Genet. Res., Camb. (1997), **69**, *pp.* 111–116. *With* 3 *figures.* Printed in the United Kingdom © 1997 Cambridge University Press

On the estimation of ancestral population sizes of modern humans

ZIHENG YANG
Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA

(*Received 17 October 1996 and in revised form 12 January 1997*)

Summary

The theory developed by Takahata and colleagues for estimating the effective population size of ancestral species using homologous sequences from closely related extant species was extended to take account of variation of evolutionary rates among loci. Nuclear sequence data related to the evolution of modern humans were reanalysed and computer simulations were performed to examine the effect of rate variation on estimation of ancestral population sizes. It is found that the among-locus rate variation does not have a significant effect on estimation of the current population size when sequences from multiple loci are sampled from the same species, but does have a significant effect on estimation of the ancestral population size using sequences from different species. The effects of ancestral population size, species divergence time and among-locus rate variation are found to be highly correlated, and to achieve reliable estimates of the ancestral population size, effects of the other two factors should be estimated independently.

1. Introduction

Takahata (1986) suggested an ingenious idea for estimating the effective population size of the common ancestor of two closely related species when homologous DNA sequences from the two species are available at many loci. A pair of homologous genes sampled from two species must have diverged in the ancestral species. Sequences from different loci share the same species divergence time, while their coalescent times in the ancestral population vary according to the population size of the ancestral species. Thus the average sequence divergence at many loci provides information on the species divergence time, while the variation among loci of sequence divergences reflects the effective population size of the ancestral species. Takahata (1986) initially based his analyses on the mean and variance of the number of substitutions between pairs of sequences sampled from two species, and the approach was later extended to a full likelihood analysis and to the case of three species, where the population sizes of the two extinct ancestors were both estimated (Takahata *et al.*, 1995). The authors estimated the population sizes of ancestral species of humans, chimpanzees and gorillas, using synonymous, pseudogene and intron differences (referred to as 'silent' differences) at many nuclear loci. They estimated the effective population size of modern humans to be around 10000, and that of the ancestor of humans and chimpanzees to be around 100000, indicating a very populous ancestral species.

As discussed by Takahata (1986) and Takahata *et al.* (1995), the analysis was based on several assumptions, such as no recombination between sites within the same locus, free recombination between loci, neutral mutation and constancy of mutation (substitution) rates among loci. The assumption of rate constancy is a worrying one, as estimation of ancestral population size relies on the variation among loci of sequence divergence between species, while this variation can be caused by different evolutionary rates at different loci. Although silent substitution rates are generally thought to be homogeneous among loci, recent studies suggest that substantial synonymous rate variation may exist (e.g. Wolfe & Sharp, 1993). In this study, I will extend the approach of Takahata *et al.* (1995) to account for possible variation of evolutionary rates among loci, and examine the effect of such rate variation on estimation of ancestral population sizes. As in Takahata *et al.* (1995), the case of one species will be considered first. The case of sampling two individuals at each locus and a constant evolutionary rate at all loci considered by Takahata *et al.* (1995) will be extended to the case of sampling many individuals for each locus and variable rates among loci. Then the case of two species will be considered, and the maximum likelihood estimation of Takahata *et al.* (1995) extended to account for

variable rates among loci. The theory for the case of three species of Takahata *et al.* (1995) can similarly be extended, but no analysis is performed here due to paucity of such data.

2. One species

Let $\theta = 4N\mu$, where N is the effective population size of the species and μ is the substitution rate per nucleotide site. Following Takahata *et al.* (1995), I first consider the case of sampling two sequences at each locus. The data are pairs of DNA sequences (or silent sites in a homologous gene) from the same species. Suppose there are p loci, and the ith locus has n_i silent sites, with $i = 1, 2, \ldots, p$. Let k_i be the number of substitutions observed in the sequence pair at the ith locus; under the infinite-sites model (Kimura, 1969), this is the number of different sites in the two sequences. The data can then be represented as $\mathbf{k} = \{k_1, k_2, \ldots, k_p\}$. Note that the substitution rate for the entire ith locus is $n_i\mu$. The probability of observing k_i substitutions at the ith locus is given by the geometric distribution as

