

## TESTING FOR GENETIC EVIDENCE OF POPULATION EXPANSION AND CONTRACTION: AN EMPIRICAL ANALYSIS OF MICROSATELLITE DNA VARIATION USING A HIERARCHICAL BAYESIAN MODEL

JAY F. STORZ<sup>1,2</sup> AND MARK A. BEAUMONT<sup>3,4</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Arizona, Biosciences West, Tucson, Arizona 85721

<sup>2</sup>E-mail: storz@email.arizona.edu

<sup>3</sup>School of Animal and Microbial Sciences, University of Reading, Whiteknights, P.O. Box 228, Reading RG6 6AJ, United Kingdom

<sup>4</sup>E-mail: m.a.beaumont@reading.ac.uk

**Abstract.**—The role of past climatic change in shaping the distributions of tropical rain forest vertebrates is central to long-standing hypotheses about the legacy of the Quaternary ice ages. One approach to testing such hypotheses is to use genetic data to infer the demographic history of codistributed species. Population genetic theory that relates the structure of allelic genealogies to historical changes in effective population size can be used to detect a past history of demographic expansion or contraction. The fruit bats *Cynopterus sphinx* and *C. brachyotis* (Chiroptera: Pteropodidae) exhibit markedly different distribution patterns across the Indomalayan region and therefore represent an exemplary species pair to use for such tests. The purpose of this study was to test alternative hypotheses about historical patterns of demographic expansion and contraction in *C. sphinx* and *C. brachyotis* using a coalescent-based analysis of microsatellite variation. Specifically, we used a hierarchical Bayesian model based on Markov chain Monte Carlo simulations to estimate the posterior distribution of genealogical and demographic parameters. The results revealed strong evidence for population contraction in both species. Evidence for a population contraction in *C. brachyotis* was expected on the basis of biogeographic considerations. However, similar evidence for population contraction in *C. sphinx* does not support the hypothesis that this species underwent a pronounced range expansion during the late Quaternary. Genetic evidence for population decline may reflect the consequences of habitat destruction on a more recent time scale.

**Key words.**—Bats, Bayesian inference, coalescent, *Cynopterus*, demographic history, Markov chain Monte Carlo, microsatellite DNA, population bottleneck, zoogeography.

Received March 6, 2001. Accepted September 20, 2001.

The role of Quaternary climatic change in shaping the geographic distributions of tropical species remains a subject of much debate (Endler 1982; Hewitt 2000; Willis and Whittaker 2000). In particular, there remains a great deal of uncertainty about the extent to which contemporary distribution patterns reflect a past history of climatically induced habitat change during the Quaternary ice ages. The Old World fruit bat genus *Cynopterus* (Chiroptera: Pteropodidae) exhibits a diversity of different distribution patterns across the Indomalayan region (Corbet and Hill 1992) and therefore represents an exemplary taxon for testing zoogeographic hypotheses about historical patterns of range expansion and contraction. The two most geographically widespread members of the genus, the short-nosed fruit bat (*Cynopterus sphinx*) and the lesser short-nosed fruit bat (*C. brachyotis*), are characterized by markedly different distribution patterns (Fig. 1). *Cynopterus sphinx* has an extensive continental distribution across tropical South Asia and is peripherally distributed along the western margin of the Malay Archipelago (Storz and Kunz 1999). In contrast, *C. brachyotis* is characterized by a predominantly insular pattern of distribution across the Malay Archipelago and is peripherally distributed along the continental margin of South Asia (Corbet and Hill 1992). The most striking peculiarity of the geographic distribution of *C. brachyotis* is the vast regional disjunction between populations in the Indian subcontinent and the more continuous portion of the species' continental range in Southeast Asia (Corbet and Hill 1992; Bates and Harrison 1997).

*Cynopterus sphinx* is widely distributed across peninsular

India, ranging from semi-arid scrub jungle to secondary forest at the mountain-plains interface. *Cynopterus sphinx* is generally restricted to low elevations and does not live in primary wet-evergreen forest (Bates and Harrison 1997; Storz and Kunz 1999; Storz et al. 2001a). In this same region, available evidence suggests that *C. brachyotis* is almost exclusively restricted to tropical wet-evergreen and semi-evergreen forests of the Western Ghats, a mountain range that parallels the western coast of peninsular India (Bates and Harrison 1997; Pradhan and Kulkarni 1997; Balasingh et al. 1999). Likewise, in Sri Lanka, *C. sphinx* is common in secondary forest and scrub habitat in the arid lowlands, whereas *C. brachyotis* (= *C. sphinx ceylonensis*) is restricted to highland wet-forest habitat (Phillips 1934, 1980; Bates and Harrison 1997). The remainder of the continental distribution of *C. brachyotis* extends at least as far west as southern Myanmar (Burma) and may extend into the tropical wet forests of the eastern Himalayas (Corbet and Hill 1992). A number of vertebrate taxa restricted to wet-forest habitat exhibit this same pattern of regional disjunction between the mountains of southwestern India and Sri Lanka and the more continuous tract of wet forest to the east or northeast of the Gangetic Plain (Blanford 1901; Hora and Jayaram 1949; Ripley 1949; Roonwal and Nath 1949; Jayaram 1974; Kurup 1974). This distributional concordance across taxa suggests that the geographic ranges of extant lineages have undergone a pronounced shift through time in response to a common vicariant event, most likely associated with Quaternary climatic change (Brandon-Jones 1996).

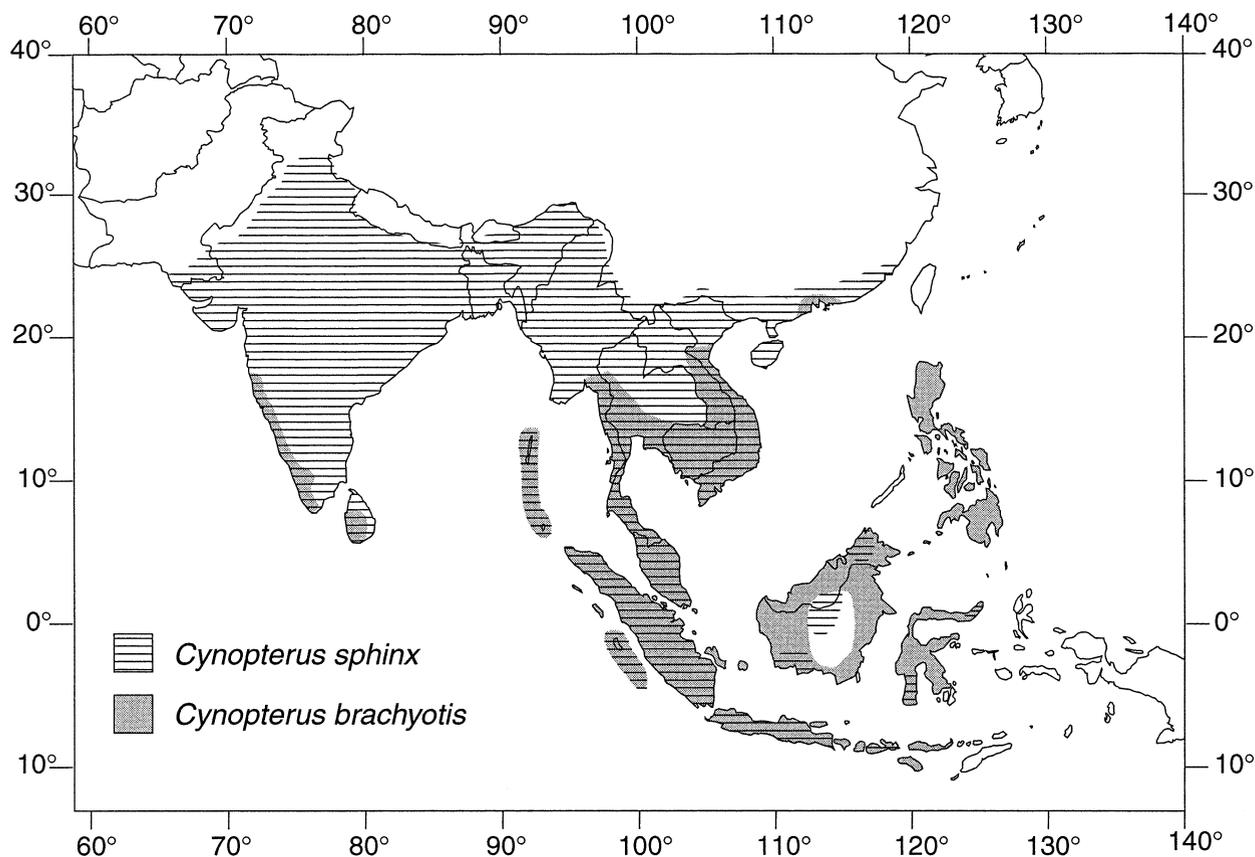


FIG. 1. Geographic distribution of *Cynopterus sphinx* and *C. brachyotis* in the Indomalayan region. The map is based on localities of record listed in Corbet and Hill (1992), Bates and Harrison (1997), Pradhan and Kulkarni (1997), Balasingh et al. (1999), and Storz and Kunz (1999).

The Indian subcontinent is thought to have been continuously forested through most of the late Quaternary (Frenzel et al. 1992). It thus appears likely that the *C. brachyotis* population in the Western Ghats represents an isolated relic of a formerly continuous population whose disappearance from the remainder of the Indian subcontinent is attributable to deforestation associated with the onset of arid climatic conditions in the late Pleistocene. Reconstructions of the vegetational and climatic history of the Indian subcontinent suggest that tropical wet-forests have undergone two major contractions in the late Quaternary: once during the last glacial maximum (20,000–16,000 years ago) and again in the mid-Holocene (5000–2000 years ago; Sukumar et al. 1993, 1995; Rajagopalan et al. 1997, 1999; Ravindranath and Sukumar 1998). Thus, whereas climatically induced changes in forest cover appear to have forced a dramatic contraction of the geographic range of *C. brachyotis*, these same changes also likely facilitated colonization of the Indian subcontinent by *C. sphinx*. This biogeographic scenario makes clear predictions about contrasting patterns of historical demography in these two species. Specifically, it predicts that *C. brachyotis* underwent a dramatic demographic contraction, while *C. sphinx* underwent an equally dramatic demographic expansion across the same region. Alternatively, human-induced environmental changes in peninsular India over the last few

thousand years (Chandran 1997) may have caused population declines in both species.

