Time Dependency of Molecular Evolutionary Rates? Yes and No

Sankar Subramanian* and David M. Lambert*

Griffith School of Environment and School of Biomolecular and Physical Sciences, Griffith University, Nathan, Queensland, Australia

*Corresponding authors: E-mail: s.subramanian@griffith.edu.au; d.lambert@griffith.edu.au.

Accepted: 12 October 2011

Abstract

Some previous studies have suggested that rates of evolution inferred using molecular sequences vary substantially depending on the time frame over which they are measured, whereas a number of other studies have argued against this proposition. We examined this issue by separating positions of primate mitochondrial genomes that are under different levels of selection constraints. Our results revealed an order of magnitude variation in the evolutionary rates at constrained sites (including nonsynonymous sites, D-loop, and RNA) and virtually an identical rate of evolution at synonymous sites, independent of the timescales over which they were estimated. Although the evolutionary rate at nonsynonymous sites obtained using the European (H1 haplogroup) mitogenomes is 9–15 times higher than that estimated using the human–chimpanzee pair, in contrast, the rates at synonymous sites are similar between these comparisons. We also show that the ratio of divergence at nonsynonymous to synonymous sites estimated using intra- and interspecific comparisons vary up to nine times, which corroborates our results independent of calibration times.

Key words: rates of evolution, natural selection, neutral evolution, time dependency, divergence times, population coalescent times.

Introduction

Rates of molecular evolution are central to understanding the genetics and molecular biology of species. In the past, evolutionary rates have typically been estimated by calibrating the levels of evolutionary divergence (usually from extant species) with likely divergence times between species/populations. Although the evolutionary rates obtained using different calibration times should in principle be similar, an earlier study first noted much higher rates based on recent calibration times, compared with the rates estimated using older calibrations (Garcia-Moreno 2004). A later study using mitochondrial sequence data from primates and birds suggested that this pattern is more universal, and the authors proposed that rates of evolution are generally time dependent (Ho et al. 2005). Several studies of populations and closely related species reported higher evolutionary rates compared with those obtained from distantly related species and thus provided support for this concept (Burridge et al. 2008; Gratton et al. 2008; Howell et al. 2008; Henn et al. 2009; Soares et al. 2009). On the other hand, a number of studies refuted the idea of time dependency based on the methodological artifacts associated with the Bayesian Markov Chain Monte Carlo (MCMC) approach used to estimate evolutionary rates (Emerson 2007; Bandelt 2008; Debruyne and Poinar 2009; Navascues and Emerson 2009). These studies showed an upward bias in the rates estimated by the complex Bayesian MCMC methods using the data from populations or closely related species (Navascues and Emerson 2009), which was attributed to the poor or low signal in the data (Debruyne and Poinar 2009). Note that the rate of evolution mentioned above (and throughout this article) refers only to the inferred number of substitutions (or mutations) per site divided by the coalescence/divergence times. Hence, this does not suggest that the rate of mutation itself varies with time.

Population genetic theories predict a time-dependent rate at constrained sites (but not at neutral sites) due to the removal of slightly deleterious mutations over time. Therefore, the concept of time dependency needs to be re-examined by separating neutral and other constrained sites that are under different magnitudes of selective constraints. Furthermore, a simple method of rate estimation that uses a minimal number of parameters and assumptions is needed to avoid methodological biases reported previously. Hence, to examine this, we assembled data sets consisting of mitochondrial genome sequences belonging to human populations, Neanderthal, and chimpanzee.

Table 1
Estimates of Rates of Molecular Evolution

Group (Number of Genomes)	Calibration Times (Intervals), kyr	Rate of Evolution ($\times 10^{-8}$ s/s/year)				
		Synonymous Sites	Nonsynonymous Sites	RNA	D-Loop	d*N*/d*S*(SE)
European—H1 (83)	18 (11–25)	5.1 (8.4–3.7)	2.2 (3.5–1.6)	2.8 (4.6–2.0)	12.0 (19.0–8.6)	0.423 (0.173)
Australian (33)	45 (40–65)	6.9 (7.8–4.8)	1.8 (2.1–1.3)	2.1 (2.4–1.5)	18.0 (20.0–12.0)	0.264 (0.067)
Humans (100)[a]	150 (100–200)	3.7 (5.5–2.8)	0.8 (1.2–0.6)	1.2 (1.9–0.9)	7.8 (12.0–5.9)	0.210 (0.041)
Human–Neanderthal	500 (400–600)	3.4 (5.7–2.5)	0.5 (0.8–0.3)	0.6 (1.0–0.4)	4.4 (7.3–3.1)	0.136 (0.006)
Human–chimpanzee	6,000 (5,000–7,000)	4.3 (5.2–3.7)	0.2 (0.2–0.2)	0.7 (0.9–0.6)	4.3 (5.2–3.7)	0.047 (0.0005)
Ratio H1/human–chimpanzee		1.2 (1.0–1.6)	10.8 (9.1–14.8)	3.8 (3.2–5.2)	2.8 (2.3–3.8)	9.1 (5.4–12.5)

NOTE.—SE, standard error.
[a] Five mitogenomes each were taken from 20 major haplogroups.

