

CALIBRATING A MOLECULAR CLOCK FROM PHYLOGEOGRAPHIC DATA: MOMENTS AND LIKELIHOOD ESTIMATORS

MICHAEL J. HICKERSON,^{1,2} MICHAEL A. GILCHRIST,^{3,4} AND NAOKI TAKEBAYASHI^{1,5}

¹Department of Biology, Duke University, Box 90338, Durham, North Carolina 27708

²E-mail: mjh2@duke.edu

³Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

⁵E-mail: ntakebay@duke.edu

Abstract.—We present moments and likelihood methods that estimate a DNA substitution rate from a group of closely related sister species pairs separated at an assumed time, and we test these methods with simulations. The methods also estimate ancestral population size and can test whether there is a significant difference among the ancestral population sizes of the sister species pairs. Estimates presented in the literature often ignore the ancestral coalescent prior to speciation and therefore should be biased upward. The simulations show that both methods yield accurate estimates given sample sizes of five or more species pairs and that better likelihood estimates are obtained if there is no significant difference among ancestral population sizes. The model presented here indicates that the larger than expected variation found in multitaxa datasets can be explained by variation in the ancestral coalescence and the Poisson mutation process. In this context, observed variation can often be accounted for by variation in ancestral population sizes rather than invoking variation in other parameters, such as divergence time or mutation rate. The methods are applied to data from two groups of species pairs (sea urchins and *Alpheus* snapping shrimp) that are thought to have separated by the rise of Panama three million years ago.

Key words.—Ancestral population size, coalescent, DNA substitution rate, geminate species, phylogeography.

Received October 22, 2002. Accepted April 22, 2003.

For a wide range of evolutionary studies that use DNA sequence data, the rate of DNA substitution is an essential, yet often unstated, parameter. Whether dating evolutionary processes, estimating ancestral population parameters, or testing hypotheses regarding shared history among species that presently share geographic ranges, the conclusions researchers make are heavily dependent on the assumed rate of DNA substitution. The assumed rate of DNA substitution will affect the dating of speciation events, population divergences, range expansions, colonization, selective sweeps, origin of alleles, the origins of endemic ecotypes (Cunningham and Collins 1994; Klicka and Zink 1997; Schneider et al. 1998; Avise 2000; Knowles 2000; Slatkin 2000; Hellburg et al. 2001; Wares and Cunningham 2001), and other estimates of ancestral population parameters such as ancestral population size or the extent of subdivision (Takahata et al. 1995; Wakeley 2000). In practice, DNA substitution rates are often estimated independently from sister taxa that arose at a time estimated from geological data. In taxonomic groups with poor fossil records, these independent estimates may be the only hope for obtaining good estimates of DNA substitution rates. In the best scenarios, a single estimate can be obtained from multiple species pairs that are closely related and arose from a single geologically dated event. Estimating a DNA substitution rate from multiple species pairs is, of course, preferable to obtaining an estimate from a single species pair because the error associated with the estimate should be reduced by a larger sample size. Examples of such geological events include the rise of the Panamanian Isthmus, the separation of Spain and North Africa, and the emergence of the Hawaiian Islands (Knowlton 1993; Kambysellis et al. 1995;

Wagner and Funk 1995; Fleischer et al. 1998; Roderick and Gillespie 1998; Marko and Moran 2002).

However, these multiple-species-pair datasets often show pairwise genetic distances that exceed what is expected given a Poisson model of equal mutation rates and equal divergence times (Knowlton 1993; Tarr and Fleischer 1993; Cunningham and Collins 1994; Bermingham et al. 1997; Fleischer et al. 1998; Knowlton and Weigt 1998; Lessios et al. 2001b; Wares 2001; Marko 2002; Marko and Moran 2002). Although these studies have invoked extrinsic explanations to the disparity in genetic distances such as variation in the timing of divergence, variation in the DNA substitution rate, or incorrect sister species status due to extinction (Jackson et al. 1993; Roopnarine 2001; Marko 2002), these data have not been analyzed with a model that explicitly incorporates variation in gene coalescent time within the ancestral species. Assuming the DNA substitution rate is equal among closely related species pairs and that these sister species pairs arose from allopatric speciation at a single time, coalescent theory predicts that extrinsic sources of variance may need not be invoked to explain the empirical observations (Edwards and Beerli 2000). This intrinsic variance would come from two independent processes: (1) the variance in the time to the most recent common ancestor (MRCA) or coalescence time of pairs of alleles in the ancestral species; and (2) the variance associated with the number of mutations along the lineages of descent (Kingman 1982a,b; Tajima 1983; Hudson 1990; Takahata and Satta 1997; Edwards and Beerli 2000; Wakeley 2000). Although empirical studies have accounted for the latter source of variance, which is expected to follow a Poisson distribution, ignoring the former source of variance causes the estimates of DNA mutation rates to be substantially upwardly biased. For example, at a haploid locus the variance in the MRCA between the two genetic lineages that are an-

⁴ Present address: Department of Biology, University of New Mexico, 167A Castetter Hall, Albuquerque, New Mexico 87131-1091; E-mail: mike@unm.edu.

central to today's sample in the sister species is the square of the ancestral female population size.

The work presented here uses a neutral coalescent model to derive moments and likelihood estimators of a DNA substitution rate and ancestral population size. Our methods are for data from closely related multiple sister species pairs that arose by physical separation at assumed times. Our strategy is analogous to methods developed to estimate the divergence time and the effective population size in the common ancestor of two sister species from multiloci data (Takahata 1986; Takahata et al. 1995; Takahata and Satta 1997; Yang 1997). However, instead of using data from multiple genes in two sister species, we are using single genes from multiple species pairs that arose at assumed times to estimate a DNA substitution rate and ancestral effective population sizes. Although the model that we specify is for haploid uniparental genes such as mitochondrial genes, the model can be modified and be applied to nuclear autosomal genes.

To illustrate our approach, we use two datasets of multiple sister species pairs (*Echinoidea* and *Alpheus*) that we will assume to have arisen from the same geologically dated barrier to gene flow, the Isthmus of Panama, three million years ago (Bermingham and Lessios 1993; Knowlton et al. 1993; Knowlton and Weigt 1998; Lessios et al. 1999; McCartney et al. 2000; Lessios et al. 2001b). Both the moments and likelihood methods assume that a group of closely related species all have the same DNA substitution rate. The likelihood method also allows us to test whether ancestral population sizes significantly differed among species pairs. The methods use data from multiple two-tip trees (one individual sampled from each extant species) and therefore do not assume reciprocal monophyly. The methodology presented here is appropriate for estimating rates of DNA substitution when the speciation events are within an order of magnitude of ancestral population size. Using older events such as the breakup of Gondwana or ancient fossil specimens to estimate DNA substitution rates requires a different set of statistical models (Huelsenbeck et al. 2000; Tavaré et al. 2002). Although coalescent methodology for estimating rates of DNA substitution is available, these methods require knowledge of the population size (Lundstrom et al. 1992; Kuhner et al. 1998) or assume that ancestral population sizes are the average of the effective population sizes estimated in the extant sister species (Nei and Li 1979).