Fortunately, such hypotheses can now be tested using genetic data. Recent advances in population genetic theory permit inferences about demographic history from patterns of DNA variation in contemporary populations (Hudson 1990; Donnelly and Tavaré 1995; Donnelly 1999). Using data in the form of multilocus microsatellite genotypes, the frequency distribution of allelic length variants can be viewed from a genealogical perspective to make inferences about historical changes in effective population size. The advent of genealogical modeling has greatly facilitated the estimation of demographic and mutational parameters using microsatellite data (Chakraborty and Kimmel 1999; Feldman et al. 1999; King et al. 2000). A typical approach is to infer parameters by comparing statistics calculated from an empirical dataset with a distribution generated by Monte Carlo simulations of the coalescent process. This method-of-moments approach to microsatellite analysis has been used successfully to infer demographic and mutational parameters in populations of humans and other species (Di Rienzo et al. 1994, 1998; Harpending et al. 1998; Kimmel et al. 1998; Reich and Goldstein 1998; Goldstein et al. 1999; Gonser et al. 2000). When historical information about the population of interest is available, an alternative is to use a Bayesian approach to

statistical inference. Historical information can be used to construct a prior distribution of parameter values, which can then be combined with methods of calculating the likelihoods of demographic and mutational parameters from genetic data (Beaumont 1999; Beaumont and Bruford 1999). The purpose of this study was to use a hierarchical Bayesian analysis of microsatellite variation to test alternative hypotheses about historical patterns of demographic expansion and contraction in Indian populations of *C. sphinx* and *C. brachyotis*.

## MATERIALS AND METHODS

### Sampling

Bats were trapped on roosting and foraging grounds as described in Storz et al. (2000a, b, 2001a). For this study, a total of 357 bats (246 *C. sphinx* and 111 *C. brachyotis*) were sampled from five localities along the eastern flanks of the Western Ghats. *Cynopterus sphinx* was sampled from a single locality (Pune, Maharashtra, 18°32'N, 73°51'E). *Cynopterus brachyotis* was sampled from four localities at elevations ranging from 675 to 1050 m, in tropical wet-evergreen and semi-evergreen forest (Hosabale, Karnataka, 14°12'N, 74°41'E; Mudigere, Karnataka, 13°08'N, 75°35'E; Appan-gala, Karnataka, 12°23'N, 75°43'E; and Kalakad Mundanthurai Tiger Reserve, Tamil Nadu, 8°50'N, 77°30'E). Bats could be unambiguously identified to species using diagnostic criteria provided in Bates and Harrison (1997) and Storz and Kunz (1999). Only adults were used in the genetic analysis. Tissue samples for DNA analysis were collected as described in Storz et al. (2001b, c).

### Genotyping

Genetic analysis of *C. sphinx* was based on eight polymorphic microsatellite loci: two di-, three tri-, and three tetranucleotide repeats. Primer sequences and polymerase chain reaction protocols were reported previously (Storz 2000; Storz et al. 2001b). The genetic analysis of *C. brachyotis* was based on a subset of six of these loci, five of which were polymorphic. For both species, sequencing of alleles confirmed that length polymorphism at each locus was attributable to variation in the copy number of a single repeat motif. Genetic subdivision among the four population samples of *C. brachyotis* was tested using an exact *G*-test of genotypic frequencies (Goudet et al. 1996). Using the program FSTAT (ver. 2.9.1; updated from Goudet 1995), 10,000 randomizations (permuting genotypes across populations) were used to generate a probability distribution under the null hypothesis of no differentiation. Available genetic marker data provide no evidence for introgressive hybridization between *C. sphinx* and *C. brachyotis* in regions of range overlap in peninsular India (J. F. Storz, unpubl. data). Thus, microsatellite datasets for each of the two species are independent.

### The Model

We consider the demographic history of a closed population that increases or decreases exponentially from an initial size  $N_1$  to the current size  $N_0$ , over a time interval  $x_a$ . The population size at any time  $x$  is given by

$$N_x = N_0 \left( \frac{N_1}{N_0} \right)^{x/x_a}, \quad (1)$$

where time is increasing into the past. The model makes use of multilocus microsatellite genotypes representing a population sample of chromosomes. The loci are assumed to be evolving according to a strict single-step mutation model, with mutation rate  $\mu$ . Allelic frequencies are used to infer the model parameters  $\Phi = \{N_0, N_1, x_a, \mu\}$  using a Bayesian approach: the posterior probability density of the parameters is estimated given a prior density and the data. This can be accomplished several different ways (reviewed in Stephens and Donnelly 2000). Beaumont (1999) introduced a method based on microsatellite data for detecting changes in population size using Markov chain Monte Carlo (MCMC). Here we analyze the microsatellite data using two different models. The first method (the basic model) is very similar to that described in Beaumont (1999), but the implementation is slightly modified (point 1 below). The second method (the hierarchical model) differs from the original model (Beaumont 1999) in three respects: (1) In the original model, multiple loci were accommodated by estimating posterior densities of the parameters for each locus separately and then taking the product of the independent densities. In the present study, posterior densities for both models are estimated using all loci in the same MCMC simulation. This should not lead to any difference in the expected results between the original model and the basic model, but it reduces error associated with density estimation. (2) In the original model, only three identifiable parameters were inferred: mutation rate,  $\mu$ , scaled by current population size,  $N_0$  ( $\theta = 2N_0\mu$ ); the ratio of current to ancestral population size ( $r = N_0/N_1$ ); and number of generations,  $t_a$ , over which the population has been changing in size scaled by current population size ( $t_f = t_a/N_0$ ). Here we consider four natural parameters  $N_0, N_1, x_a$  ( $= g \times t_a$ , where  $g$  is the generation length), and  $\mu$ . These parameters are inferred separately using priors, following the approach of Tavaré et al. (1997) and Wilson and Balding (1998). (3) The original model was based on the assumption that all parameters other than mutation rate were identical across loci. In the hierarchical model, however, parameters are free to vary from one locus to the next. The extent of interlocus variation is set by priors, as described below. Theoretically, demographic parameters should be uniform across loci. However, in reality, a past history of selection at linked sites or variation in mutational dynamics may produce discrepancies among loci that adversely affect the fit of the model.

Both models are available for downloading via <http://www.rubic.rdg.ac.uk/~mab/software.html>

### Rationale for the model

In the Bayesian approach to statistics, the data and parameters are treated equally as random variables that have some joint distribution, which can be explicitly modeled (i.e., we can calculate the relative frequency of occurrence of any given dataset and parameter values). Two concepts are useful here: marginal and conditional distributions. A conditional distribution is the probability distribution of one set of variables given that the others have specific values. For example,

the likelihood is the conditional distribution of the data for a given set of parameter values. A marginal distribution is the sum (integral) of the joint distribution over particular variables. It represents a best guess of the probability distribution of some variables if we have no information on the values of others. For example, the Bayesian prior is the marginal distribution of the parameters over the data, that is, our best guess for the parameter values in the absence of data. In fact, the joint distribution discussed here is simply the product of the prior and the likelihood. The goal of a Bayesian analysis is to estimate the conditional distribution of the parameter values given the data, which is called the posterior distribution. The increasing popularity of Bayesian analysis is largely attributable to the fact that the integration necessary to calculate marginal and conditional distributions has only recently become computationally feasible through the use of methods such as MCMC.

In our model, the random variables whose joint distribution we can specify consist of the genetic data, the genealogical history that gave rise to the data, and the mutational and demographic parameters described earlier. We condition on the genetic data that we have obtained and use MCMC to simulate random draws of the parameters from the posterior distribution. In general, because we know nothing of the genealogical history, we report the posterior distributions of the demographic parameters and mutation rate marginal to the genealogical history.

The hierarchical modeling approach described here allows the posterior distribution of the mutation rate and the demographic parameters to vary among loci. Allowing inter-locus variation in demographic parameters makes our analysis more robust, and it enables us to detect aberrant loci that could potentially bias the results. In a model where the parameters are not allowed to vary, aberrant loci will have a strong effect because the likelihoods for individual loci are multiplied together. In contrast, in a hierarchical model the likelihoods are combined in a more additive fashion. In many ways our analysis is analogous to the Lewontin-Krakauer test (Lewontin and Krakauer 1973; Beaumont and Nichols 1996), where the direct or indirect effects of selection at a particular locus can be detected against a baseline level of neutral variation at unlinked loci.

#### Specification of the priors

In the basic model, rectangular priors were assumed for the parameters, with bounds of  $(-5, 5)$  for  $\log_{10}(\theta)$ ,  $\log_{10}(r)$ , and  $\log_{10}(t_f)$ . These limits were chosen to be sufficiently broad that the high density region of the posterior distribution would be relatively unaffected by the prior. In the hierarchical model, the parameters  $\Phi$  for each locus were assumed to be drawn from log-normal distributions, each with means (on a log<sub>10</sub> scale) of  $M = \{M_{N0}, M_{N1}, M_{\mu}, M_{xa}\}$  and standard deviations (SDs)  $V = \{V_{N0}, V_{N1}, V_{\mu}, V_{xa}\}$ . These distributions were then used to calculate  $p(\Phi | V, M)$ .

Prior distributions for the means were assumed to be specified by normal distributions with means  $\alpha_{N0}, \alpha_{N1}, \alpha_{\mu},$  and  $\alpha_{xa}$ , and SDs  $\sigma_{N0}, \sigma_{N1}, \sigma_{\mu},$  and  $\sigma_{xa}$ . Priors for the SDs were taken to be normal distributions truncated at zero with means  $\beta_{N0}, \beta_{N1}, \beta_{\mu},$  and  $\beta_{xa}$ , and SDs  $\tau_{N0}, \tau_{N1}, \tau_{\mu},$  and  $\tau_{xa}$ . It was

not necessary to calculate a normalizing constant for these truncated distributions because the integration constant cancels out in equation (2) below. This set of parameters is designated  $H$ . These distributions were then used to calculate  $p(V, M | H)$ .