Materials and Methods

Mitochondrial genome sequences of chimpanzee, Neanderthal, and human populations were obtained from GenBank. We obtained 83 sequences for which the names of the haplogroup (H1) were explicitly mentioned. Similarly, there were 33 sequences with explicit references to native Australian (Aborigine) in the source fields. To estimate the rate of evolution within humans, 100 mitogenomic sequences were used by collecting five representative sequences from 20 major haplogroups A, B, C, D, G, H, I, J, K, L0, L1, L2, L3, T, U, V, W, X, Y, and Z. Hence, 216 human genomes plus the genomes of Neanderthal and chimpanzee were used in this study. Individual orthologous protein sequences from 218 genomes were aligned, and this was used to align cDNA. Similarly, individual tRNA, rRNA genes, and D-loop regions were aligned. Note that the D-loop regions were available only for 20 and 31 genomes belonging to Australians and Europeans (H1), respectively. All positions with alignment gaps were excluded. To estimate rates at synonymous positions, 4-fold and 2-fold sites from all protein-coding genes were concatenated. Likewise, the 0-fold sites were used to estimate rates at nonsynonymous sites. All tRNA and rRNAs were concatenated into a single alignment.

To estimate the rate of evolution for the population data from European (H1), Australian, and within humans, the program "MCMCcoal" was employed using the option of "data analysis from one species" (Rannala and Yang 2003). All the default prior settings in the control file MCMCcoalYu2001.ctl (supplied with the program) were used. The summary statistics were extracted from the output file (mcmc.out) using the program "ds," which is provided along with the MCMCcoal software. This program estimates the coalescence distances or root heights (μt_{MRCA}) of the most recent common ancestor for a given set of sequences from a species (supplementary table S1, Supplementary Material online). The rates of evolution were determined by dividing μt_{MRCA} by the respective population coalescence times given in table 1.

To estimate rates of evolution between species, pairwise distances (and standard errors) between human–Neanderthal and human–chimpanzee were estimated (supplementary table S1, Supplementary Material online). In order to reduce estimation errors and to avoid any bias, the mean distance estimates between the 100 human genomes and chimpanzee or Neanderthal were obtained. For protein-coding genes, the "codeml" program of "PAML" (Yang 2007) was used to estimate d*N* and d*S* between species. For RNA and D-loop, the software "PAUP" (Swofford 2003) was used to estimate pairwise distances using the Hasegawa-Kishino-Yano plus Gamma plus invariant sites model. The above model was determined using the software "Modeltest" (Posada and Crandall 1998) using Bayesian Information criterion. Finally, the mean pairwise distances were divided by the species divergence times to obtain the rates of evolution. The program MCMCcoal was not used to obtain divergence between species. This is because this program assumes equal rate of evolution among sites and uses the simple Jukes–Cantor model to correct multiple substitution. These assumptions are sufficient only when the divergence is small (<0.2). Because the divergence between human and chimpanzee is >0.5 (at synonymous sites and D-loop), simple models will underestimate evolutionary distances (Nei and Kumar 2000). Evolutionary distance estimates and number of sites used are given in supplementary table S1, Supplementary Material online. Methods related to the rate estimations based on Bayesian MCMC analysis are given in the supplementary methods (supplementary tables S2 and S3, Supplementary Material online).

Results and Discussion

The data set was partitioned into synonymous sites, nonsynonymous sites, RNAs, and D-loop. The evolutionary distances between species pairs were estimated by PAUP (Swofford 2003) using the parameter estimates obtained from Modeltest (Posada and Crandall 1998). For the population data, the coalescence distance of the most recent common ancestor (μt_{MRCA}) or the root height was estimated using the software MCMCcoal (Rannala and Yang

2003), which uses the Jukes–Cantor model for multiple-hit corrections. The pairwise divergences and coalescence distances were obtained for all the four types of sites (supplementary table S1, Supplementary Material online), and rates of evolution were estimated by simply dividing these distances by the respective divergence/coalescence times.

We used five calibration time points spanning 18, 45, 150, 500, and 6,000 kyr, which adequately captures the rate of evolution in a wide range of timescales (table 1). The coalescence time of Europeans (H1 haplogroup) was based on the last glacial maximum (Achilli et al. 2004; Endicott and Ho 2008) and that for Australians was based on the oldest fossils of humans in Australia (Bowler et al. 2003). Other time points are based on the divergence time estimated using the nuclear data from human populations (Green et al. 2006; Gutenkunst et al. 2009) and from human/Neanderthal (Green et al. 2006) comparisons, respectively. We used the widely accepted 6 (5–7) million years for human/chimpanzee divergence.