METHODS

Model of Allopatric Divergence

Our model of allopatric divergence presented here is based on haploid genes that are transmitted through uniparental lineages. Our model can be applied to autosomal nuclear loci, except that when there are equal sex ratios the effective population size is four times the effective population size of haploid uniparental loci. Our model is divided into two time periods, τ , the number of generation since two species have diverged, and T , the number of generations two haploid genetic lineages sampled in the two sister species will merge prior to τ (see Fig. 1). The time of divergence τ is assumed to be known. When the ancestral population is panmictic, the

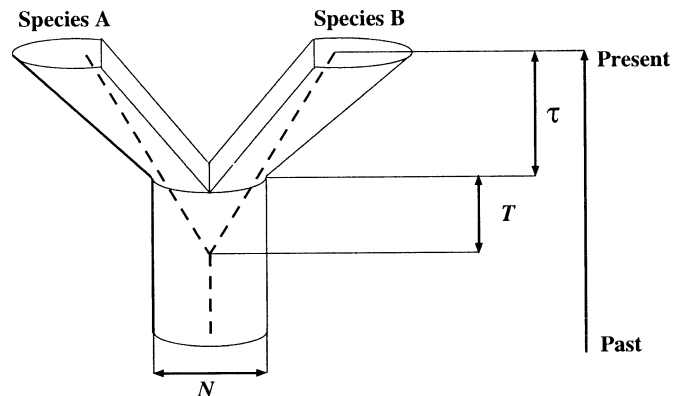


FIG. 1. A model of a species splitting into two sister species (A and B) at time τ generations in the past. The gene divergence (dashed line) predates the species divergence by T generations. The effective size of the panmictic ancestral population is N females (mitochondrial gene).

distribution of T is approximately exponential. Accordingly, the mean and variance of T are

$$E(T) = N \quad \text{and} \quad (1)$$

$$\text{Var}(T) = N^2, \quad (2)$$

where N is the effective female population size (mitochondrial loci) of the ancestral panmictic population (Hudson 1990). An additional source of variance is the stochastic mutational process. Because this model assumes that the mutation rate per site per generation is small ($\mu < 10^{-7}$), we use the infinite sites model. The mutations occur at n nucleotide sites within a locus and follow a Poisson process under a molecular clock. The expected number of mutations that accrue between a pair of sister species, k , can be calculated as a function of

$$E[k] = 2\mu n(\tau + N), \quad (3)$$

where μ is the number of DNA substitutions per site per generation, assuming there are no multiple mutations per site. Incorporating both the MRCA and mutation, the total variance in k is

$$\begin{aligned} \text{Var}[k] &= E[\text{Var}[k | T]] + \text{Var}[E[k | T]] \\ &= E[2\mu n\tau + 2\mu nT] + \text{Var}[2\mu n\tau + 2\mu nT] \\ &= 2\mu n\tau + 2\mu nN + 4(\mu nN)^2. \end{aligned} \quad (4)$$

This model assumes that the ancestral species consisted of a single panmictic population.

This model of allopatric divergence has been previously used to estimate the ancestral population size and divergence time between humans and other hominid lineages (Takahata et al. 1995; Takahata and Satta 1997; Yang 1997). Following these authors' work, we calculate the conditional probability of observing k substitutions between a pair of sister species separated τ generations ago given N and μ ,

$$\text{Pr}(k | N, \mu) = \int_0^{\infty} \frac{e^{-2\mu n(\tau+T)} [2\mu n(\tau+T)]^k e^{-T/N}}{Nk!} dT. \quad (5)$$

This conditional probability is a function of both the expo-

nential and Poisson processes, signifying the independent stochastic affects of the MRCA and mutation, respectively.

Moments Estimator

The average number of DNA substitutions between different sister species pairs is often used to estimate μ within a taxonomic group, even though genetic divergence is expected to predate the known time at which a biogeographic barrier arose by T generations. Ideally, one would want to correct for this coalescent time within the ancestral populations to obtain an unbiased estimate of μ . Although this can be done by using Nei and Li's (1979) net genetic divergence, the correction it employs assumes that ancestral polymorphism equals the average polymorphism in the two extant sister species. Takahata's (1986) approach, in contrast, explicitly uses the variance in ancestral coalescence times to estimate ancestral population parameters.

Takahata's (1986) method estimates the effective population size of the common ancestor of two extant species by using the difference between the sample mean and sample variance in the number of pairwise differences found in different genes. Whereas Takahata (1986) was interested in estimating the ancestral population size and the species divergence time between one pair of species using sequence data from multiple genes, we are interested in estimating the DNA substitution rate from multiple species pairs that were separated at assumed times. Following Takahata's (1986) approach, we use the difference between the sample mean and sample variance in the numbers of pairwise differences between multiple sister species pairs separated at τ generations ago to estimate μ . Assuming that all species pairs arose at time τ , generation time is equal among species, and the ancestral population of the species pairs were of equal size and panmictic, the expected difference between the variance and mean in the number of substitutions k can be obtained from equations (3) and (4), to yield

$$\frac{\text{Var}[k] - E[k]}{n^2} = (2\mu N)^2. \quad (6)$$

Thus, an estimate of μ is simply achieved by equating the sample variance ($\text{Var}_s[k]$) and sample mean ($E_s[k]$) with $\text{Var}[k]$ and $E[k]$ in equation (6) to first obtain an estimate of $\hat{\theta}$, (where θ is equal to $2N\mu$). Subsequently, $E_s[k]$ and $\hat{\theta}$ are used to obtain $\hat{\mu}$ from equation (3), such that

$$\hat{\mu} = \frac{E_s[k] - \hat{\theta}n}{2n\tau}. \quad (7)$$

Likelihood Estimator

Another means of estimating DNA substitution rates is to use a likelihood approach, similar to the ones developed by Takahata and Satta (1997) and Yang (1997). Again, while they were interested in estimating the ancestral population size and the species divergence time between one pair of species using sequence data from multiple genes, we are interested in estimating the DNA substitution rate for a single gene from multiple species pairs that were separated at τ . The likelihood of parameters given the observation of a particular event is, by definition, proportional to the probability

of observing that event. From equation (5), it follows that the likelihood function for observing k substitutions given N and μ is

$$L(\mu, N | k) = \int_0^\infty \frac{e^{-2\mu n(\tau+T)} [2\mu n(\tau+T)]^k e^{-T/N}}{Nk!} dT. \quad (8)$$