$$p(k_i) = \int_0^\infty \frac{\mathrm{e}^{-n_i\mu\cdot 2t}(n_i\mu\cdot 2t)^{k_i}}{k_i!} \times \frac{1}{2N}\mathrm{e}^{-\frac{1}{2N}t}\,\mathrm{d}t$$
$$= \frac{1}{1+n_i\theta}\left(\frac{n_i\theta}{1+n_i\theta}\right)^{k_i} \quad (1)$$

(Watterson, 1975). Nucleotide substitutions are assumed to occur independently at different loci, and the log-likelihood function is given by

$$l(\theta;\mathbf{k}) = \sum_{i=1}^{p} \log\{p(k_i)\} \quad (2)$$
$$= \sum_{i=1}^{p}\left\{k_i \log\left(\frac{n_i\theta}{1+n_i\theta}\right) - \log(1+n_i\theta)\right\} \quad (3)$$

(Takahata *et al.*, 1995). The maximum likelihood estimate of θ can be obtained using a numerical optimization algorithm.

To account for the variation of evolutionary rates among loci, the gamma distribution is used as an approximation. The substitution rate (per site) for a locus may be represented as θr, where r is a gamma variate normalized to have mean one, so that μ and θ reflect the average substitution rate across loci. The density function of r is then

$$f(r) = \frac{\alpha^\alpha}{\Gamma(\alpha)} r^{\alpha-1}\,\mathrm{e}^{-\alpha r}, \quad r > 0, \alpha > 0. \quad (4)$$

The shape parameter α is inversely related to the extent of rate variation with $\mathrm{var}(r) = 1/\alpha$; $\alpha = \infty$ corresponds to the case of a constant rate for all loci.

Ohta (1995: table 3) estimated synonymous substitution rates in 49 protein-coding genes for the three branches in a star tree of primates, artiodactyls and rodents. If we ignore possible errors (both sampling and systematic) involved in those estimates, crude estimates of α are obtainable. For the primate branch, the estimated number of synonymous substitutions per site ranges from 0·029 to 0·313 among loci, with mean $\bar{x} = 0{\cdot}1340$ and standard deviation $\sigma = 0{\cdot}0616$. Using a method of moments to fit the gamma distribution gives $\hat{\alpha} = (\bar{x}/\sigma)^2 = 4{\cdot}7$. The estimate for the artiodactyl branch is $\hat{\alpha} = (0{\cdot}1835/0{\cdot}0657)^2 = 7{\cdot}8$ while the estimate for the rodent branch is $\hat{\alpha} = (0{\cdot}3536/0{\cdot}1032)^2 = 11{\cdot}7$. (Estimates of α for non-synonymous substitution rates across loci using Ohta's data are 0·97, 1·08 and 0·95 for the primate, artiodactyl and rodent branches, respectively; these estimates are much smaller than those for synonymous sites, indicating much more severe rate variation at the non-synonymous sites.) Although the genes used by Ohta (1995) are not the same as those used by Takahata *et al.*, $\alpha = 5$ or 10 seem to be reasonable values for nuclear synonymous sites. In later analysis, several values are used for α to examine its effect.

Conditional on the rate (r_i) for the ith locus, the probability of observing k_i substitutions at the locus is given by

$$p(k_i\,|\,r_i) = \frac{1}{1+n_i\theta r_i}\left(\frac{n_i\theta r_i}{1+n_i\theta r_i}\right)^{k_i}. \quad (5)$$

The unconditional probability is given by

$$p(k_i) = \mathscr{E}\{p(k_i\,|\,r_i)\}$$
$$= \int_0^\infty \frac{1}{1+n_i\theta r_i}\left(\frac{n_i\theta r_i}{1+n_i\theta r_i}\right)^{k_i} f(r_i)\,\mathrm{d}r_i. \quad (6)$$

