First course in stochastic processes. Academic Press, New York.

Kendall, D. G. 1948. On the generalized "birth-and-death" process. Annals of Mathematical Statistics 19:1–15.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Nee, S., A. Ø. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proceedings of the National Academy of Sciences, USA 89:8322–8326.

Nee, S., R. M. May, and P. H. Harvey. 1994a. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London B 339:139–146.

Nee, S, E. C. Holmes, R. M. May, and P. H. Harvey. 1994b. Estimating extinction from molecular phylogenies. Pp. 164–182 in J. L. Lawton and R. M. May, eds. Estimating extinction rates. Oxford University Press, Oxford.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Ohta, T., and M. Kimura. 1971. On the constancy of the evolutionary rate of cistrons. Journal of Molecular Evolution 1:18–25.

Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and classification of birds. Yale University Press, New Haven, Conn.

Takahata, N. 1987. On the overdispersed molecular clock. Genetics 116:169–179.

Zuckerkandl, E., and L. Pauling. 1962. Molecular disease, evolution, and genetic heterogeneity. Pp. 189–225 in M. Kasha and B. Pullman, eds. Horizons in biochemistry. Academic Press, New York.

Corresponding Editor: D. Maddison

APPENDIX A

Here we give a self-contained derivation of equation (1), on which all the biological arguments presented here are based. Kendall (1948) analyzed the model using a forward equation for a generating function, and Harvey et al. (1994a) and Nee et al. (1994a) started their analyses from Kendall's result. In the following, we show a different proof by treating the model as branching processes using a backward equation for the generating function.

Let $P_k(t,T)$ be the probability that a single taxon randomly chosen at time t will have k descendant taxa at time T. The generating function is:

$$g(s; t, T) = E[s^{N_T} | N_t = 1] = \sum_{k=0}^{\infty} s^k P_k(t, T), \quad (A1)$$

where s is a parameter satisfying $0 < s \le 1$. This can be decomposed according to the events during time interval of length Δt

$$g(s; t - \Delta t, T) = b(t) \Delta t \, g(s; t, T)^2 + c(t) \Delta t$$
$$+ [1 - b(t) \Delta t - c(t) \Delta t]g(s; t, T) + o(\Delta t^2), \quad (A2)$$

where the first term of the right-hand side is for branching of a taxon into two, occurring with probability $b(t)\Delta t$; the second term is for the extinction of the taxon with probability $c(t)\Delta t$; and the third term is for the event where the number of taxa remains one with probability $[1 - b(t)\Delta t - c(t)\Delta t$. When we take the limit $\Delta \to 0$ with some rearrangement of terms, equation (A2) becomes the following partial differential equations:

$$-\frac{\partial g}{\partial t} = b(t)g^2 + c(t) - [b(t) - c(t)]g$$
$$= [b(t)g - c(t)](g - 1). \quad (A3a)$$

Because the terminal condition is $N_T = 1$, we have

$$g(s; T, T) = s. \quad (A3b)$$

By putting $G = 1/(1 - g)$, equations (A3a) and (A3b) become

$$-\frac{\partial G}{\partial t} = -[b(t) - c(t)]G + b(t) \quad \text{for } t < T,$$

$$G(T) = \frac{1}{1 - s}$$

which can be solved as

$$G(t) = \frac{1}{1 - s}\exp\left\{-\int_t^T [b(u) - c(u)] \, du\right\}$$

$$+ \int_t^T b(u)\exp\left\{-\int_t^u [b(z) - c(z)] \, dz\right\} du,$$

$$= \left(\frac{1}{1 - s} + \alpha\right)\frac{1}{\beta}, \quad \text{for } t \le T \quad (A4)$$

where α is given by equation (1b) and β is given by (1c). By Taylor expansion, equation (A4) is rewritten as:

$$g(s; t, T) = 1 - \frac{1}{G}$$

$$= \frac{1 + \alpha - \beta}{1 + \alpha} + \frac{\beta}{\alpha(1 + \alpha)}\left[\frac{\alpha}{1 + \alpha}s + \left(\frac{\alpha}{1 + \alpha}\right)^2 s^2\right.$$

$$\left. + \left(\frac{\alpha}{1 + \alpha}\right)^3 s^3 + \ldots\right]. \quad (A5)$$

Hence the probability P^k as the coefficient of s^k ($k = 0, 1, 2, 3, \ldots$) is given by equation (1) in the text.

APPENDIX B

Case 3: Explosive Branching

First we note the following equation:

$$\exp\left[\int_u^T b(z) \, dz\right] = \begin{cases} \exp[b_0(T - u) + b_1] & \text{for } u < \tau. \\ \exp[b_0(T - u)] & \text{for } u > \tau. \end{cases} \quad (B1)$$

Hence,

$$-\frac{d}{du}\exp\left[\int_u^T b(z) \, dz\right] = e^{b_0(T-\tau)}(e^{b_1} - 1)\delta(u - \tau)$$

$$+ \begin{cases} b_0\exp[b_0(T - u) + b_1] & \text{for } u < \tau. \\ b_0\exp[b_0(T - u)] & \text{for } u > \tau. \end{cases}$$

Then, equation (1b) becomes

$$\alpha = \int_t^T b(u)\exp\left\{\int_u^T [b(z) - c_0]\, dz\right\}\, du$$

$$= \int_t^T e^{-c_0(T-u)}\left(-\frac{d}{du}\exp\left\{\int_u^T b(z)\, dz\right\}\right)\, du.$$