When analyzing data from multiple species pairs, we can choose a number of different models. Here we analyzed two extreme cases. In one case we analyzed the data under the biologically reasonable assumption that the size of each of the ancestral populations varied independently among species pairs (independent- N model). In the other case we analyzed the data under the assumption that ancestral population sizes were equal among species pairs (uniform- N model). For the i th of x total species pairs, let N_i and k_i be ancestral population size and observed numbers of substitutions since the MRCA, respectively. We denote the vector (N_1, N_2, \dots, N_x) by \vec{N} , and the vector (k_1, k_2, \dots, k_x) by \vec{k} . The likelihood function for the independent- N model is

$$L(\mu, \vec{N} | \vec{k}) = \prod_{i=1}^x \int_0^\infty \frac{e^{-2\mu n(\tau+T)} [2\mu n(\tau+T)]^{k_i} e^{-T/N_i}}{N_i k_i!} dT, \quad (9)$$

which can be transformed into the finite summation,

$$L(\mu, \vec{N} | \vec{k}) = \prod_{i=1}^x \frac{e^{-2n\mu\tau}}{2Nn\mu + 1} \sum_{j=0}^{k_i} \frac{1}{j!} \left(\frac{2Nn\mu}{2Nn\mu + 1} \right)^{k_i - j} (2Nn\mu\tau)^j. \quad (10)$$

The finite summation formulation in (10) is generally easier to evaluate numerically than the indefinite integral formulation in (9). Note that the uniform- N model is a special case of the independent- N model such that $N_1 = N_2 = \dots = N_x = N$. We found the maximum-likelihood estimate (MLE) of μ using the following approach. We first began with an initial value for μ and then for each species pair x , numerically searched for the MLE of N_i given k_i and our chosen value of μ . Using this approach we then searched for the μ value with the maximum likelihood value.

One of the problems with using the independent- N model is that there will be more parameters than datapoints (μ and one N per species pair). Although population sizes certainly varied among ancestral species, it is unclear whether differences among ancestral population sizes have had a significant effect on k . Before using the uniform- N model, one must demonstrate that the independent- N model does not significantly fit the data better than the uniform- N model (Hilborn and Mangel 1997). The two models can be explicitly compared using a likelihood-ratio test to see if the estimation of μ using the uniform- N model can be justified. Approximately two times the negative log of the likelihood ratio of the two models is asymptotically χ^2 distributed with the number of degrees of freedom (df) equal to the difference in the number of estimated parameters between the two models. In our case, this is the number of species pairs minus one (Hilborn and Mangel 1997).

Simulations

We examined the reliability of our estimators under different parameter values for ancestral population sizes using

TABLE 1. Summary of the means and standard deviations of the moments (A) and likelihood (B, C) estimators of μ and N on sets of 100 replicate simulations of samples of x species pairs. The simulations were generated using the following fixed parameters: $\mu = 2.00 \times 10^{-8}$ per site per generation, generation time of two years, 500 base pairs, and divergence time τ of 1.5×10^6 generations. N is the effective size of the female ancestral population and was either fixed at 500,000 or was drawn from a uniform distribution ranging from 50,000 to 1 million individuals. The average and standard deviation of each set of 100 estimates is under the columns for $\hat{\mu}$, and SD.

(A) Simulation conditions		$\hat{\mu}$	SD	\hat{N}	SD
$N = 500,000$	$x = 5$	2.10×10^{-8}	3.82×10^{-9}	442,672	424,337
	$x = 10$	2.06×10^{-8}	2.68×10^{-9}	449,400	260,076
	$x = 20$	2.04×10^{-8}	1.80×10^{-9}	448,441	193,359
	$x = 40$	2.04×10^{-8}	1.41×10^{-9}	462,726	152,406
$N = 50,000$ to 1 million	$x = 5$	1.93×10^{-8}	3.88×10^{-9}	1,490,054	1,991,906
	$x = 10$	1.92×10^{-8}	3.01×10^{-9}	1,435,000	1,659,677
	$x = 20$	1.88×10^{-8}	2.55×10^{-9}	1,011,666	2,955,557
	$x = 40$	1.85×10^{-8}	1.84×10^{-9}	960,315	7,709,001
(B) Independent- N model		$\hat{\mu}$	SD	\hat{N}	SD
$N = 500,000$	$x = 5$	2.23×10^{-8}	3.28×10^{-9}		
	$x = 10$	2.17×10^{-8}	2.00×10^{-9}		
	$x = 20$	2.19×10^{-8}	1.32×10^{-9}		
	$x = 40$	2.19×10^{-8}	1.09×10^{-9}		
Uniform- N model					
$N = 500,000$	$x = 5$	2.19×10^{-8}	3.74×10^{-9}	354,581	377,551
	$x = 10$	2.06×10^{-8}	2.84×10^{-9}	411,013	286,781
	$x = 20$	2.01×10^{-8}	1.77×10^{-9}	511,733	204,192
	$x = 40$	2.01×10^{-8}	1.09×10^{-9}	492,915	132,669
(C) Independent- N model		$\hat{\mu}$	SD	\hat{N}	SD
$N = 50,000$ to 1 million	$x = 1$	2.73×10^{-8}	8.80×10^{-9}		
	$x = 5$	2.23×10^{-8}	3.49×10^{-9}		
	$x = 10$	2.18×10^{-8}	2.07×10^{-9}		
	$x = 20$	2.16×10^{-8}	1.93×10^{-9}		
	$x = 40$	2.14×10^{-8}	8.74×10^{-10}		
Uniform- N model					
$N = 50,000$ to 1 42million	$x = 5$	2.12×10^{-8}	3.41×10^{-9}	414,101	452,606
	$x = 10$	2.02×10^{-8}	2.83×10^{-9}	536,862	401,274
	$x = 20$	1.94×10^{-8}	1.64×10^{-9}	566,583	239,478
	$x = 40$	1.91×10^{-8}	1.37×10^{-9}	630,918	181,177

computer simulations of allopatric speciation. Additionally, we examined how sample size affected our estimator of μ as well as how reliable these estimators were when some assumptions were violated.

In each simulation of a sister species pair diverging at a common time in the past τ , and a given set of parameters N and μ , we generated a random variable T from an exponential distribution with a mean N . We then drew a Poisson random variable with a parameter value equal to $2\mu n(\tau + T)$, which is the expected number of mutations accruing on the branches leading to the two extant genetic lineages. This was stored as a value of k DNA substitutions for each simulated species pair. Samples of replicate sister species pairs were generated in which ancestral N was held constant and in which N was drawn from a uniform distribution ranging from 50,000 to one million individuals. The value of μ and range of N values reflected what is often found in empirical phylogeographic studies.

We explored the properties of the moments and likelihood estimators of μ and N ($\hat{\mu}$ and \hat{N}) using simulated data generated from a panmictic model that either specified equal population sizes ($N = 500,000$ individuals) or allowed ancestral population sizes to vary uniformly from 50,000 to one million individuals. The estimator under each set of as-

sumptions was applied to 100 simulated replicate datasets of sample size five, 10, 20, and 40 species pairs (see Table 1).

RESULTS

Properties of the Moments Estimator

When the ancestral population sizes are equal among species pairs, the moments estimator of μ converges to be within 2% of the true value at a sample size of 20 species pairs, while the estimator of N has a slight downward bias (see Table 1A). These slight biases asymptotically disappear with larger but unrealistic sample sizes (e.g., 1000 species pairs). When the simulated ancestral population sizes varied, the estimator of μ had a downward bias (see Table 1A), while the estimate of effective population size had a larger upward bias. Overall, the moment estimator of μ has a downward bias when the assumption of uniform ancestral population sizes is violated, but it performs well.