The log-likelihood function is given by (2). Analytical solution of the above integral seems difficult; instead, the discrete-gamma model of Yang (1994) can be used for efficient calculation. Several equal-probability categories are used to approximate the gamma distribution, with the mean of each category used to represent all rates in that category. In this study, 16 categories are used to achieve a good approximation. Thus in place of (6) we have

$$p(k_i) = \sum_{j=1}^{16} f(\bar{r}_j)\,p(k_i\,|\,\bar{r}_j), \quad (7)$$

where $f(\bar{r}_j) = \frac{1}{16}$ and the rate ($\bar{r}_j$) for category j is calculated as a function of parameter α (Yang, 1994).

When more than two sequences are available at a locus, the likelihood function can be calculated similarly. Suppose that m_i individuals are sampled from the population at the ith locus. The probability of observing k_i substitutions at n_i sites in a sample of size m_i is given by

$$p(k_i) = \frac{m_i - 1}{n_i\theta}\sum_{j=0}^{m_i-2}(-1)^j\binom{m_i-2}{j}\left(\frac{n_i\theta}{n_i\theta+j+1}\right)^{k_i+1} \quad (8)$$

(Tavaré, 1984). The probability under the model of variable rates among loci can be calculated in a similar way to (6) and (7), by conditioning on the rate for the locus.

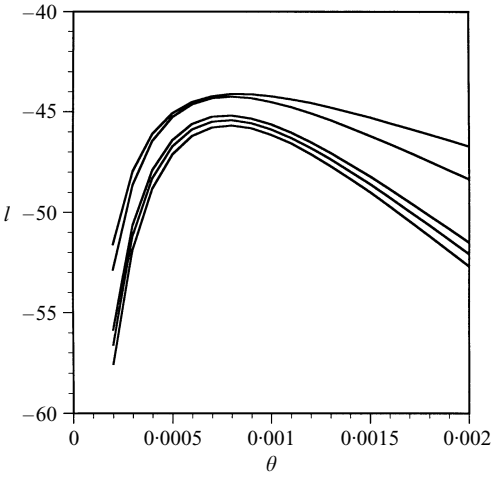

Fig. 1. Log-likelihood curves as functions of $\theta = 4N\mu$ for different values of the gamma parameter α. Pairs of human sequences at 49 loci are analysed. The five curves correspond, from top to bottom, to $\alpha = \infty$, 10, 5, 1 and 0·5.

In the following, I apply the gamma distribution model to the data analysed by Takahata *et al.* (1995: table 1). The data are numbers of silent differences between pairs of sequences at 49 loci, compiled by Li & Saddler (1991). Fig. 1 shows the likelihood curves as functions of θ for different values of α. The estimate of θ under the constant rate model ($\alpha = \infty$) is $0{\cdot}000\,78 \pm 0{\cdot}000\,20$ (the standard error is calculated by the 'curvature' method, which inverts the second-order derivative of the log-likelihood function with respect to the parameter). The estimate is essentially identical to that of Takahata *et al.* (0·0008), although those authors did not calculate the standard errors of parameter estimates. With an average silent substitution rate of 10^{-9} substitutions per site per year and a generation gap of 20 years (Takahata *et al.*, 1995), this estimate of θ gives an estimate of the human effective population size of $0{\cdot}000\,78/(4 \times 10^{-9} \times 20) = 9750$ with a standard error of 2500. The maximum likelihood estimates of θ are $0{\cdot}000\,78 \pm 0{\cdot}000\,21$, $0{\cdot}000\,77 \pm 0{\cdot}000\,21$ and $0{\cdot}0079 \pm 0{\cdot}000\,26$ for $\alpha = 10$, 5 and 1, respectively. The estimate of θ is quite robust to changes in the value of α used, although the standard error of the estimate is slightly underestimated when the among-locus rate variation exists but is ignored.