Simulations described here were based on the following assumptions for both bat species:  $\sigma_{N0} = \sigma_{N1} = 3$ ,  $\beta_{N0} = \beta_{N1} = \beta_{\mu} = \beta_{xa} = 0$  (i.e., density is monotonically declining from  $x = 0$ ),  $\tau_{N0} = \tau_{N1} = \tau_{xa} = 0.5$ ,  $\alpha_{\mu} = -3.3$ ,  $\sigma_{\mu} = 0.5$ ,  $\tau_{\mu} = 2$ ,  $\alpha_{xa} = 3.7$ , and  $\sigma_{xa} = 1$ . For *C. sphinx* we assumed  $\alpha_{N0} = 6.8$  and  $\alpha_{N1} = 6.1$ , and for *C. brachyotis* we assumed  $\alpha_{N0} = 6.1$  and  $\alpha_{N1} = 6.8$ . The prior means for the current (effective) population sizes were chosen by estimating the geographical area of the Indian subcontinent ( $3.29 \times 10^6$  km<sup>2</sup>), the fraction of the subcontinent currently occupied by each species (0.8 and 0.17 for *C. sphinx* and *C. brachyotis*, respectively; Corbet and Hill 1992; Storz and Kunz 1999), population density (5 adults/km<sup>2</sup>; Storz et al. 2000a, b; J. F. Storz, unpubl. data), and the ratio of effective population size to adult census number ( $\sim 0.5$ ; Storz et al. 2001c). The values of  $\sigma_{N0}$  and  $\sigma_{N1}$  were chosen to allow for a large degree of uncertainty in these estimates. Because geographic range shifts may also cause changes in population density and spatial connectivity, changes in range size may not necessarily entail proportional changes in effective population size (Whitlock and Barton 1997; Wakeley 1999). However, it seems reasonable to assume that range expansions or contractions on a continental scale would generally affect population numbers in the same direction. The prior mean time since the population started changing in size was taken to be the start of the mid-Holocene wet-forest contraction ( $\approx 5000$  years ago). The SD around this value was chosen such that recent changes and earlier periods of deforestation (i.e., 20,000–16,000 years ago) also had support under the prior. We also considered a model with  $\alpha_{xa} = 1.7$  and  $\sigma_{xa} = 2$ , to investigate the effect of the prior on the posterior distribution of  $x_a$ . The former is referred to as the ‘‘longer-time’’ model and the latter is called the ‘‘shorter-time’’ model. We assumed a point prior of  $g = 8.34$  years per generation (average of male and female generation times; Storz et al. 2001c). The mean mutation rate among loci was assumed to be  $5 \times 10^{-4}$ , a value widely assumed in demographic models (e.g., Goldstein et al. 1995) but a little lower than pedigree-based estimates for autosomal microsatellites in humans (Brinkmann et al. 1998; Ellegren 2000b). The prior SD of the mean mutation rate was chosen so that values in the region  $5 \times 10^{-3}$  to  $5 \times 10^{-5}$  would have reasonable support. The value chosen for  $\tau_{\mu}$  means that  $V_{\mu} \approx 2$  would be frequent under the prior so that the expected ratio of mutation rates for pairs of loci would be about 700-fold. The values chosen for  $\tau_{N0}, \tau_{N1},$  and  $\tau_{xa}$  imply that the expected ratio of the corresponding demographic parameters for pairs of loci would be about 5-fold under the prior. In each case, expected ratios that were much smaller or larger than these values also had support under the prior. Thus, the model allows for a wide degree of variation among loci. Overall, the main strategy was to have vague priors on the demographic parameters and differences among loci, but to have more informative priors on the mutation rate.

### Calculation of Metropolis-Hastings acceptance probability

In the Metropolis-Hastings algorithm, parameters are updated by choosing trial values and deciding whether they should be accepted. The simulated sequence of parameter values tends to the posterior distribution of the parameters as the sequence length becomes large. For the hierarchical model, the acceptance probability was calculated as

$$P = \frac{p(V', M' | H')}{p(V, M | H)} \times \frac{p(\Phi' | V', M')}{p(\Phi | V, M)} \times \frac{p(G' | \Phi')}{p(G | \Phi)} \times \frac{p(V, M, \Phi, G | V', M', \Phi', G')}{p(V', M', \Phi', G' | V, M, \Phi, G)}, \quad (2)$$

where the prime denotes a trial updated value. The distributions in the first two terms are given above. The probability of a genealogical history given the data,  $p(G | \Phi)$ , is given in Beaumont (1999). The last term in the expression is the Hastings correction, which ensures that there is no bias in the updating procedure toward any particular region of parameter space. The Hastings corrections for updates in  $G$  are given in Beaumont (1999); corrections for the other parameters are given below. For  $P \geq 1$ , the trial value was accepted, otherwise it was accepted with probability  $P$ . With the updating protocol given below, only a small subset of equation (2) was evaluated at any one time. For example,  $p(G | \Phi)$  was only evaluated for the full sequence of events in the genealogy when the  $T_i$  or components of  $\Phi$  were changed, as discussed in Beaumont (1999). For the basic model, the acceptance probability was similar to equation (2), but without the terms in  $H$ , and the priors were simple rectangular distributions as discussed above.

### Updating the parameters

For each locus, the parameters that need to be updated are the sequence of events in the genealogical history, the times of the events,  $T_i$ , and the components of  $\Phi$ :  $N_0$ ,  $N_1$ , and  $\mu$ . We followed the random-sweep updating protocol (Brooks 1998). The general scheme will be outlined first, followed by the details. The components of  $\Phi$  were parameterized on the  $\log_{10}$  scale. The parameters were updated in a random order with (conditional) probabilities given in brackets below. This scheme was developed by trial and error to obtain good rates of convergence.

#### Basic model:

- (0.99) Update sequence of events.
- (0.01) Update some of  $T_i$ ,  $\log_{10}(\theta)$ ,  $\log_{10}(r)$ ,  $\log_{10}(t_f)$ .
  - (0.05)  $T_i$  only.
  - (0.95) Some of  $T_i$ ,  $\log_{10}(\theta)$ ,  $\log_{10}(r)$ ,  $\log_{10}(t_f)$ .
    - (0.05)  $\log_{10}(\theta)$ ,  $\log_{10}(r)$ ; (0.05)  $\log_{10}(\theta)$ ,  $\log_{10}(r)$ ,  $T_i$ .
    - (0.2)  $\log_{10}(\theta)$ ,  $\log_{10}(t_f)$ ; 0.2  $\log_{10}(\theta)$ ,  $\log_{10}(t_f)$ ,  $T_i$ .
    - (0.0125)  $\log_{10}(r)$ ; (0.0125)  $\log_{10}(r)$ ,  $T_i$ .
    - (0.2)  $\log_{10}(\theta)$ ,  $\log_{10}(r)$ ,  $\log_{10}(t_f)$ ; 0.2  $\log_{10}(\theta)$ ,  $\log_{10}(r)$ ,  $\log_{10}(t_f)$ ,  $T_i$ .
    - (0.025)  $\log_{10}(\theta)$ ,  $T_i$ .
    - (0.025)  $\log_{10}(t_f)$ ; (0.025)  $\log_{10}(t_f)$ ,  $T_i$ .

The sequence of events was updated as specified in Beaumont (1999). The times were updated by simulating from the

coalescent, and the log-transformed values of demographic and mutational parameters were updated by adding normal random variables with a mean of zero and a SD of 0.43. As described in Beaumont (1999), when the demographic and mutational parameters were updated jointly in the scheme above, the same random deviate was used for each updated parameter, but the sign was reversed for  $\log_{10}(t_f)$ . Additionally, when  $\log_{10}(t_f)$  was updated alone (or in conjunction with  $T_i$ ), a normal random variable with a SD of 2.2 was used. This was done to ensure better mixing of the simulated Markov chain. For these updates, the Hastings correction = 1.0. If parameter values were generated that were outside the rectangular priors, the prior probability was zero.

#### Hierarchical model:

- (0.5) Update sequence of events.
- (0.02) Update some of  $T_i$ ,  $\log_{10}(N_0)$ ,  $\log_{10}(N_1)$ ,  $\log_{10}(\mu)$ ,  $\log_{10}(x_a)$ .
- (0.5) Update parameters for all loci and then update prior means in  $M$  corresponding to the parameters that are changed.
- (0.5)  $T_i$  only.
- (0.5) Some of  $T_i$ ,  $\log_{10}(N_0)$ ,  $\log_{10}(N_1)$ ,  $\log_{10}(\mu)$ ,  $\log_{10}(x_a)$ .
  - (0.05)  $\log_{10}(N_1)$ ,  $\log_{10}(\mu)$ ; (0.05)  $\log_{10}(N_1)$ ,  $\log_{10}(\mu)$ ,  $T_i$ .
  - (0.05)  $\log_{10}(\mu)$ ,  $\log_{10}(x_a)$ ; (0.05)  $\log_{10}(\mu)$ ,  $\log_{10}(x_a)$ ,  $T_i$ .
  - (0.4)  $\log_{10}(N_0)$ ,  $\log_{10}(N_1)$ ,  $\log_{10}(\mu)$ ,  $\log_{10}(x_a)$ .
  - (0.05)  $\log_{10}(N_1)$ ;  $\log_{10}(N_1)$ ,  $T_i$ .
  - (0.05)  $\log_{10}(N_0)$ ;  $\log_{10}(N_0)$ ,  $T_i$ .
  - (0.1)  $\log_{10}(\mu)$ ,  $T_i$ .
  - (0.05)  $\log_{10}(x_a)$ ; 0.05  $\log_{10}(x_a)$ ,  $T_i$ .
- (0.5) Update one locus (as above, but with one locus only)
- (0.05) Update one of  $V_{N0}$ ,  $V_{N1}$ ,  $V_\mu$ ,  $V_{x_a}$ .