Properties of the Likelihood Estimator

The likelihood estimates of μ and N were obtained given both the uniform- N and independent- N models (see Tables 1B, C). Examples of likelihood surfaces from simulated data

given the uniform- N model are shown in Figures 2A–C. Likelihood-ratio tests rarely showed the independent- N model to significantly fit the simulated datasets better than the uniform- N model given our simulation conditions. The likelihood estimates of μ based on the uniform- N model tended to be more accurate than those based on the over parameterized independent- N model, even when simulated N varied (see Table 1B, C). However, when using the uniform- N model on data simulated with variation in N , a downward bias in $\hat{\mu}$ and an upward bias in \hat{N} is apparent with increasing sample size (see Table 1C). This is consistent with the biases in the moments estimator when the simulated N varied. In these cases, the downward bias is due to $\hat{\mu}$ being confounded with \hat{N} . Therefore true variation in ancestral N will tend to yield overestimates of N if one assumes the uniform- N model, and consequently μ will be underestimated. Estimates of μ based on the independent- N model were consistently upwardly biased (see Tables 1B, C).

Application to Data from Panama Geminate Species Pairs

One of the most notable geologic events allowing a DNA substitution rate to be estimated from the allopatric splitting of multiple species is the emergence of the Panamanian Isthmus. Jordan (1908) coined the term “geminate” species to describe marine taxa in both oceans that arose through allopatric speciation by the emergence of the land bridge, and he viewed these replicate speciations as a valuable natural biogeographic experiment. This barrier is thought to have separated marine species into the Pacific and Atlantic Oceans approximately three million years ago (Farrell et al. 1995; Coates and Obando 1996; Collins 1996), although potential for gene flow between the two oceans may have occurred as recently as two million years ago (Cronin and Dowsett 1996). Not surprisingly, DNA sequence data has been gathered from geminate species since the dawn of the molecular genetic era (Lessios 1979; Vawter et al. 1980).

Here we apply the moments and likelihood estimators to data gathered from eight geminate pairs of sea urchins (Echinoids; Bermingham and Lessios 1993; Lessios et al. 1999; McCartney et al. 2000; Lessios et al. 2001b; Lessios et al. 2003; Z. S. Zigler and H. A. Lessios, unpubl. ms.) and 15 geminate pairs of the snapping shrimp (*Alpheus*; Knowlton et al. 1993; Knowlton and Weigt 1998). The amounts of observed variation in average genetic divergence between species pairs ranges from under three-fold in the sea urchins to nearly seven-fold in *Alpheus* (Bermingham and Lessios 1993; Knowlton et al. 1993; Knowlton and Weigt 1998; Lessios 1998; Lessios et al. 1999, 2001b; McCartney et al. 2000; Lessios et al. 2003; M. J. Hickerson, Z. S. Zigler, and H. A. Lessios, unpubl. ms.; Z. S. Zigler and H. A. Lessios, unpubl. ms.). The observed variation in the sea urchins and *Alpheus* is greater than expected under a Poisson molecular clock given a single genetic divergence time. Consequently, previous researchers have explained the observed variation by subsequently proposing that either μ varies between species (Bermingham and Lessios 1993), or τ varies between species pairs (Knowlton and Weigt 1998; Lessios et al. 2001b). Variation in τ can also result from incorrect geminate species status due to possible extinction of any true geminate species

(Jackson et al. 1993; Roopnarine 2001; Marko 2002) as well as gene flow occurring as recently as two million years ago in a subset of the species (Cronin and Dowsett 1996). Variation in μ could also occur if there are large differences in the number of germ-line replications per generation. Although these post hoc explanations are possible, we show that they are unnecessary if one considers variation in coalescent times within the ancestral populations. Given variation in the coalescence parameter (N) among ancestral species, the independent- N and uniform- N models presented here could be sufficient to explain the observed variation.

The sea urchin sequence data used to demonstrate this methodology all came from a homologous cytochrome oxidase I (COI) region of mtDNA that was approximately 640 bp in length (Lessios et al. 1999; McCartney et al. 2000; Lessios et al. 2001b; M. J. Hickerson, Z. S. Zigler, and H. A. Lessios, unpubl. ms.; Lessios et al. 2003; Z. S. Zigler and H. A. Lessios, unpubl. ms.). For the urchins, we assume that the pairs of sampled taxa arose from the rise of the Panamanian Isthmus three million years ago. Because there has been speciation since the rise of the isthmus in some of the genera, some of the pairs of sampled taxa are subsets of geminate clades rather than subsets of geminate species. The estimates were conducted on the observed number of DNA base substitutions between geminate species pairs as well as the number of substitutions corrected for multiple hits using the Kimura’s (1980) two-parameter model following Lessios et al. (2001a).

By applying equation (6) and (7), the moments estimate of μ for the sea urchin geminates yields a DNA divergence rate of 1.85% per million years (1.95% per million years if corrected for multiple hits using Kimura’s two-parameter model). However, this moments estimator ignores variation in ancestral population size. Based on simulation results, $\hat{\mu}$ is likely to be an underestimate of μ (see Table 1A).

The likelihood surface of μ and N obtained from the eight sea urchin geminates is shown in Figure 2D. The independent- N model does not fit the data significantly better than the uniform- N ($P = 0.34$ and 0.36 for corrected and uncorrected data, respectively). The confidence region around the MLE of μ yields a DNA divergence rate between approximately 0.59% per million years and 2.07% per million years when using the uncorrected data. Lessios et al. (2001b) attributed the observed variation in genetic distances to two different divergence times thereby yielding estimates of the divergence rate ranging from 1.6% to 2.6% per million years. Although this confidence region might be satisfactory for distinguishing alternative hypothesized divergence times that differ by an order of magnitude, divergence times that are subtly different will be difficult to distinguish.

One way to improve upon this estimate of μ would be to incorporate prior information on some of the parameters, such as is done in Bayesian statistics. For example, if one assumes there is a predictive relationship between the estimates of N and μ in extant species and these parameters in ancestral species, then the observation of reciprocal monophyly in the sea urchin geminate pairs can be used to set an upward bound to a prior distribution of ancestral population sizes. Because haploid mtDNA loci are expected to reach reciprocal monophyly in $2N$ generations (Tajima 1983) and the divergence