3. Two species

Suppose that one sequence is sampled from each of the two species at each locus, and p loci are sampled. Let $\theta_0 = 4N_0\mu$, where N_0 is the effective population size of the common ancestor of the two species. Since only one individual is sampled at a locus for each species, the effective population sizes of the two extant species do not enter the analysis (Takahata *et al.*, 1995). Let τ be the divergence time of the two species, and $\gamma = \mu\tau$. The coalescent time (t) of the two sequences within the ancestral population is a random variable with an exponential distribution

$$f(t) = \frac{1}{2N_0}\exp\left\{-\frac{t}{2N_0}\right\}. \tag{9}$$

Since the two sequences diverged time $\tau + t$ ago,

$$p(k_i \mid t) = \frac{\{2n_i\mu(\tau+t)\}^{k_i}}{k_i!}\mathrm{e}^{-2n_i\mu(\tau+t)}. \tag{10}$$

The unconditional probability is then

$$p(k_i) = \int_0^\infty p(k_i \mid t) f(t)\,\mathrm{d}t \tag{11}$$

$$= \frac{1}{1+n_i\theta_0}\mathrm{e}^{-2n_i\gamma}\sum_{j=0}^{k_i}\frac{1}{j!}(2n_i\gamma)^j\left(\frac{n_i\theta_0}{1+n_i\theta_0}\right)^{k_i-j}, \tag{12}$$

where all loci are assumed to have the same substitution rate (Takahata *et al.*, 1995). When rates are variable among loci, we have

$$p(k_i \mid r_i) = \frac{1}{1+n_i\theta_0 r_i}\mathrm{e}^{-2n_i\gamma r_i} \times \sum_{j=0}^{k_i}\frac{1}{j!}(2n_i\gamma r_i)^j\left(\frac{n_i\theta_0 r_i}{1+n_i\theta_0 r_i}\right)^{k_i-j}. \tag{13}$$

The unconditional probability is given by

$$p(k_i) = \mathscr{E}\{p(k_i \mid r_i)\} = \int_0^\infty p(k_i \mid r_i) f(r_i)\,\mathrm{d}r_i. \tag{14}$$

As in the case of one species, the discretized gamma distribution (Yang, 1994) is used for efficient calculation. I consider two possibilities. The first is to estimate both parameters θ_0 and γ by maximizing the likelihood function, as done by Takahata *et al.*, while the second is to estimate θ_0 only, with γ estimated independently, for example from phylogenetic analysis.

Log-likelihood values and maximum likelihood estimates of θ_0 and γ are shown in Fig. 2 when α is fixed at different values. The data are numbers of silent differences between 15 pairs of sequences from humans and chimpanzees (Takahata *et al.*, 1995: table 2). Generally a smaller α (meaning more variable rates among loci) leads to a smaller estimate of θ_0 and a larger estimate of γ. The estimates under the constant-rate model ($\alpha = \infty$) are $\hat{\theta}_0 = 0{\cdot}004\,98 \pm 0{\cdot}002\,99$, $\hat{\gamma} = 0{\cdot}004\,63 \pm 0{\cdot}001\,54$. Use of the same substitution rate and generation time as in the one-species analysis with this estimate of θ_0 leads to a population size for the ancestor of humans and chimpanzees of $62\,200 \pm 37\,300$, similar to the estimate obtained by Takahata *et al.* (1995). However, for $\alpha = 10$, the estimates are $\hat{\theta}_0 = 0{\cdot}0025 \pm 0{\cdot}0047$, $\hat{\gamma} = 0{\cdot}0057 \pm 0{\cdot}0025$, which lead to an ancestral population