The parameters were updated in the same way as for the basic model described above. The demographic and mutational parameters were updated by adding a normal random variable with mean zero and SD that varied among different types of update as follows. When updating  $\log_{10}(x_a)$  alone (or in conjunction with  $T_i$ ) across all loci, a SD of 2.2 was used. Otherwise it was 0.43 for any set of parameters across all loci. When updating a single locus, a SD of 0.22 was used for all parameters. When updating the demographic and mutational parameters jointly, the same random variable was used but with differing sign for  $\mu$  (this leads to equivalent joint updates as for  $\theta$ ,  $r$ , and  $t_f$  above). When all loci were updated, the means of the priors of the altered parameters were updated by adding the same random variable used to update the parameters. Thus, when all loci were updated, the prior probability of the parameters remained unchanged (although that of the priors—the hyperprior probability—did change). The SDs of the priors were updated by drawing a log-normal random variable with median centered on the current SD and a SD (on a  $\log_{10}$  scale) of 0.5. In this case, the Hastings correction was not equal to 1.0.

TABLE 1. Summary statistics for microsatellite markers used in the genetic analysis of *Cynopterus sphinx* and *C. brachyotis* sampled from peninsular India.  $N$ , number of bats genotyped per locus;  $N_A$ , number of alleles per locus;  $H_O$ , observed heterozygosity; and  $H_E$ , expected heterozygosity.

Locus	<i>Cynopterus sphinx</i>					<i>Cynopterus brachyotis</i>				
	Allele size range	$N$	$N_A$	$H_O$	$H_E$	Allele size range	$N$	$N_A$	$H_O$	$H_E$
Locus 3	174–194	246	10	0.70	0.76	170–204	104	16	0.85	0.86
CSP-1	191–218	246	9	0.71	0.70	174–227	111	12	0.61	0.69
CSP-2	113–131	246	6	0.77	0.74	101	20	1	0	0
CSP-3	95–107	246	5	0.40	0.42	—	—	—	—	—
CSP-5	110–170	246	12	0.72	0.73	110–166	111	11	0.29	0.34
CSP-7	231–265	246	16	0.81	0.82	229–263	111	17	0.72	0.86
CSP-8	150–202	242	14	0.70	0.74	—	—	—	—	—
CSP-9	278–298	246	5	0.49	0.49	270–282	111	4	0.45	0.49

### Implementation

To achieve reliable convergence in a reasonable amount of time, a subset of 100 chromosomes from each dataset was taken. The chain was run for  $2 \times 10^9$  steps, recording parameter values every  $10^5$  steps to give 20,000 draws from the posterior distribution. Five independent chains were run for each analysis presented in the Results. The output was checked for convergence using the Gelman-Rubin statistic calculated in CODA (Best et al. 1995), as implemented in R (Ihaka and Gentleman 1996; available via <http://www.r-project.org/>). We examined plots of the value of the Gelman-Rubin statistic against iteration number to check that it was reliably converging toward its final value as recommended by Brooks and Gelman (1998). In addition, posterior densities from individual runs were examined to check for overall consistency in shape. The last half of each run was then combined to produce an overall set of 50,000 points. Each run took about 70 h on a 700-Mhz Pentium (a small Beowulf cluster of machines was used to run chains in parallel). Density estimation was carried out using the program Locfit (Loader 1996), implemented in R. This was used to estimate modes and highest probability density (HPD) limits. The HPD limits specify points of equal probability density enclosing a region (or possibly disjoint regions in multimodal data) with probability equal to some specified value. The strength of evidence of population decline versus population growth was assessed using Bayes factors as described in Beaumont (1999). The Bayes factor is a ratio in which the numerator is the posterior probability of one model (e.g., population contraction) divided by its prior probability and the denominator is the posterior probability of another model (e.g., population expansion) divided by its prior probability. The Bayes factor expresses the relative likelihood of the two different models. Bayes factors greater than  $\exp(2)$  are generally considered significant. The ratio of posterior probabilities can be estimated from the simulated chain by counting the proportion of iterations in which the population has expanded and then dividing this by the proportion of iterations in which it has contracted. Although extreme ratios vary among replicate chains due to sampling effects, ratios of 10 or less are generally similar among replicate chains for reasonable values of the Gelman-Rubin statistic. Datasets from each of the two species were analyzed with the basic model and the hierarchical model. Additionally, the hierarchical

model was run with a different prior on the time over which the population changed in size.

## RESULTS

### Summary Statistics

For the total sample of *C. sphinx* ( $n = 246$  bats), mean number of alleles per locus was 9.6 (range = 5–16) and mean observed heterozygosity ( $H_O$ ) was 0.66 (range = 0.40–0.81; Table 1). Although preliminary screening of 20 individuals indicated that CSP-2 was monomorphic in *C. brachyotis*, the remaining five loci segregated an average of 11.3 alleles (range = 4–17) and mean  $H_O$  was 0.60 (range = 0.45–0.72;  $n = 111$  bats). The exact  $G$ -test revealed no evidence of heterogeneity in genotypic frequencies among the four population samples of *C. brachyotis* ( $P > 0.05$  for each locus). Allelic frequencies for both species are given in Table 2.

### Variation in Parameters among Loci

The degree of consistency among loci was quantified by the posterior distributions of  $V = \{V_{N0}, V_{N1}, V_{\mu}, V_{xa}\}$ . The simulated posterior distributions of these parameters converged rapidly with both the *C. sphinx* and *C. brachyotis* data, in both the longer-time and shorter-time models. Estimates of the Gelman-Rubin statistic and their 97.5% quantiles were  $\leq 1.02$  for each parameter in all simulations. Results from both models were very similar, and those from the longer-time model are described here.

The microsatellite data for *C. sphinx* revealed no significant discrepancies among loci. The 90% quantiles of the posterior distributions did not exceed those of the priors (Table 3). Modes of the posterior distributions for  $V_{N0}$ ,  $V_{N1}$ , and  $V_{xa}$  tended to be at or close to zero, and joint posterior distributions of different pairs of parameters in  $V$  showed no strong patterns of correlation (data not shown). The mode for  $V_{\mu}$  was approximately 0.4, but the posterior density estimated at zero remained  $>0.75$  of the modal value. For parameters considered singly or in pairwise combination, a significant degree of interlocus variation was indicated if a modal value of zero fell outside the estimated 90% HPD limits. Applying this criterion to the *C. sphinx* data, we were unable to reject the null hypothesis of zero variance among loci. However, the posterior distributions were consistent with a substantial degree of variation, which may be manifest in all parameters

TABLE 2. Microsatellite allelic frequencies for *Cynopterus sphinx* and *C. brachyotis*. For each locus,  $n$  denotes number of chromosomes (*C. sphinx*/*C. brachyotis*). *Cynopterus brachyotis* was not genotyped at CSP-3 or CSP-8.

Locus	Allele	Frequency		
		<i>C. sphinx</i>	<i>C. brachyotis</i>	
Locus 3 ( $n = 492/208$ )	170	—	0.0048	
	172	—	0.0192	
	174	0.0061	0.0913	
	176	—	0.3077	
	178	0.0020	0.0673	
	180	0.0041	0.0144	
	182	0.0427	0.0577	
	184	0.2012	0.0577	
	186	0.2114	0.1106	
	188	0.3882	0.0769	
	190	0.0732	0.0721	
	192	0.0203	0.0337	
	194	0.0508	0.0481	
	196	—	0.0240	
	198	—	0.0096	
	204	—	0.0048	
	CSP-1 ( $n = 492/222$ )	176	—	0.0045
		179	—	0.0090
182		—	0.2432	
185		—	0.0180	
191		0.0407	—	
194		0.0264	—	
197		0.0244	—	
200		0.1382	0.0045	
203		0.1423	0.4910	
206		0.0325	0.0586	
209		0.4980	0.0045	
212		0.0041	—	
218		0.0935	0.0045	
221		—	0.0045	
224		—	0.0631	
227		—	0.0946	
CPS-2 ( $n = 492/40$ )		101	—	1.0000
		113	0.0020	—
	119	0.2459	—	
	122	0.1118	—	
	125	0.3598	—	
	128	0.2317	—	
	131	0.0488	—	
	CSP-3 ( $n = 492/0$ )	95	0.7358	—
		98	0.0610	—
		101	0.0020	—
104		0.1992	—	
107		0.0020	—	
CPS-5 ( $n = 492/222$ )	110	0.0020	0.0090	
	114	—	0.0045	
	118	—	0.0450	
	122	—	0.8108	
	126	—	0.0631	
	130	0.0122	0.0045	
	134	0.0020	—	
	138	0.3841	—	
	142	0.0020	—	
	146	0.0020	0.0180	
	150	0.0041	0.0225	
	154	0.2642	0.0090	
	158	0.2195	0.0090	
	162	0.0772	—	
166	0.0244	0.0045		
170	0.0061	—		
CPS-7 ( $n = 492/222$ )	229	—	0.0541	
	231	0.0102	0.0090	
	233	0.0325	0.0901	
	235	0.0081	0.1937	
	237	0.2744	0.2658	

TABLE 2. Continued.

Locus	Allele	Frequency	
		<i>C. sphinx</i>	<i>C. brachyotis</i>
CPS-8 ( $n = 484/0$ )	239	0.1057	0.0631
	241	0.0020	0.1126
	243	0.1626	0.0450
	245	0.0020	0.0721
	247	0.0041	0.0360
	249	0.2480	0.0180
	251	0.0020	0.0090
	253	0.0081	0.0045
	255	—	0.0090
	257	0.0244	0.0090
	261	0.0996	0.0045
	263	0.0102	0.0045
	265	0.0061	—
	150	0.0207	—
	154	0.0165	—
	158	0.0558	—
	162	0.0372	—
	166	0.0103	—
170	0.0021	—	
174	0.0062	—	
178	0.0103	—	
182	0.0702	—	
186	0.1136	—	
190	0.4711	—	
194	0.0930	—	
198	0.0847	—	
202	0.0083	—	
CSP-9 ( $n = 492/222$ )	270	—	0.0045
	274	—	0.6171
	278	0.0020	0.3649
	282	—	0.0135
	286	0.1931	—
	290	0.6809	—
	294	0.1138	—
	298	0.0102	—

or may appear to be manifest in all parameters through weak correlations with one variable parameter such as mutation rate (see Discussion).