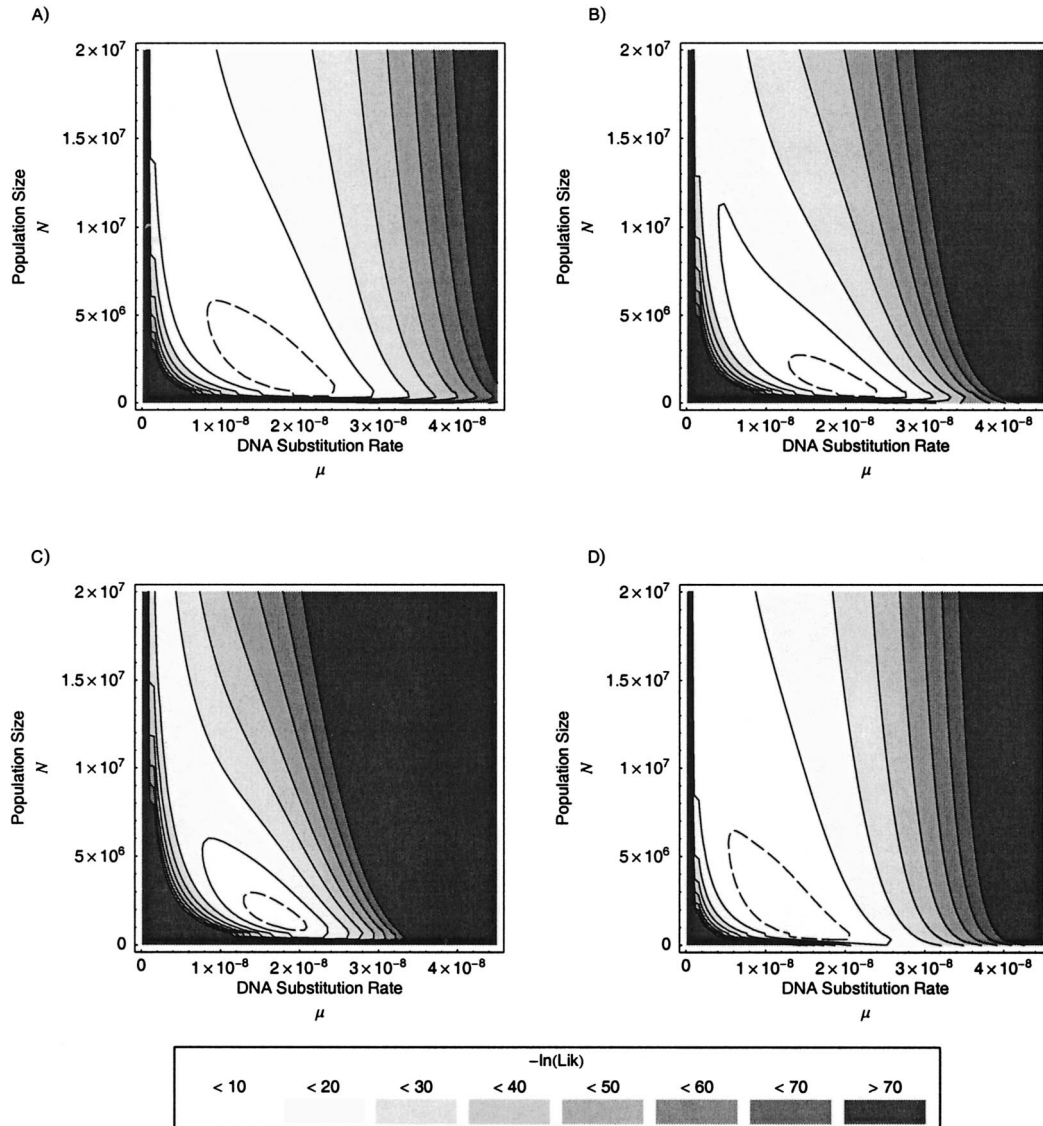


FIG. 2. Log likelihood surfaces of $\hat{\mu}$ and \hat{N} estimated from simulated data (A, B, and C) and eight sea urchin geminate species (D). Likelihood values were calculated using the uniform- N model in equations (9) and (10). Panels (A), (B), and (C), are based on simulated data of 10, 20, and 40 species pairs, respectively, that were separated from panmictic ancestral populations 1.5 million generations in the past. Simulated data was based on a model of 500 nucleotides (mtDNA), a generation time of two years, a DNA mutation rate, μ , of 2.0×10^{-8} per site per generation, and the panmictic ancestral populations were drawn from a uniform distribution of 50,000 to one million females (mitochondrial gene). Panel (D) is based on data from eight sea urchin species pairs that are assumed to have been separated approximately 1.5 million generations in the past. In all panels, the dashed contour lines represent approximate 95% confidence regions determined using likelihood-ratio tests.

time of these geminate species pairs is assumed to be 1.5×10^6 generations (given an average generation time of two years), 750,000 individuals is a conservative upper limit for ancestral population sizes. Although this upper limit is based on a model with no growth, any bottlenecks that could have occurred since species divergence would have increased the likelihood of reciprocal monophyly (Lessios et al. 2001a). This is because under a model with population growth, the distribution of coalescent times is consistent with a distribution given a model of smaller population sizes and no population growth (Kuhner et al 1998). We are currently developing a Bayesian estimator that will incorporate a prior distribution for ancestral population size.

To further demonstrate these methods, we estimated a DNA substitution rate from an approximately 564-bp region of the mtDNA COI gene from 15 *Alpheus* species pairs that were sequenced in previous studies (Knowlton et al. 1993; Knowlton and Weigt 1998). Application of these two estimators to the *Alpheus* dataset highlights where this method is less appropriate due to excessive variance in the data. The moments estimator of μ for the *Alpheus* geminates yielded a DNA divergence rate of 1.80% per one million years. On the other hand, the likelihood approach identified that this dataset might be too variable to estimate μ with as much confidence as we did in the sea urchins because the uniform- N model was rejected in certain cases.

TABLE 2. (A) The estimates of the two parameters ($\hat{\mu}$, \hat{N}) from the 15 geminate species pairs of snapping shrimp (*Alpheus*) given models 1 through 4. Models 1 and 3 assume that all 15 geminates diverged at three million years ago. Given models 2 and 4, $\hat{\tau}$ is estimated for the two mangrove geminates while assuming that the remaining 13 geminates diverged at three million years ago. Models 1 and 2 assume that ancestral population sizes (N) were equal among the 15 species pairs (uniform- N model). Models 3 and 4 assume that this parameter varies independently amongst the 15 geminates (independent- N model). (B) The results of the likelihood-ratio tests among nested models. In parentheses indicate the level of significance (P) and degrees of freedom (df) used in the likelihood-ratio tests. Approximately negative two times the log of the likelihood ratio of the two compared models is asymptotically χ^2 distributed with the number of degrees of freedom (df) equal to the difference in the number of estimated parameters between the two models.

(A)							
Model	Number of assumed τ values	Number of assumed N values	-Log likelihood	$\hat{\mu}$	$\hat{\tau}$	\hat{N}	
1	1	1	73.05	1.98×10^{-8}	n/a	1.87×10^6	
2	2	1	70.86	2.40×10^{-8}	1.66×10^6	1.37×10^6	
3	1	15	61.80	2.51×10^{-8}	n/a	$1.32 \times 10^{6*}$	
4	2	15	58.75	2.81×10^{-8}	1.92×10^6	$1.64 \times 10^{6**}$	
						$1.26 \times 10^{6**}$	
						$1.47 \times 10^{6**}$	
(B)							
Model	1	2	3	4			
1	—	$P = 0.036$; df = 1	$P = 0.069$; df = 14	$P = 0.018$; df = 15			
2	—	—	—	$P = 0.043$; df = 14			
3	—	—	—	$P = 0.014$; df = 1			
4	—	—	—	—			

* Average.