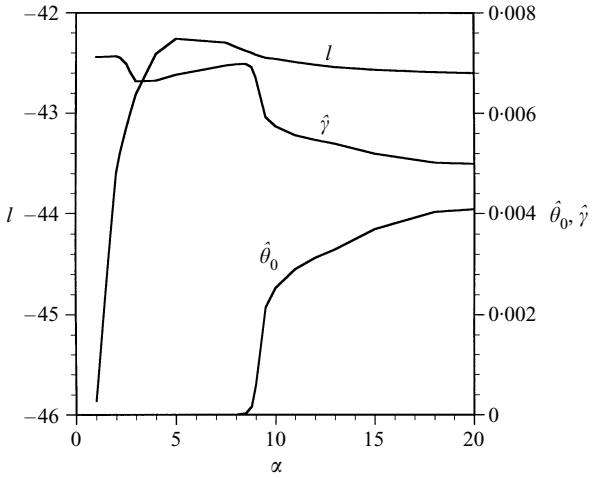

Fig. 2. Log-likelihood value (l) and maximum likelihood estimates of parameters θ_0 and γ for given values of the gamma parameter α. The parameters are $\theta_0 = 4N_0\mu$ and $\gamma = \mu\tau$, where N_0 is the effective population size of the ancestor of humans and chimpanzees and τ is the divergence time between the two species. Fifteen pairs of sequences from humans and chimpanzees are analysed.

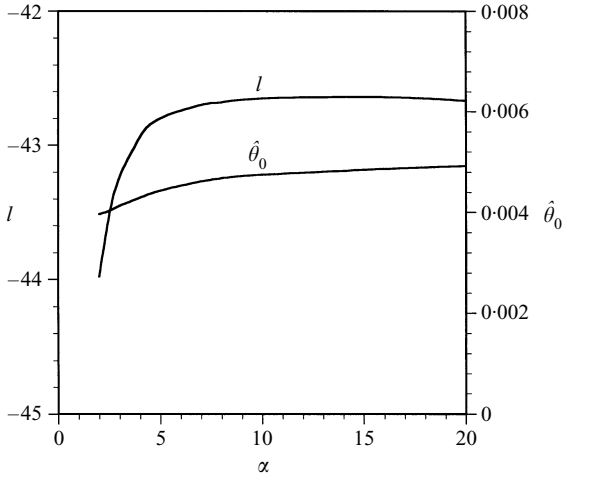

Fig. 3. Log-likelihood value (l) and maximum likelihood estimate of parameter θ_0 for given values of α. Parameter γ is fixed at 0·0045. See legend to Fig. 2.

of only half the size. When $\alpha < 8$, estimates of θ_0 are zero (Fig. 2). Estimates of θ_0 and thus N_0 are thus highly sensitive to the assumed rate variation among loci. These results contrast with the robustness of estimates of the population size of modern humans when samples from the single species are used. The reason is that estimation of the population size for a single species is largely an estimation of the average sequence divergence, which is not highly dependent on whether rate variation among loci is assumed in the model due to the low divergence of the sequences, while estimation of the population size of the common ancestor of two extant species relies on the variation of sequence divergences among loci. This variation can be caused by variable substitution rates.

Maximum likelihood estimates of θ_0 when γ is fixed at 0·0045 (corresponding to a human–chimpanzee separation of 4·5 million years ago) are shown in Fig. 3. Unlike the case where both θ_0 and γ are estimated (Fig. 2), estimates of θ_0 are much less sensitive to the value of α used, although the overall effect is the same in that use of a larger value for α leads to a smaller estimate for θ_0. For $\alpha = 10$ and 5, θ_0 is estimated to be $0{\cdot}00474 \pm 0{\cdot}00221$ and $0{\cdot}00443 \pm 0{\cdot}00251$, respectively, which correspond to population sizes of $59\,200 \pm 27\,600$ and $55\,300 \pm 31\,350$, respectively. These estimates are much greater than those for the modern human population, but the standard errors are so large that no reliable inference can be made about the population dynamics.

4. Computer simulations

To validate the estimation procedure and the empirical findings described above, a computer simulation was performed, the results of which are shown in Table 1. I used $\theta_0 = 0{\cdot}005$ and $\gamma = 0{\cdot}0045$, values typical of the human–chimpanzee data set. Three values were used for the number of loci (p): 50, 100 and 200. The gamma parameter α is fixed at 10 or 5. For each locus i, a random integer in the range (100, 5000) is generated for the number of sites (n_i); fixing the sequence length for each locus at the average (2550) was found to give similar results (not shown). A random variate from the discrete-gamma distribution (Yang, 1994) was generated for the rate at the locus (r_i), and a random variate from the exponential distribution with mean $\theta_0/2$ is generated for μt, where t is the coalescent time in the ancestral population (see equation 9). The number of differences at locus i (k_i) is then generated as a Poisson variate with parameter $2n_i(\gamma + \mu t)\, r_i$ (see equation 10). Each simulated data set (that is, p pairs of n_i and k_i) is used to estimate θ_0 and γ by maximum likelihood, with α fixed at ∞ (no rate variation), 10 or 5. Parameter θ_0 is also estimated with γ fixed at its true value (0·0045).