For the *C. brachyotis* dataset, the marginal posterior distributions for  $V$  were generally wider than those for *C. sphinx* (Table 3). In the case of  $V_{N0}$  and  $V_{xa}$ , posterior distributions were slightly wider than the priors. In particular, the joint posterior distribution of  $V_{N0}$  and  $V_{\mu}$  showed a negative correlation, and density of the region where  $V_{N0} + V_{\mu}$  was less than about 0.5 fell below the critical density of the 90% HPD limit. Thus, the point where  $V_{N0}$  and  $V_{\mu}$  were both zero was outside the 90% HPD limit. The 90%, 50%, and 10% HPD regions of the joint distribution of  $V_{N1}$  and  $V_{xa}$  were narrower than those of the prior, with no evidence of correlation, and the mode was at (0, 0). There was no evidence against the

TABLE 3. Upper 90% quantiles for the priors and upper 90% quantiles of the posterior distribution of  $V = \{V_{N0}, V_{N1}, V_{\mu}, V_{xa}\}$  for *Cynopterus sphinx* and *C. brachyotis*.

	Prior	<i>C. sphinx</i>	<i>C. brachyotis</i>
$V_{N0}$	0.82	0.62	1.01
$V_{N1}$	0.82	0.77	0.75
$V_{\mu}$	3.29	0.85	2.09
$V_{xa}$	0.82	0.74	0.86

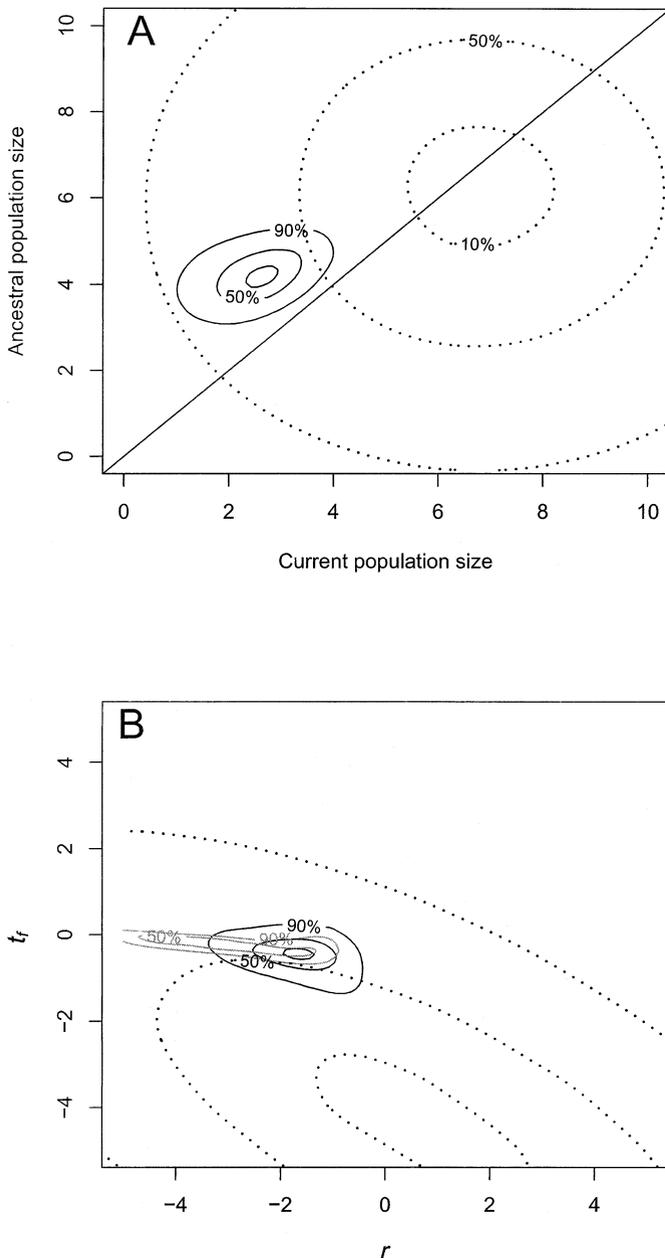


FIG. 2. Joint posterior distribution of  $M_{N0}$  (contemporary population size) and  $M_{N1}$  (ancestral population size) based on data from *Cynopterus sphinx*. (A) The 90%, 50%, and 10% highest probability density (HPD) limits are plotted for the joint distribution of  $M_{N0}$  and  $M_{N1}$  on a  $\log_{10}$  scale. The diagonal line corresponds to  $M_{N0} = M_{N1}$ . (B) HPD limits for the joint distribution of  $r$  and  $t_f$  in the basic and hierarchical models are plotted together. HPD limits for the prior in the hierarchical model are shown as dotted lines.

null hypothesis of zero variance among loci when  $N_1$  and  $x_d$  were considered together, but there was evidence against the hypothesis of zero variance in  $N_0$ ,  $\mu$ , or both.

#### Evidence for Historical Changes in Population Size

The data indicate that *C. sphinx* has undergone a pronounced population contraction (Fig. 2A). In the case of the

longer-time model, the Gelman-Rubin convergence statistics for the four demographic and mutational parameters were (upper 97.5% quantile is given in parentheses here and in the following)  $M_{N0}$ : 1.02 (1.04);  $M_{N1}$ : 1.00 (1.00);  $M_\mu$ : 1.00 (1.01); and  $M_{x_d}$ : 1.01 (1.03). The posterior distribution of the mean mutation rate  $M_\mu$  had a mean of  $-3.58$  (SD = 0.46), and was therefore very similar to the prior. This was not surprising given the broad priors on  $M_{N0}$  and  $M_{N1}$ , and it should be emphasized that the results presented here on population sizes (but not their ratios) and times of events are highly dependent on the priors set for  $M_\mu$ . The 90%, 50%, and 10% HPD regions are plotted for the joint posterior distribution of  $M_{N0}$  and  $M_{N1}$  in Figure 2A. It can be seen that the estimates of population size were generally much less than expected from the priors, with a modal estimate of about  $10^{2.5}$  for the current size and about  $10^4$  for the ancestral size. The Bayes factor in favor of decline versus growth was  $365/0.77 = 474$ .

The parameters were transformed and the posterior distribution of  $r$  and  $t_f$  (on a  $\log_{10}$  scale) was compared with that from the basic model in Figure 2B (for comparison with simulation results, see fig. 5D in Beaumont 1999). The Gelman-Rubin statistics for the basic model were  $r$ : 1.14 (1.34);  $t_f$ : 1.03 (1.08); (the single-locus statistics for  $\theta$  were similar to those for  $r$ ). The relatively high value of the Gelman-Rubin statistic for  $r$  in the basic model was due to the skew in its posterior distribution (Fig. 2B) toward low values of  $r$ . Examination of posterior densities from individual runs showed that the width of the tail (i.e., the amount of posterior probability within the tail) varied among simulations, which also caused the mean to vary. The distribution of  $t_f$  was similar in both models, although it was broader in the hierarchical model. The distribution of  $r$  did not extend into low values and it had a mode at about  $10^{-1.5}$ . The reason for this is that low values of  $r$  were observed only at low values of  $\theta$ , but the priors used on  $\mu$  and  $N_0$  in the hierarchical model made such values improbable. The Bayes factor in favor of decline versus growth was approximately 5000 (only one observation out of 50,000 had  $\log_{10}[r] > 0$ ). The general broadness of the posterior distribution in the hierarchical model, both to higher values of  $r$  and in the width of the distribution of  $t_f$ , reflects the increased uncertainty that results when parameters are allowed to vary among loci.

The *C. brachyotis* data also showed evidence of a population contraction (Fig. 3A), but the apparent decline was not as pronounced as in *C. sphinx*. The Gelman-Rubin statistics were:  $M_{N0}$ : 1.17 (1.38);  $M_{N1}$ : 1.00 (1.01);  $M_\mu$ : 1.00 (1.00); and  $M_{x_d}$ : 1.00 (1.01). Posterior densities of  $M_{N0}$  from individual runs were characterized by a long tail toward high values. The width of the tail (which generally had low posterior probability) varied among simulations, which resulted in the relatively large value of the Gelman-Rubin statistic for  $M_{N0}$ . The mean value of  $M_\mu$  was estimated at  $-3.27$  (SD = 0.47). As with *C. sphinx*, estimated population size was substantially lower than the prior (Fig. 3A). The modal value of  $M_{N0}$  was about 103.5 and that of  $M_{N1}$  was about  $10^4$ . The Bayes factor in favor of decline versus growth was  $18.6/1.34 = 13.9$ .

Results of the basic and hierarchical models are compared in Figure 3B. Gelman-Rubin statistics were  $r$ : 1.01 (1.02)

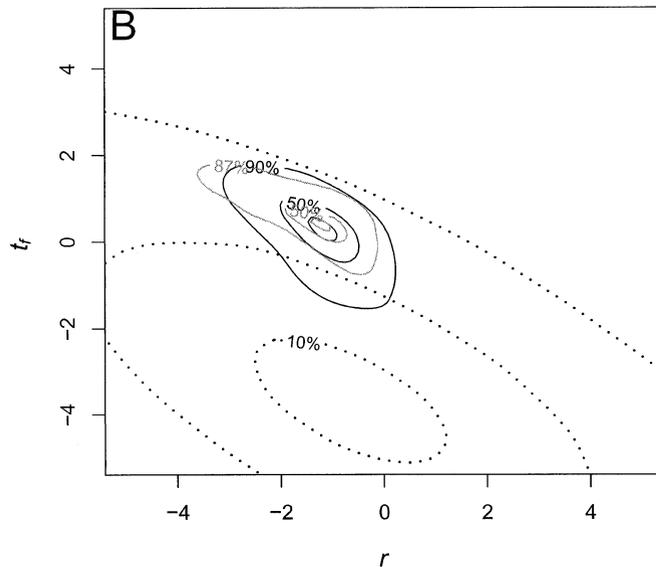
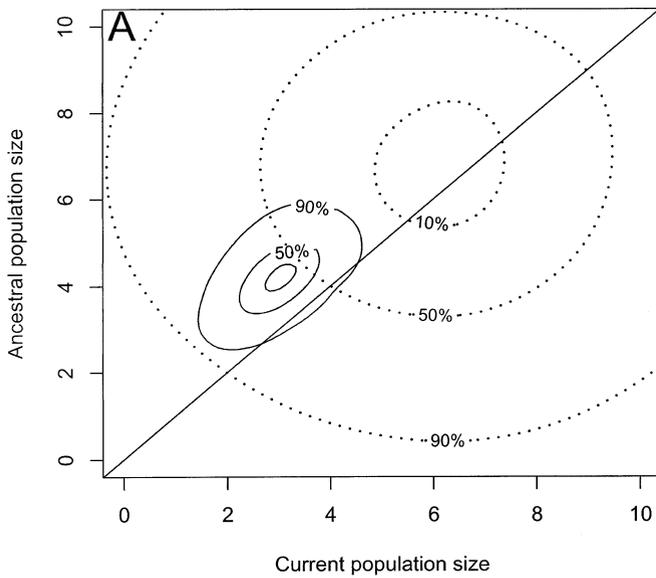