** Standard deviation.

In addition to comparing the uniform- N to the independent- N model assuming that τ was uniform (three million years ago) among the 15 species pairs, we compared these two models given that τ was uniform (three million years ago) among the 13 deeper water species pairs while τ was estimated in the two mangrove species pairs. This was done to address whether the two mangrove species pairs diverged as recently as two million years ago as has been suggested (Cronin and Dowsett 1996; Knowlton and Weigt 1998). In all four models, a single μ was estimated from all 15 species pairs. These four models were compared with each other using likelihood-ratio tests when the comparisons involved nested models (see Table 2). These data from *Alpheus* were also analyzed before and after correcting for multiple hits following the author's methodology (Knowlton and Weigt 1998).

Given either the uniform or independent- N models, τ for the two mangrove species was found to be significantly lower than three million years ago and the MLE of this τ was lower than two million years ago (see Table 2). If we compared uniform and independent- N models while estimating τ for the two mangrove species and μ for all 15 species pairs, the ancestral population sizes did significantly differ ($P = 0.043$). If we use a P -value of 0.05 as a criterion for significance, then we are forced to use the over parameterized and heavily biased independent- N model for estimating μ .

These nested model comparisons suggest that there is weak support for using the uniform- N model estimator of μ in *Alpheus*. They also suggest that estimates of μ are very sensitive to which statistical model is used. If we assume that τ was uniform for all species pairs, then the MLE estimates of the divergence rates are 1.98 % per million years given the uniform N model and 2.51% per million years given the independent- N model, respectively. If we estimate τ for the two mangrove species pairs, then the MLE estimates of the

divergence rates are 2.40% per million years given the uniform- N model and 2.81% per million years given the independent- N model, respectively. However, the weak support for using the uniform- N model suggests that there is excessive variation in the parameters influencing k , the number of substitutions among geminate species pairs.

Although there is a large disparity between the various estimates of μ , they are higher than Knowlton and Weigt's (1998) estimate of 1.4% per million years. Knowlton and Weigt (1998) obtained this estimate based on the pair of *Alpheus* that had the lowest sequence divergence under the assumption that this pair diverged three million years ago and subsequently used this estimate to conclude that the remaining 14 pairs diverged at earlier dates without incorporating ancestral coalescence. However, given the high level of observed variation, data from the *Alpheus* species pairs are probably not suitable for estimating μ without making further assumptions regarding the various parameters influencing k , such as assuming that N is gamma distributed with the shape parameter being estimated. This strategy is currently being developed.

DISCUSSION

Recent empirical studies suggests that DNA substitution rates are similar among closely related taxa, even if they have different generation times (Kumar and Subramanian 2002). Thus, the methods presented here could have wide applicability. The model presented here is related to earlier coalescence models used for estimating the divergence time and ancestral population size of two recently diverged species such as humans and chimpanzees (Takahata 1986; Takahata et al. 1995; Takahata and Satta 1997; Yang 1997). In addition to being useful for estimating parameters from multispecies phylogeographic datasets, our model can be expanded for use

in clinical scenarios, such as accounting for inpatient divergence when estimating the rate of DNA substitution in HIV (Leitner and Albert 1999).

Although μ could be considered the least interesting of ancestral population parameters to investigate compared with ancestral population sizes and divergence times, these latter parameters rely on estimates of μ . The simulation analysis suggests several strategies for estimating μ . When applying the likelihood estimate of μ to test other phylogeographic hypotheses, caution should be exercised by directly using the confidence region to demonstrate if rejecting a hypothesis is robust to the range of uncertainty in the estimation of μ . It may be that competing phylogeographic hypotheses which differ in regard to the timing of evolutionary events will only be distinguishable if the alternative hypotheses differ by orders of magnitude with regard to dating of events. Indeed, even when a single DNA substitution rate is biologically plausible, estimates of divergence time that incorporate this estimate will also have wide confidence intervals because of the heavily stochastic nature of the coalescent and Poisson process (Edwards and Beerli 2000; Nielsen and Wakeley 2001).

The likelihood estimator of μ accommodates some variation in the size of the ancestral populations and hence should also accommodate some variation in generation times among species because these two parameters scale together. For example, the distribution of T in species with small population sizes and long generation times should be similar to species that have large ancestral populations and short generation times. If μ is in fact equal among a group of species pairs, then variation in generation time will have the same effect on k as variation in ancestral population size. However, if generation time varies and μ is correlated with generation time due to different numbers of germ-line replications per generation, then the coalescent model we use here would clearly break down.

Although we chose to ignore possible size changes in the ancestral population, the extra parameters used in a size change model would not have greatly improved our estimation of μ . Kuhner et al. (1998) showed that the parameters of population size and of exponential growth rate are heavily correlated and that a growth model is difficult to statistically reject when there really is no growth. Furthermore, bottlenecks in ancestral populations will improve our method by decreasing the expectation and variance in coalescent times among species pairs.

The contrasting datasets of the sea urchins and *Alpheus* illustrate the varying degrees to which our likelihood method is appropriate. The method is more applicable to the former dataset because we are more confident in our assumptions about the parameters that have influenced the observed variation in k . Regarding the assumption of τ , the sea urchin geminates most likely all arose at the same time because they share dispersal characteristics. However, there is weak support for using the uniform- N model for the *Alpheus* data, because there are too many violations in the assumptions of our model. For example, while the assumption of panmixia is probably not violated in the sea urchin, the higher possibility of ancestral subdivision in *Alpheus* could have hindered our ability to estimate the parameters (McClure and Green-

baum 1999; but for example of *Alpheus* species without subdivision see Mathews et al. 2002). Under a multiple-deme model ancestral subdivision will be manifested as excessive variance in k , while under a two-deme model μ will be overestimated (Nath and Griffiths 1993; Wakeley 2000; Rosenberg and Feldman 2002). Another plausible reason for elevated variation is that geminate species status is incorrect due to extinction of true geminate species (Jackson et al. 1993; Roopnarine 2001; Marko 2002).

Despite the potential for there being multiple causes for the elevated variation in k in such datasets, it is not the purpose of this work to claim that models that are more complicated than the uniform or independent- N model do not fit geminate datasets better than these two models. Instead of making such an erroneous claim, we attempt to demonstrate that the simpler uniform- N model can often fit geminate data sufficiently and yield good estimates of μ , even when N actually varied among closely related ancestral species. Although variation in a number of parameters (τ , μ , N , or subdivision) or incorrect geminate status could all be contributing to the large empirical variance in the *Alpheus* dataset, it is reasonable to first use a model allowing variation in N because it is usually the parameter that is most biologically plausible to vary among closely related species pairs. Thus, we first compare the independent and uniform- N models. Although we do not explicitly compare the uniform- N model to a variable μ model, the correlation of N and μ apparent from equation (3) and (4) means that our model comparison is relevant to evaluating a variable- μ model. Data that is sufficiently variable to cause rejection of the uniform- N model when compared to the independent- N model would also likely cause rejection of the uniform- N model if compared it to a variable- μ model. Likewise, we did not explicitly compare a uniform parameter model to a model in which τ varied among all species pairs. Although τ is not correlated with N in our model, large variation in τ would also result in inflated variation in k . This would likely result in the uniform- N model being rejected when compared to the independent- N model using the likelihood-ratio test. It is important to note that the assumption of τ occurring at three million years ago and being uniform among all species pairs is certainly an oversimplification. Cessation of transisthmus gene flow likely occurred at different times between 2–3.5 million years ago such that τ likely had some variation, even if dispersal characteristics were largely similar such as in the sea urchins.