For $t > \tau$,

$$\alpha(t) = \int_t^T e^{-c_0(T-u)}b_0 e^{b_0(T-u)}\, du$$

$$= \frac{b_0}{b_0 - c_0}[e^{(b_0-c_0)(T-t)} - 1], \qquad (B2a)$$

for $t < \tau$,

$$\alpha(t) = \alpha(\tau + 0) + e^{(b_0-c_0)(T-\tau)}(e^{b_1} - 1) + \int_t^T b_0 e^{(b_0-c_0)(T-u)}e^{b_1}\, du$$

$$= \frac{b_0}{b_0 - c_0}[e^{(b_0-c_0)(T-\tau)} - 1] + e^{(b_0-c_0)(T-\tau)}(e^{b_1} - 1)$$

$$+ \frac{b_0}{b_0 - c_0}e^{b_1}e^{(b_0-c_0)(T-\tau)}[e^{(b_0-c_0)(\tau-t)} - 1], \qquad (B2b)$$

which has a jump at $t = \tau$.

Case 4: Mass Extinction

Now $c(t)$ includes the delta function.

$$\exp\left[-\int_u^T c(z)\, dz\right] = \begin{cases} \exp[-c_0(T - u) - c_1] & \text{for } u < \tau. \\ \exp[-c_0(T - u)] & \text{for } u > \tau. \end{cases} \qquad (B3)$$

Then equation (1b) is

$$\alpha(t) = \int_t^T b_0\exp\left\{\int_u^T [b_0 - c(u)]\, dz\right\}\, du.$$

For $t > \tau$,

$$\alpha(t) = \int_t^T b_0 e^{(b_0-c_0)(T-u)}\, du$$

$$= \frac{b_0}{b_0 - c_0}[e^{(b_0-c_0)(T-t)} - 1]. \qquad (B4a)$$

For $t < \tau$,

$$\alpha(t) = \alpha(\tau + 0) + \int_t^T b_0 e^{(b_0-c_0)(T-u)}e^{-c_1}\, du$$

$$= \frac{b_0}{b_0 - c_0}[e^{(b_0-c_0)(T-\tau)} - 1]$$

$$+ \frac{b_0}{b_0 - c_0}e^{-c_1}e^{(b_0-c_0)(T-\tau)}[e^{(b_0-c_0)(\tau-t)} - 1], \qquad (B4b)$$

APPENDIX C

After some computations, equation (3) can be rewritten as

$$\ln N_t = \ln N_T - \int_t^T [b(z) - c(z)]\, dz$$

$$- \ln\left(1 + \int_t^T c(u)\exp\left\{-\int_t^u [b(z) - c(z)]\, dz\right\}\, du\right). \qquad (C1)$$

Now we place the following two assumptions:

$$c(t) < c_{max}, \qquad (C2a)$$

$$b(x) - c(x) > \delta > 0. \qquad (C2b)$$

Equation (C2a) indicates that the rate of extinction is bounded. Equation (C2b) indicates that the branching rate is always higher than the extinction rate so that their difference is larger than a positive constant δ. Then, the last term of equation (C1) satisfies

$$(0 <)\ln\left(1 + \int_t^T c(u)\exp\left\{-\int_t^u [b(z) - c(z)]\, dz\right\}\, du\right)$$

$$< \ln\left[1 + \int_t^T c_{max}\exp\left(-\int_t^u \delta\, dz\right)\, du\right]$$

$$< \ln\left(1 + \frac{c_{max}}{\delta}\right). \qquad (C3)$$

In contrast, the second term of equation (C1) is

$$\int_t^T [b(z) - c(z)]\, dz > \int_t^T \delta\, dz = \delta(T - t). \qquad (C4)$$

From (C4) and (C3), the third term of (C1) can be negligibly small compared with the second term if

$$T - t \gg \frac{1}{\delta}\ln\left(1 + \frac{c_{max}}{\delta}\right). \qquad (C5)$$

Hence we can say that as times goes back sufficiently far from (C1) with the third term neglected, we have an approximate formula, equation (14) in the text.

APPENDIX D

Consider first the case in which branching rate is a constant b_2, but the extinction rate changes with time $c_2(t)$. Then think about the possibility of finding the time-dependent branching rate $b_1(t)$ and the constant extinction rate c_1 that give the same pattern of N_t. The condition for equivalence is given again by equation (8), but we should then regard $b_1(t)$ as an unknown function. Equation (8b) specifies a differential equation:

$$-\frac{d}{dt}\ln b_1(t) = [b_2 + c_1 - c_2(t)] - b_1(t) \qquad (\text{for } t \leq T), \qquad (D1)$$

which can be rewritten as

$$-\frac{d}{dt}\frac{1}{b_1} = [b_2 + c_1 - c_2(t)]\frac{1}{b_1} - 1$$

and be integrated using terminal condition (8a), as

$$b_1(t) = \left(\frac{1}{b_2}\exp\left\{-\int_t^T [b_2 + c_1 - c_2(u)]\, du\right\}\right.$$

$$+ \left.\int_t^T \exp\left\{-\int_t^u [b_2 + c_1 - c_2(z)]\, dz\right\}\, du\right)^{-1}, \qquad (D2)$$

which is positive for any b_2, $c_2(t)$, and c_1. Hence, the pattern of ancestral phylogeny generated by any time-dependent extinction rate can also be generated by the time-dependent branching rate given by (D2).