FIG. 3. Joint posterior distribution of  $M_{N0}$  (contemporary population size) and  $M_{N1}$  (ancestral population size) based on data from *Cynopterus brachyotis*. Details are the same as in Figure 2. In (B), the 87% highest probability density limit is shown for the basic model for reasons discussed in the text.

and  $t_f$ : 1.00 (1.01), with values of  $\theta$  at each locus generally similar to those of  $r$ . The Bayes factor in favor of decline was 19.4 for the nonhierarchical model. Outside the peak of high density, the joint distribution of  $r$  and  $t_f$  (on a  $\log_{10}$  scale) was very flat (for explanation, see Beaumont 1999). This makes the 90% HPD region difficult to estimate, so the 87% region was estimated instead, reflecting the edge of the high-density region. It can be seen that the central peak in density is broader in the hierarchical model, due to the fact that the demographic parameters were allowed to vary among

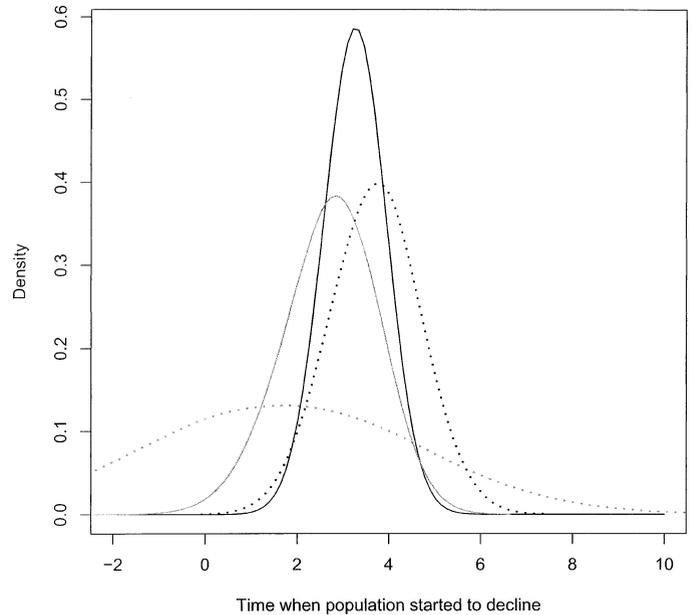


FIG. 4. The posterior distribution of the time when the *Cynopterus sphinx* population started to decline,  $M_{xd}$ , is plotted for the longer-time (darker lines) and shorter-time (lighter lines) models. Prior distributions are shown as dotted lines.

loci. However, tails of the distribution generated by the hierarchical model were narrower because more informative priors were used. The modal value of  $r$  was approximately 0.1 in both models.

#### Time of Change in Population Size

There is little information on the time scale of population decline in either species. The posterior distributions of the time since the populations started to decline (conditional on having declined rather than expanded) are plotted in Figures 4 and 5. The posterior distributions of times from the shorter-time model were also plotted for comparison. Simulated distributions for this model converged well with both datasets. In both cases, the highest 97.5% quantile observed in the four demographic and mutation parameters was 1.06.

In the case of *C. sphinx*, it can be seen that in the longer-time model, the posterior distribution was very similar to that of the prior, with a mode (on a  $\log_{10}$  scale) of 3.4. Overall, the posterior distribution favored a slightly shorter time scale of population decline relative to the prior. In the shorter-time model, the posterior distribution was only slightly shifted in the direction of a shorter time scale, with a mode of 2.9. The model with a prior for shorter times gave stronger support for population decline (Bayes factor = 1509).

Results for *C. brachyotis* tended to support a longer time scale of population decline (Fig. 5, Table 4). The mode of the posterior distribution in the longer-time model was 4.5. The shorter-time model gave slightly less support for population decline (Bayes factor = 8.35). In this model, the modal time on a  $\log_{10}$  scale was 4.9. Thus, the effect of the model with a broader prior supporting shorter times actually increased the time over which the population is inferred to have declined. This is explained by the fact that if a longer

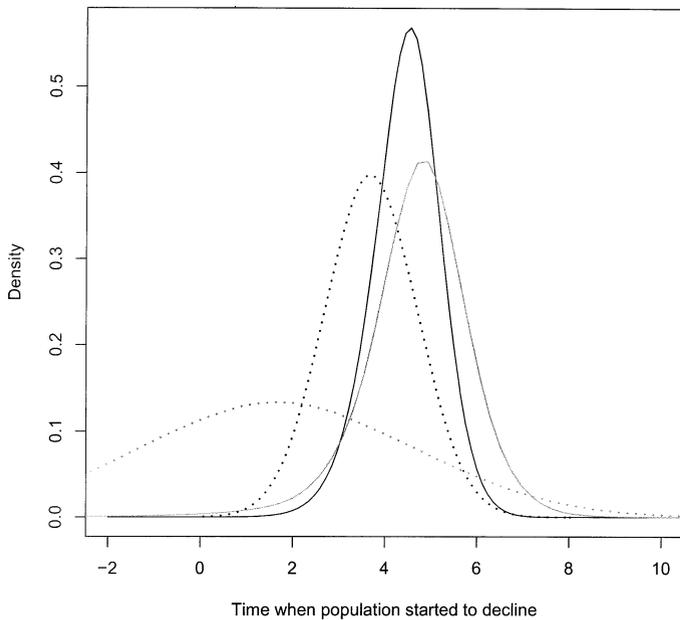


FIG. 5. The posterior distribution of the time when the *Cynopterus brachyotis* population started to decline,  $M_{xa}$ , is plotted for the longer-time (darker lines) and shorter-time (lighter lines) models. Prior distributions are shown as dotted lines.

time scale is supported by the data, this can be accommodated with the broader prior (Fig. 5).

#### DISCUSSION

Historical reductions in tropical forest cover in the Indian subcontinent are thought to have occurred in response to changes in temperature, precipitation, and levels of atmospheric  $\text{CO}_2$  (Ravindranath and Sukumar 1998). Several independent lines of evidence indicate that arid climatic conditions prevailed during and immediately following the last glacial maximum (20,000–16,000 years ago) and again in the mid-Holocene (5000–2000 years ago). Sources of evidence include stable-carbon isotope data (Sukumar et al. 1993, 1995; Rajagopalan et al. 1997, 1999), palynological data (Swain et al. 1983; Caratini et al. 1991, 1994), and the monsoon upwelling record from the Arabian Sea (Naidu 1996). Similarly, palynological evidence and climate-model simulations indicate that the southwest (summer) monsoon was severely attenuated during these periods (Prell and Kutzbach 1987; Phadtare 2000). There is also strong evidence to suggest that the more recent period of climatically induced deforestation may have been augmented by the activity of agri-pastoralist people in the Neolithic (5000–3000 years ago; Chandran 1997). The contemporary distribution of savannah and scrubland across much of the Indian subcontinent is thought to reflect a past history of forest-clearing by humans, beginning in the Neolithic and continuing apace to the present day (Pascal 1988; Chandran 1997; Prasad 1998).

The microsatellite data suggest that Indian populations of both *C. sphinx* and *C. brachyotis* have undergone pronounced demographic contractions in the past. Given the marked differences between the two species in geographic distribution and contemporary habitat associations, it is surprising that

TABLE 4. Modes and quantiles of the posterior distribution (on a  $\log_{10}$  scale) of  $M_{xa}$  for *Cynopterus sphinx* and *C. brachyotis* using two different priors.

	Model	Mode	Quantile		
			2.5%	50%	97.5%
<i>C. sphinx</i>	longer-time	3.4	1.91	3.25	4.58
	shorter-time	2.9	0.53	2.77	4.67
<i>C. brachyotis</i>	longer-time	4.5	2.86	4.48	5.87
	shorter-time	4.9	2.07	4.77	6.77

they appear to share such similar patterns of demographic history. Results of the microsatellite analysis are consistent with the hypothesis that the *C. brachyotis* population in the Western Ghats represents an isolated relic of what was formerly a widespread, continuously distributed population. Genetic evidence for a population contraction in this species confirms our expectations based on considerations of the vegetational history and paleoclimate of the Indian subcontinent. However, similar evidence for a population contraction in *C. sphinx* came as a surprise. The available evidence suggests that Quaternary climatic changes promoted a dramatic expansion of secondary forest-thornscrub and savannah at the expense of primary wet-evergreen forest across much of the Indian subcontinent. If contemporary habitat associations are any indication, this biogeographic scenario predicts that *C. sphinx* should have undergone a pronounced range expansion since the late Quaternary. Results of our analysis do not support this prediction: The microsatellite data indicated an even more pronounced population contraction in *C. sphinx* than in *C. brachyotis*. Thus, a model of uniform range expansion does not apply to either species.

A species' ability to adapt to long-term environmental change suggests that contemporary habitat associations should not be considered an infallible guide to historical inference. The multifarious nature of climatically induced habitat change may often produce unexpected effects on a species' geographic range size and/or effective population size. It is also possible that populations of both species have been adversely affected by destruction of roosting and foraging habitat as a result of human activity on a more recent time scale. Thus, genetic evidence for population decline in *C. sphinx* may also reflect the consequences of human-induced habitat destruction in the relatively recent past.

A novel feature of the model used in this analysis is the ability to allow for interlocus variation in demographic parameters. Factors that could potentially produce such variation include rare mutations that involve multistep changes in microsatellite allele lengths (discussed below), hitchhiking effects due to selection at linked sites (Hudson and Kaplan 1988; Kaplan et al. 1989), and background selection in genomic regions with low recombination rates (Charlesworth et al. 1993; Hudson and Kaplan 1995). Ideally, such possibilities should be modeled explicitly. However, in practice it is impossible to construct general likelihood-based models to incorporate all these possibilities, and it would be more consistent with general statistical practice to use simpler models that allow goodness-of-fit to be assessed.