Regardless of the cause for inflated variation in k , the uniform- N model would usually be rejected in such cases and we would not be justified in using estimates based on the uniform- N model. As in the case of the *Alpheus* geminate pairs, we are forced to use the more biased estimator based on the overparameterized independent- N model in which there are more parameters than datapoints. In contrast, sea urchins are probably largely panmictic and might lack substantial variation in τ , μ , or N . Consequently, it is not surprising that we could justify using the simpler and less biased uniform- N model to estimate μ . However, a comparison between the uniform- N and the overparameterized independent- N model using the likelihood-ratio test could violate the assumptions of the χ^2 approximation because the latter model

is one in which parameters are added as more datapoints are collected.

To make this methodology applicable to the greatest number of phylogeographic datasets, our model uses one individual per species instead of incorporating intraspecific information. One could expand our model such that it uses data from multiple alleles in the descendent species, thereby decreasing the variance in the estimate of μ . In the simple model presented here, $E[k] = 2\mu n(\tau + N)$, τ could be partitioned to include estimates of $2N\mu$ obtained from each descendent species (Kuhner et al. 1998). This strategy is currently being developed for the analysis of intraspecific data from geminate species pairs.

Although this method does not explicitly account for multiple mutations at single sites, intragene rate heterogeneity has a negligible importance at the time scale at which the estimators are applicable (Takahata and Satta 2002). Indeed we found that correcting for multiple hits using the best-fit model of evolution from the sequence data (Posada and Crandall 1998), makes little difference. It should also be noted that this methodology assumes that the DNA sequences used to estimate μ have evolved neutrally. However, the moments and likelihood estimators partly accommodate selection in the ancestral species because they estimate the rate of genetic change after the screening of selection. Balancing or positive selection (including hitchhiking effect) occurring in the ancestral species could be accommodated because these would either increase or decrease respectively the expected variance in k among the ancestral species. If selection causes an excessive variance in k , then our method would be shown to be inappropriate from the likelihood-ratio test. Furthermore, if either of these forms of selection differentially occur in the ancestral and descendent species, then the likelihood models used here would be invalid.

Despite the potential difficulties associated with selection and the various processes that can contribute to observed variation in k , it is encouraging that our simulation study shows that reasonable estimates of μ can be obtained from fewer than 10 species pairs when the assumption of uniform ancestral population sizes is violated. When estimating μ from multiple species pairs that exhibit variation in k , we have shown that much variation can be accommodated with a simple coalescent model in which N can vary. It is most biologically reasonable to invoke variation in N among related taxa and should be explored before invoking variation in the other parameters that can influence variation in k .

ACKNOWLEDGMENTS

We thank H. Lessios and K. Zigler for useful discussions and providing sea urchin DNA sequence data. For key discussions at the initial stages of the study, we thank M. Levine and J. Niegel. We thank S. Baird, C. Cunningham, C. Moritz, Y. Satta, M. Slatkin, N. Takahata, J. Wares, and two anonymous reviewers for discussion or comments on the manuscript. Support for MJH was provided by National Science Foundation Dissertation Enhancement Grant NSF-INT-02-03094. Support for MAG was provided by the U.S. Department of Energy under contract W-7405-ENG-36. Mathematica routines that estimate μ using the likelihood methods

described here are available from MAG, and the C code used for the simulations is available from MJH.

LITERATURE CITED

- Avise, J. C. 2000. *Phylogeography: the history and formation of species*. Harvard Univ. Press, Cambridge, MA.
- Bermingham, E., and H. A. Lessios. 1993. Rate variation of protein and mitochondrial DNA evolution as revealed by sea urchins separated by the Isthmus of Panama. *Proc. Natl. Acad. Sci. USA* 90:2734–2738.
- Bermingham, E., S. S. McCafferty, and A. P. Martin. 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. Pp. 113–128 in T. D. Kocher and C. A. Stepien, eds. *Molecular systematics of fishes*. Academic Press, San Diego, CA.
- Coates, A. G., and J. Obando. 1996. The geological evolution of the Central American Isthmus. Pp. 21–56 in J. B. C. Jackson, A. F. Budd, and A. G. Coates, eds. *Evolution and environment in tropical America*. Univ. of Chicago Press, Chicago, IL.
- Collins, T. M. 1996. Molecular comparisons of transisthmian species pairs: rates and patterns of evolution. Pp. 303–334 in J. B. C. Jackson, A. F. Budd, and A. G. Coates, eds. *Evolution and environment in tropical America*. Univ. Chicago Press, Chicago, IL.
- Cronin, T. M., and H. J. Dowsett. 1996. Biotic and oceanographic response to the Pliocene closing of the central American Isthmus. Pp. 76–104 in J. B. C. Jackson, A. F. Budd, and A. G. Coates, eds. *Evolution and environment in tropical America*. Univ. of Chicago Press, Chicago, IL.
- Cunningham, C. W., and T. M. Collins. 1994. Developing model systems for molecular biogeography: vicariance and interchange in marine invertebrates. Pp. 405–433 in B. Schierwater, B. Streit, G. Wagner, and R. DeSalle, eds. *Molecular ecology and evolution: approaches and applications*. Birkhauser Verlag, Basel, Switzerland.
- Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54:1839–1854.
- Farrell, J., I. Raffi, T. R. Janecek, D. W. Murray, M. Levitan, K. A. Dadey, K.-C. Emeis, M. Lyle, J.-A. Flores, and S. Hovan. 1995. Late Neogene sedimentation patterns in the eastern equatorial Pacific Ocean. *Proc. Ocean Drill. Prog. Sci. Res.* 138: 717–756.
- Fleischer, R. C., C. E. Mcintosh, and C. L. Tarr. 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Mol. Ecol.* 7:533–545.
- Hellburg, M. E., D. P. Balch, and K. Roy. 2001. Climate driven range-expansion and morphological evolution in a marine gastropod. *Science* 292:1707–1710.
- Hilborn, R., and M. Mangel. 1997. *The ecological detective: confronting models with data*. Princeton Univ. Press, Princeton, NJ.
- Hudson, R. R. 1990. Gene genealogies and the coalescent process. Pp. 1–44 in D. Futuyma and J. Antonovics, ed. *Oxford Surveys in Evolutionary Biology*. Vol. 7. Oxford Univ. Press, Oxford, U.K.
- Huelsenbeck, J. P., B. Larget, and D. A. Swofford. 2000. A compound Poisson process for relaxing the molecular clock. *Genetics* 154:1879–1892.
- Jackson, J. B. C., P. Jung, A. G. Coates, and L. S. Collins. 1993. Diversity and extinction of tropical American mollusks and emergence of the Isthmus of Panama. *Science* 260:1624–1626.
- Jordan, D. S. 1908. The law of the geminate species. *Am. Nat.* 42: 73–80.
- Kambysellis, M. P., K.-F. Ho, E. M. Craddock, F. Piano, M. Parisi, and J. Cohen. 1995. Patterns of ecological shifts in the diversification of Hawaiian *Drosophila* inferred from molecular phylogeny. *Curr. Biol.* 5:1129–1139.
- 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.