The simulations (Table 1) show similar patterns to the empirical results obtained from the human–chimpanzee data set (Figs. 1, 2). When both θ_0 and γ are estimated, the estimates ($\hat{\theta}_0$ and $\hat{\gamma}$) are very sensitive to the value of α assumed. Ignoring rate variation among loci leads to overestimation of θ_0 and underestimation of γ. Estimates obtained when the true value of α is used suggest that $\hat{\theta}_0$ and $\hat{\gamma}$ have negative and positive biases, respectively, for small numbers of loci, especially for small α. Fixing γ is noted to reduce greatly the effect of α and to lead to more accurate estimates of θ_0 ($\tilde{\theta}_0$). The curvature approximation to standard errors of maximum likelihood estimates is also found to be quite reliable (results not shown).

Table 1. *Effects of the assumed α on maximum likelihood estimates of θ_0 and γ*

Assumed α	$E(\hat{\theta}_0) \pm$ SE	$E(\hat{\gamma}) \pm$ SE	$E(\tilde{\theta}_0) \pm$ SE
	50 loci, true $\alpha = 10$		
∞	0·00753 ± 0·00014	0·00326 ± 0·00005	0·00550 ± 0·00011
10	0·00478 ± 0·00018	0·00464 ± 0·00008	0·00505 ± 0·00011
5	0·00219 ± 0·00020	0·00595 ± 0·00010	0·00479 ± 0·00011
	50 loci, true $\alpha = 5$		
∞	0·00946 ± 0·00015	0·00234 ± 0·00005	0·00595 ± 0·00012
10	0·00734 ± 0·00018	0·00343 ± 0·00008	0·00542 ± 0·00012
5	0·00446 ± 0·00027	0·00484 ± 0·00012	0·00514 ± 0·00013
	100 loci, true $\alpha = 10$		
∞	0·00758 ± 0·00010	0·00326 ± 0·00004	0·00555 ± 0·00007
10	0·00498 ± 0·00013	0·00457 ± 0·00006	0·00510 ± 0·00008
5	0·00254 ± 0·00017	0·00582 ± 0·00008	0·00484 ± 0·00008
	100 loci, true $\alpha = 5$		
∞	0·00923 ± 0·00012	0·00240 ± 0·00004	0·00584 ± 0·00009
10	0·00719 ± 0·00014	0·00345 ± 0·00005	0·00531 ± 0·00009
5	0·00465 ± 0·00019	0·00470 ± 0·00008	0·00503 ± 0·00010
	200 loci, true $\alpha = 10$		
∞	0·00755 ± 0·00006	0·00324 ± 0·00003	0·00549 ± 0·00005
10	0·00502 ± 0·00008	0·00451 ± 0·00003	0·00504 ± 0·00006
5	0·00277 ± 0·00010	0·00567 ± 0·00004	0·00478 ± 0·00006
	200 loci, true $\alpha = 5$		
∞	0·00930 ± 0·00008	0·00236 ± 0·00003	0·00583 ± 0·00007
10	0·00731 ± 0·00009	0·00338 ± 0·00003	0·00530 ± 0·00007
5	0·00495 ± 0·00012	0·00454 ± 0·00005	0·00503 ± 0·00007

Means and standard errors of maximum likelihood estimates are calculated from 100 simulated replicates. The true values of parameters used in the simulation are $\theta_0 = 0{\cdot}005$ and $\gamma = 0{\cdot}0045$. $\tilde{\theta}_0$ is the estimate when γ is fixed at its true value.