Results of the present analysis suggest that the data are consistent with either extensive variation among loci or the

absence of variation among loci in most parameters (the only exceptions are mutation rate and/or current population size in *C. brachyotis*). This ambiguous conclusion is often a feature of hierarchical models (see discussion in Gelman et al. 1995, pp. 144–148). It is possible to demonstrate evidence against a hypothesis of zero variation in parameters, but it is difficult to conclude that the variation is zero. This latter possibility could be tested using Bayes factors in a reversible jump model, which would be a straightforward extension of the current method. However, Gelman et al. (1995) caution against such an approach because of the sensitivity of the conclusions to the priors (Gelman et al. 1995, pp. 176–177). Much of the apparent variation in parameter values among loci may stem from the priors chosen to model the loci. For both species, simulations with  $\tau_{N0} = \tau_{N1} = \tau_{xa} = 0.05$  and  $\tau_{\mu} = 1$  give posterior distributions for  $V_{N0}$ ,  $V_{N1}$ , and  $V_{xa}$  that are identical to the priors, but give posterior distributions for  $V_{\mu}$  that are markedly different from the priors, with low density at zero (results not shown). This suggests that results based on less informative priors could be explained entirely by variation in mutation rate among loci. In light of the results obtained, the less informative priors (which have strong support for the possibility of a five-fold ratio of demographic parameters among loci) are probably unrealistic. It is perhaps better to construct an initial model with priors that favor less variation in the demographic parameters, thereby allowing the possibility of substantial variation in mutation rate among loci. The simulations with lower values of  $\tau$  gave prior support for an expected difference of about 20% in demographic parameters among pairs of loci. Despite this tighter prior, a strong discrepancy among loci in a demographic parameter would be readily apparent, because the distribution of  $V$  would be pulled away from zero. Interestingly, demographic inferences from the model with lower values of  $\tau$  were very similar to those presented here. For both species, the model indicated somewhat stronger support for population contraction, tighter estimates for  $t_f$ , and very similar estimates for  $M_{xa}$ . Thus, inferences from the model are generally robust to the assumed priors for the degree of interlocus variation in demographic parameters, provided that mutation rate is allowed to vary. This is supported by the similarity in posterior distributions between the basic model, which allowed no variation in the demographic parameters, and the hierarchical model (Figs. 3B, 4B). Although we would have reached similar conclusions if we had used the basic model alone, the hierarchical model allowed us to assess whether our results had been influenced by discrepancies among loci.

Inferences about demographic history are potentially biased if the assumptions of the stepwise mutation model are violated (Gonser et al. 2000). For example, mutations that involve multistep changes in allele length can produce a pattern of variation similar to that produced by a population bottleneck. The relative frequency of such mutations is the subject of considerable debate (reviewed by Ellegren 2000b). Indirect studies based on frequency distributions of allelic length variants have provided inconclusive evidence about the general applicability of the stepwise mutation model to microsatellite data (Estoup and Cornuet 1999; Balloux et al. 2000). However, analyses of Y-chromosome microsatellites based on the stepwise mutation model have provided strong

evidence for demographic expansion of the human population, corroborating inferences from other markers (Pritchard et al. 1999; Thomson et al. 2000). Direct pedigree analysis suggests that more than 90–95% of microsatellite mutations involve single step changes and the remainder typically involve a small number of steps (Brinkmann et al. 1998). Such studies indicate that interlocus variation in mutation rate is largely attributable to variation in mean allele length (Ellegren 2000a; Schlötterer 2000). Given the current state of knowledge about mutational dynamics of microsatellite loci, it is perhaps advisable to give less weight to inferences drawn from population data because results may be confounded by historical demographic events that are not accounted for by the models. If multistep mutational changes are infrequent relative to single step changes, as indicated by the available pedigree evidence, then only a small number of loci should be affected. The hierarchical model will tend to give less weight to rare discrepant loci. Furthermore, although it is possible that there is variability in demographic parameters among loci, zero variance in demographic parameters has the highest posterior probability density when we assume that mutation rate varies among loci.

The present analysis suggests that the contemporary effective sizes of the *C. sphinx* and *C. brachyotis* populations are  $10^3$  to  $10^4$  times lower than the estimated census numbers. One possible explanation is that the priors for the mean mutation rate have been set too high. As pointed out in the Results, all estimates of population sizes and times depend on the chosen prior for the mean mutation rate among loci. The mean of this prior is somewhat lower than most current estimates of mutation rates for autosomal microsatellite loci in humans (Brinkmann et al. 1998; Ellegren 2000b), but it is possible that microsatellite mutation rates in bats are even lower. It is also possible that our estimates of census population size are inaccurate or that the model of exponential decline fits the data poorly (in which case a different demographic model might give higher estimates of effective population size). For example, a model of linear decline gives different parameter estimates from a model of exponential decline (Beaumont 1999). However, pronounced disparities between long-term effective population size and contemporary census number have been documented in several vertebrate species (e.g., Avise et al. 1988; O’Ryan et al. 1998). At the spatial and temporal scale relevant to the present study of *C. sphinx* and *C. brachyotis*, long-term effective population size may be reduced to levels well below the census population size by fluctuations in population size and/or extinction-recolonization dynamics (Hedrick and Gilpin 1997; Vucetich et al. 1997; Whitlock and Barton 1997; Nunney 1999; Wang and Caballero 1999).

This study illustrates the way in which genetic data can be integrated with prior information in a Bayesian statistical framework to test alternative hypotheses about demographic history. Similar analyses of other codistributed taxa will provide a clearer picture of the zoogeographic history of the Indian subcontinent.

#### ACKNOWLEDGMENTS

We thank J. Huelsenbeck, B. Payseur, and one anonymous reviewer for insightful comments on a previous draft of the

manuscript. For invaluable assistance in the field, JFS thanks H. R. Bhat, J. Balasingh, P. T. Nathan, A. A. Prakash, and D. P. Swami Doss. JFS acknowledges support from an NRSA Postdoctoral Fellowship from the National Institutes of Health and a Fellowship in Computational Molecular Biology from the Alfred P. Sloan Foundation and U.S. Department of Energy.

## LITERATURE CITED

- Awise, J. C., R. M. Ball, and J. Arnold. 1988. Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Mol. Biol. Evol.* 5: 331–344.
- Balasingh, J., J. Ronald, P. T. Nathan, and S. S. Isaac. 1999. Occurrence of *Cynopterus brachyotis* (Chiroptera: Pteropodidae) in Kalakad Mundanthurai Tiger Reserve, southern India. *Curr. Sci.* 76:1542.
- Balloux, F., H. Brüner, N. Lugon-Moulin, J. Hausser, and J. Goudet. 2000. Microsatellites can be misleading: an empirical and simulation study. *Evolution* 54:1414–1422.
- Bates, P. J. J., and D. L. Harrison. 1997. Bats of the Indian subcontinent. Harrison Zoological Museum, Sevenoaks, Kent, U.K.
- Beaumont, M. A. 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153:2013–2029.
- Beaumont, M. A., and M. W. Bruford. 1999. Microsatellites in conservation genetics. Pp. 165–182 in D. B. Goldstein and C. Schlötterer, eds. *Microsatellites: evolution and applications*. Oxford Univ. Press, Oxford, U.K.
- Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B* 263:1619–1626.
- Best, N. G., M. K. Cowles, and S. K. Vines. 1995. CODA Manual version 0.30. MRC Biostatistics Unit, Cambridge, U.K.
- Blanford, W. T. 1901. Distribution of vertebrate animals in India, Ceylon, and Burma. *Phil. Trans. R. Soc. Lond. B.* 194:335–436.
- Brandon-Jones, D. 1996. The Asian Colobinae (Mammalia: Cercopithecidae) as indicators of quaternary climatic change. *Biol. J. Linn. Soc.* 59:327–350.
- Brinkmann, B., M. Klintschar, F. Neuhuber, J. Hühne, and B. Rolf. 1998. Mutation rate in human microsatellites: influence of structure and length of the tandem repeat. *Am. J. Hum. Genet.* 62: 1408–1415.
- Brooks, S. P. 1998. Markov chain Monte Carlo method and its application. *Statistician* 47:69–100.
- Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7:434–455.
- Caratini, C., M. Fontugne, J. P. Pascal, C. Tissot, and I. Bentaleb. 1991. A major change at ca. 3,500 years BP in the vegetation of the Western Ghats in North Karnara, Karnataka. *Curr. Sci.* 61:669–672.
- Caratini, C., I. Bentaleb, M. Fontugne, M. T. Morzadeckefour, J. P. Pascal, and C. Tissot. 1994. A less humid climate since ca. 3,500 yr BP from marine cores off Karwar, western India. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 109:371–384.
- Chakraborty, R., and M. Kimmel. 1999. Statistics of microsatellite loci: estimation of mutation rate and pattern of population expansion. Pp. 139–150 in D. B. Goldstein and C. Schlötterer, eds. *Microsatellites: evolution and applications*. Oxford Univ. Press, Oxford, U.K.
- Chandran, M. D. S. 1997. On the ecological history of the Western Ghats. *Curr. Sci.* 73:146–155.
- Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303.
- Corbet, G. B., and J. E. Hill. 1992. The mammals of the Indomalayan region: a systematic review. British Museum Publications, Oxford Univ. Press, Oxford, U.K.
- Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin, and N. B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. USA* 91:3166–3170.
- Di Rienzo, A., P. Donnelly, C. Toomajian, B. Sisk, A. Hill, M. L. Petzl-Erler, G. K. Haines, and D. H. Barch. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics* 148: 1269–1284.
- Donnelly, P. 1999. The coalescent and microsatellite variability. Pp. 116–128 in D. B. Goldstein and C. Schlötterer, eds. *Microsatellites: evolution and applications*. Oxford Univ. Press, Oxford, U.K.
- Donnelly, P., and S. Tavaré. 1995. Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.* 29:410–421.
- Ellegren, H. 2000a. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat. Genet.* 24:400–402.
- . 2000b. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 16:551–558.
- Ender, J. A. 1982. Pleistocene forest refuges: fact or fancy? Pp. 641–657 in G. T. Prance, ed. *Biological diversification in the tropics*. Columbia Univ. Press, New York.
- Estoup, A., and J. M. Cornuet. 1999. Microsatellite evolution: inferences from population data. Pp. 49–65 in D. B. Goldstein and C. Schlötterer, eds. *Microsatellites: evolution and applications*. Oxford Univ. Press, Oxford, U.K.
- Feldman, M. W., J. Kumm, and J. Pritchard. 1999. Mutation and migration in models of microsatellite evolution. Pp. 98–115 in D. B. Goldstein and C. Schlötterer, eds. *Microsatellites: evolution and applications*. Oxford Univ. Press, Oxford, U.K.
- Frenzel, B., M. Pésci, and A. A. Velichko, eds. 1992. *Atlas of paleoclimates of the Northern Hemisphere: late Pleistocene-Holocene*. Gustav Fischer Verlag, Stuttgart, Germany.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian data analysis*. Chapman and Hall, London.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA.* 92:6723–6727.
- Goldstein, D. B., G. W. Roemer, D. A. Smith, D. E. Reich, A. Bergman, and R. K. Wayne. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* 151:797–801.
- Gonser, R., P. Donnelly, G. Nicholson, and A. Di Rienzo. 2000. Microsatellite mutations and inferences about human demography. *Genetics* 154:1793–1807.
- Goudet, J. 1995. FSTAT (Version 1.2): a computer program to calculate *F*-statistics. *J. Hered.* 86:485–486.
- Goudet, J., M. Raymond, T. de Meeüs, and F. Rousset. 1996. Testing differentiation in diploid populations. *Genetics* 144: 1933–1940.
- Harpending, H. C., M. A. Batzer, M. Gurven, L. B. Jorde, A. R. Rogers, and S. T. Sherry. 1998. Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA.* 95:1961–1967.
- Hedrick, P. W., and M. E. Gilpin. 1997. Genetic effective size of a metapopulation. Pp. 165–181 in I. A. Hanski and M. E. Gilpin, eds. *Metapopulation dynamics: ecology, genetics, and evolution*. Academic Press, San Diego, CA.
- Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913.
- Hora, S. L., and K. C. Jayaram. 1949. Remarks on the distribution of snakes of peninsular India with Malayan affinities. *Proc. Natl. Acad. Sci. India.* 15:399–402.
- Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* 7:1–44.
- Hudson, R. R., and N. L. Kaplan. 1988. The coalescent process in models with selection and recombination. *Genetics* 120: 831–840.
- . 1995. Deleterious background selection with recombination. *Genetics* 141:1605–1617.
- Ihaka, R., and R. Gentleman. 1996. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5:299–314.
- Jayaram, K. C. 1974. Ecology and distribution of fresh-water fishes, amphibia and reptiles. Pp. 517–584 in M. S. Mani, ed. *Ecology*