- Kingman, J. F. 1982a. The coalescent. *Stoch. Proc. Appl.* 13: 235–248.
- . 1982b. On the genealogy of large populations. *J. Appl. Prob.* 19:27–43.
- Klicka, J., and R. M. Zink. 1997. The importance of recent ice ages in speciation: a failed paradigm. *Science* 277:1666–1669.
- Knowles, L. L. 2000. Tests of Pleistocene speciation in montane grasshoppers from the sky islands of western North America (genus *Melanoplus*). *Evolution* 54:1337–1348.
- Knowlton, N. 1993. Sibling species in the sea. *Annu. Rev. Ecol. Syst.* 24:189–216.
- Knowlton, N., and L. A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proc. R. Soc. Lond. B* 265:2257–2263.
- Knowlton, N., L. A. Weigt, L. A. Solorzano, D. K. Mills, and E. Bermingham. 1993. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. *Science* 260:1629–1632.
- Kuhner, M. K., J. Yamato, and J. Felsenstein. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149:429–434.
- Kumar, S., and S. Subramanian. 2002. Mutation rates in mammalian genomes. *Proc. Natl. Acad. Sci. USA* 99:803–808.
- Leitner, T., and J. Albert. 1999. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc. Natl. Acad. Sci. USA* 96:10752–10757.
- Lessios, H. A. 1979. Use of Panamanian sea urchins to test the molecular clock. *Nature* 280:599–601.
- Lessios, H. A., B. D. Kessing, D. R. Robertson, and G. Paulay. 1999. Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution* 53: 806–817.
- Lessios, H. A., M. J. Garrido, and B. D. Kessing. 2001a. Demographic history of *Diadema antillarum*, a keystone herbivore on Caribbean reefs. *Proc. R. Soc. Lond. B* 268:2347–2353.
- Lessios, H. A., B. D. Kessing, and J. S. Pearse. 2001b. Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution* 55:955–975.
- Lessios, H. A., J. Kane, and D. R. Robertson. 2003. Phylogeography of the pantropical sea urchin *Tripneustes*: contrasting patterns population structure between oceans. *Evolution* 57:2026–2036.
- Lundstrom, R., S. Tavaré, and R. H. Ward. 1992. Estimating substitution rates from molecular data using the coalescent. *Proc. Natl. Acad. Sci. USA* 89:5961–5965.
- Marko, P. B. 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Mol. Biol. Evol.* 19:2005–2021.
- Marko, P. B., and A. L. Moran. 2002. Correlated evolutionary divergence of egg size and a mitochondrial protein across the Isthmus of Panama. *Evolution* 56:1303–1309.
- Mathews, L. M., C. D. Schubart, J. E. Neigel, and D. L. Felder. 2002. Genetic, ecological, and behavioral divergence between two sibling snapping shrimp species (Crustacea: Decapoda: *Alpheus*). *Mol. Ecol.* 11:1427–1437.
- McCartney, M. A., G. Keller, and H. A. Lessios. 2000. Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus *Echinometra*. *Mol. Ecol.* 9: 1391–1400.
- McClure, M. R., and I. F. Greenbaum. 1999. Allozymic variation and biogeography of snapping shrimp (*Alpheus*) from the Gulf of Mexico and northwestern Atlantic coasts. *Southwest. Nat.* 44: 462–469.
- Nath, H. B., and R. C. Griffiths. 1993. The coalescent in two colonies with symmetric migration. *J. Math. Biol.* 31:841–852.
- Nei, M., and W. Li. 1979. Mathematical model for studying variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76:5269–5273.
- Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158: 885–896.
- Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Roderick, G. K., and R. G. Gillespie. 1998. Speciation and phylogeography of Hawaiian terrestrial arthropods. *Mol. Ecol.* 7: 519–531.
- Roopnarine, P. D. 2001. A history of diversification, extinction, and invasion in tropical America as derived from species-level phylogenies of chionine genera (family Veneridae). *J. Paleontol.* 75: 644–657.
- Rosenberg, N. A., and M. W. Feldman. 2002. The relationship between coalescence times and population divergence times. Pp. 130–164 in M. Slatkin and M. Veuille, eds. *Modern developments in theoretical population genetics*. Oxford Univ. Press, Oxford, U.K.
- Schneider, C. J., M. Cunningham, and C. Moritz. 1998. Comparative phylogeography and the history of endemic vertebrates in the wet tropics rainforests of Australia. *Mol. Ecol.* 7:487–498.
- Slatkin, M. 2000. Allele age and a test for selection on rare alleles. *Proc. R. Soc. Lond. B* 355:1663–1668.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.
- Takahata, N. 1986. An attempt to estimate the effective size of the ancestral species common to two extant species from which homologous genes are sequenced. *Genet. Res. Camb.* 48: 187–190.
- Takahata, N., and Y. Satta. 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. USA* 94: 4811–4815.
- . 2002. Pre-speciation coalescence and the effective size of ancestral populations. Pp. 52–71 in M. Slatkin and M. Veuille, eds. *Modern developments in theoretical population genetics*. Oxford Univ. Press, Oxford, U.K.
- Takahata, N., Y. Satta, and J. Klein. 1995. Divergence time and population size in the lineage leading to modern humans. *Theor. Popul. Biol.* 48:198–221.
- Tarr, C. L., and R. C. Fleischer. 1993. Mitochondrial DNA variation and evolutionary relationships in the amakihi complex. *Auk* 110: 825–831.
- Tavaré, S., C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* 416:726–729.
- Vawter, A. T., R. Rosenblatt, and G. C. Gorman. 1980. Genetic divergence among fishes of the eastern Pacific and the Caribbean: support for the molecular clock. *Evolution* 234:705–711.
- Wagner, W. L., and V. A. Funk. 1995. Hawaiian biogeography: evolution in a hot spot archipelago. Smithsonian Institution Press, Washington, DC.
- Wakeley, J. 2000. The effects of subdivision on the genetic divergence of populations and species. *Evolution* 4:1092–1101.
- Wares, J. P. 2001. Patterns of speciation inferred from mitochondrial DNA in North American *Chthamalus* (*Cirripedia*: Balanomorpha: Chthamaloidea). *Mol. Phylogenet. Evol.* 18:104–116.
- Wares, J. P., and C. W. Cunningham. 2001. Phylogeographic and historical ecology of the North Atlantic intertidal. *Evolution* 55: 2455–2469.
- Yang, Z. 1997. On the estimation of ancestral population sizes of modern humans. *Genet. Res.* 69:111–116.

Corresponding Editor: S. Elena