5. Discussion

Estimation of ancestral population sizes relies on the variation of sequence divergences among loci. However, at least three factors contribute to this variation: (1) variation of coalescent time in the ancestral population, which is affected by the ancestral population size (N_0 or θ_0), (2) variation of substitution rates, measured by α in this study, and (3) the stochastic fluctuation in the number of nucleotide substitutions, which is affected by both the coalescent time and the species separation time (γ). In real data, effects of these factors are confounded. For example, results of Table 1 suggest a strong negative correlation between estimates of θ_0 and γ with α fixed. Although in theory all three parameters (θ_0, γ, and α) could be estimated from the data and likelihood ratio tests could be used to examine hypotheses such as rate constancy among loci, the data clearly do not contain enough information for doing this. For example, neither the assumption of a constant substitution rate for all loci ($\alpha = \infty$) nor the assumption of a zero ancestral population size ($\theta_0 = 0$) can be rejected by such tests for the human–chimpanzee data set. In this study, external estimates of α are used, and both the real data analysis and computer simulations suggest the advantage of using independent estimates for the species separation time (γ). With the quantification of these different factors and accumulation of sequence data at more loci, it may eventually become possible to obtain reliable estimates of the ancestral population sizes.

An alternative approach to accounting for the rate variation among loci is to obtain estimates of relative evolutionary rates at different loci and then use such rates as constants. If the relative rate for locus i is estimated to be r_i, the log likelihood will be calculated as $l = \sum_{i=1}^{p} p(k_i \mid r_i)$, with $p(k_i \mid r_i)$ given by (13). Such estimates may be obtained from comparison of sequences of the same loci from more-divergent species, where ancestral polymorphisms are not important. Insofar as loci with high estimated relative rates also tend to exhibit more differences in the sequence pairs from the closely related species, which is most likely to be the case, this approach can be expected to produce similar results to those obtained here. However, reliable estimates of relative rates among loci can be expected to reduce the sampling error of the estimated ancestral population size.

It may be noted that the gamma distribution is used to approximate 'substitution rates' rather than 'mutation rates' among loci, since the differences observed in the data are more or less fixations after the screening of selection. Some selection effects are thus explicitly accommodated by the variable-rates models. Nevertheless, selection may render invalid the co-

alescent analysis implicitly employed in this study. Takahata *et al.* (1995) provided a nice discussion of the effects of selection and other factors such as recombination and population subdivision.

I thank B. Rannala, M. Slatkin, N. Takahata, J. Wakeley and two anonymous referees for comments and/or discussions. I am also indebted to members of Slatkin's laboratory for ideas to debug my program, especially needed because of the strange $\hat{\gamma}$ curve in Fig. 2. No error has been found, however, and the curve may be due to a special combination of data and model. Results reported in Table 1 and those not reported here indicate that the program is correct. It is available by anonymous ftp at our ftp site: `ftp://mw511.biol.berkeley.edu/pub/Ne.c`.
This study is supported by a grant from National Institute of Health (GM40282) to M. Slatkin.

References

Kimura, M. (1969). The number of heterogeneous nucleotide sites maintained in a finite population due to steady flux of mutation. *Genetics* **61**, 893–903.

Li, W-H. & Saddler, L. A. (1991). Low nucleotide diversity in man. *Genetics* **129**, 513–523.

Ohta, T. (1995). Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *Journal of Molecular Evolution* **40**, 56–63.

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. *Genetical Research*, *Cambridge* **48**, 187–190.

Takahata. N., Satta, Y. & Klein, J. (1995). Divergence time and population size in the lineage leading to modern humans. *Theoretical Population Biology* **48**, 198–221.

Tavaré, S. (1984). Lines of descent and geneaological processes, and their application in population genetics models. *Theoretical Population Biology* **26**, 119–164.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–276.

Wolfe, K. H. & Sharp, P. M. (1993). Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *Journal of Molecular Evolution* **37**, 441–456.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**, 306–314.