- and biogeography in India. Dr. W. Junk Publishers, The Hague, The Netherlands.
- Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The "hitchhiking effect" revisited. *Genetics* 123:887–899.
- Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins, and L. B. Jorde. 1998. Signatures of population expansion in microsatellite repeat data. *Genetics* 148:1921–1930.
- King, J. P., M. Kimmel, and R. Chakraborty. 2000. A power analysis of microsatellite-based statistics for inferring past population growth. *Mol. Biol. Evol.* 17:1859–1868.
- Kurup, G. U. 1974. Mammals of Assam and the mammal-geography of India. Pp. 585–613 in M. S. Mani, ed. *Ecology and biogeography in India*. Dr. W. Junk Publishers, The Hague, The Netherlands.
- Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175–195.
- Loader, C. R. 1996. Local likelihood density estimation. *Ann. Stat.* 24:1602–1618.
- Naidu, P. D. 1996. Onset of an arid climate at 3.5 ka in the tropics: evidence from monsoon upwelling record. *Curr. Sci.* 71:715–718.
- Nunney, L. 1999. The effective size of a hierarchically structured population. *Evolution* 53:1–10.
- O'Ryan, C., E. H. Harley, M. W. Bruford, M. A. Beaumont, R. K. Wayne, and M. I. Cherry. 1998. Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Anim. Conserv.* 1:85–94.
- Pascal, J. P. 1988. Wet evergreen forests of the Western Ghats of India: ecology, structure, floristic composition, and succession. French Institute, Pondicherry, India.
- Phillips, W. W. A. 1934. The short-nosed fruit bats (*Cynopterus*) of Ceylon. *Ceylon J. Sci.* 18B:237–248.
- . 1980. Manual of the mammals of Ceylon. Part I. Wildlife Protection Society of Sri Lanka, Colombo, Sri Lanka.
- Phadtare, N. R. 2000. Sharp decrease in summer monsoon strength 4,000–3,500 cal yr BP in the central higher Himalaya of India based on pollen evidence from alpine peat. *Quat. Res.* 53:122–129.
- Pradhan, M. S., and P. P. Kulkarni. 1997. Observations on the extended distribution of *Cynopterus brachyotis ceylonensis* Gray in western India with comments on its status. *Mammalia* 61:116–118.
- Prasad, S. N. 1998. Conservation planning for the Western Ghats of Kerala. II. Assessment of habitat loss and degradation. *Curr. Sci.* 75:228–235.
- Prell, W. L., and J. E. Kutzbach. 1987. Monsoon variability over the past 150,000 years. *J. Geophys. Res.* 92:8411–8425.
- Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16:1791–1798.
- Rajagopalan, G., R. Sukumar, R. Ramesh, R. K. Pant, and G. Rajagopalan. 1997. Late Quaternary vegetational and climatic changes from tropical peats in southern India: an extended record up to 40,000 years BP. *Curr. Sci.* 73:60–63.
- Rajagopalan, G., R. Ramesh, and R. Sukumar. 1999. Climatic implications of delta C-13 and delta O-18 ratios from C3 and C4 plants growing in a tropical montane habitat in southern India. *J. Biosci.* 24:491–498.
- Ravindranath, N. H., and R. Sukumar. 1998. Climate change and tropical forests in India. *Clim. Change* 39:563–581.
- Reich, D. E., and D. B. Goldstein. 1998. Genetic evidence for a Paleolithic human population expansion. *Proc. Natl. Acad. Sci. USA* 95:8119–8123.
- Ripley, S. D. 1949. Avian relicts and double invasions in peninsular India and Ceylon. *Evolution* 3:150–159.
- Roonwal, M. L., and B. Nath. 1949. Discontinuous distribution of certain Indo-Malayan mammals, and its zoogeographical significance. *Proc. Natl. Acad. Sci. India.* 15:375–377.
- Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365–371.
- Stephens, M., and P. Donnelly. 2000. Inference in molecular population genetics. *J. R. Stat. Soc. B.* 62:605–635.
- Storz, J. F. 2000. Variation at tri- and tetranucleotide repeat microsatellite loci in the fruit bat genus *Cynopterus* (Chiroptera: Pteropodidae). *Mol. Ecol.* 9:1198–2200.
- Storz, J. F., and T. H. Kunz. 1999. *Cynopterus sphinx*. *Mamm. Species* 613:1–8.
- Storz, J. F., J. Balasingh, P. T. Nathan, K. Emmanuel, and T. H. Kunz. 2000a. Dispersion and site-fidelity in a tent-roosting population of the short-nosed fruit bat (*Cynopterus sphinx*) in southern India. *J. Trop. Ecol.* 16:117–131.
- Storz, J. F., H. R. Bhat, and T. H. Kunz. 2000b. Social structure of a polygynous tent-making bat, *Cynopterus sphinx* (Megachiroptera). *J. Zool. (Lond.)* 251:151–165.
- Storz, J. F., J. Balasingh, H. R. Bhat, P. T. Nathan, A. A. Prakash, D. P. Swami Doss, and T. H. Kunz. 2001a. Clinal variation in body size and sexual dimorphism in an Indian fruit bat, *Cynopterus sphinx* (Chiroptera: Pteropodidae). *Biol. J. Linn. Soc.* 72:17–31.
- Storz, J. F., H. R. Bhat, and T. H. Kunz. 2001b. Genetic consequences of polygyny and social structure in an Indian fruit bat, *Cynopterus sphinx*. I. Inbreeding, outbreeding, and population subdivision. *Evolution* 55:1215–1223.
- . 2001c. Genetic consequences of polygyny and social structure in an Indian fruit bat, *Cynopterus sphinx*. II. Variance in male mating success and effective population size. *Evolution* 55:1224–1232.
- Sukumar, R., R. Ramesh, R. K. Pant, and G. Rajagopalan. 1993. A  $\delta^{13}\text{C}$  record of late Quaternary climate change from tropical peats in southern India. *Nature* 364:703–706.
- Sukumar, R., H. S. Suresh, and R. Ramesh. 1995. Climate change and its impact on tropical montane ecosystems in southern India. *J. Biogeogr.* 22:533–536.
- Swain, A. M., J. E. Kutzbach, and S. Hastenrath. 1983. Monsoon climate of Rajasthan for the Holocene: estimates of precipitation based on pollen lake levels. *Quat. Res.* 19:1–17.
- Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518.
- Thomson, R., J. K. Pritchard, P. D. Shen, P. J. Oefner, and M. W. Feldman. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* 97:7360–7365.
- Vucetich, J. A., T. A. Waite, and L. Nunney. 1997. Fluctuating population size and the ratio of effective to census population size. *Evolution* 51:2017–2021.
- Wakeley, J. 1999. Nonequilibrium migration in human history. *Genetics* 153:1863–1871.
- Wang, J., and A. Caballero. 1999. Developments in predicting the effective size of subdivided populations. *Heredity* 82:212–226.
- Whitlock, M. C., and N. H. Barton. 1997. The effective size of a subdivided population. *Genetics* 146:427–441.
- Willis, K. J., and R. J. Whittaker. 2000. The refugial debate. *Science* 287:1406–1407.
- Wilson, I. J., and D. J. Balding. 1998. Genealogical inference from microsatellite data. *Genetics* 150:499–510.

Corresponding Editor: J. Huelsenbeck