The rate estimates using nonsynonymous sites, RNA, and D-loop showed a negative relationship with the calibration times (fig. 1). In contrast, the neutral evolutionary rates using synonymous sites were similar across all the timescales. The magnitude of the rate variation is highest for the nonsynonymous sites and lowest for the D-loop region. For instance, the evolutionary rate at the amino acid replacement sites estimated for the H1 European haplogroup with a coalescence age of 18 kyr was found to be 10 times (8–14 times) higher than that obtained using the human–chimpanzee pair (fig. 1*A*; table 1). The magnitude of the difference in the evolutionary rates between these time points are 7.5 (6–10) and 2.5 (2–3) times for the RNA and D-loop, respectively (fig. 1*B*). The extent of rate differences suggests the intensity of selective constraints on these regions. Although nonsynonymous sites and RNAs are well known to be under selective constraints, the present study reveals selection on the D-loop region of mitochondrial genomes. In order to examine the robustness of these results, we reanalyzed the data using Bayesian MCMC methods (supplementary table S2, Supplementary Material online) and obtained similar results (supplementary fig. S1 and table S3, Supplementary Material online).

Population genetic theories predict that the slightly deleterious mutations contribute diversity to the population for a short while, but that they are selected against over long timescales and are prevented from becoming fixed (Kimura 1983). Evidence for this prediction is very clear from figure 1. Over short timescales, evolutionary rates for nonsynonymous sequences are high probably due to the presence of such slightly deleterious mutations. Because these mutations are gradually eliminated over time, this is reflected in a steady decline in rates. In contrast, the mutations at synonymous sites are largely neutral and thus the accumulation of such mutations is relatively constant over time. This results in the similarity of neutral evolutionary rates across various timescales. Although the temporally declining rate pattern observed for RNAs is similar to that for amino acid replacement sites, the magnitude of the decline is comparatively small. This suggests relatively weak selective constraints on the former, in contrast to the latter. Interestingly, this study also reveals selection constraints in the D-loop region. Because the region is responsible for DNA replication, it is likely to harbor replication origins and other regulatory motifs associated with replication. However, the selection pressure on D-loop appears to be much weaker than that observed for the other constrained sites as the rate obtained for Europeans is only 2–3 times higher than that estimated for the human–chimpanzee pair (fig. 1*B*; table 1). This

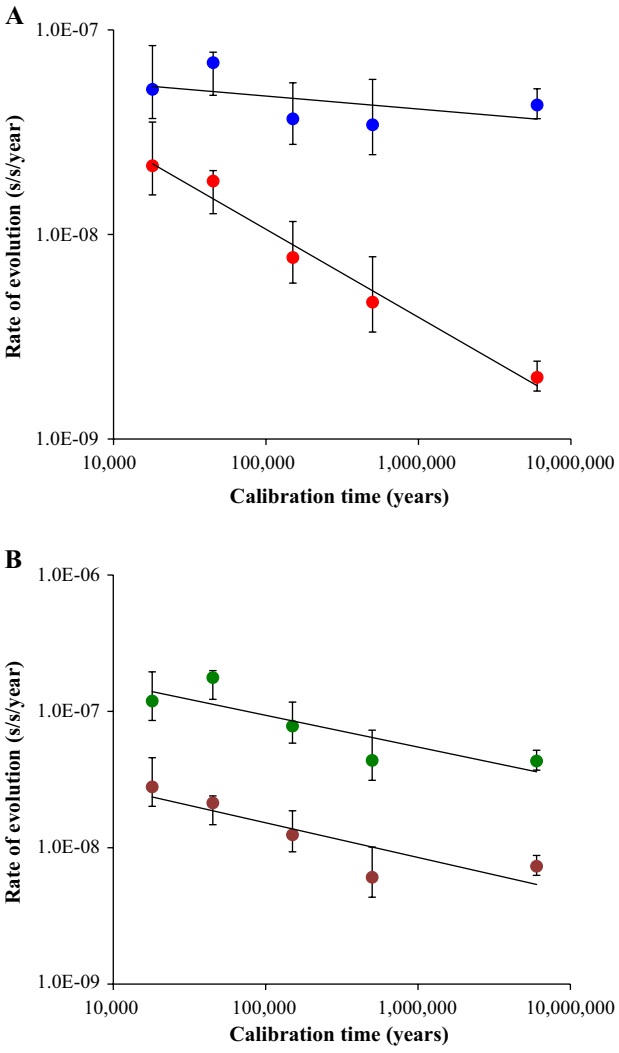

Fig. 1.—Relationships between rates of evolution and calibration times. (*A*) Rates of evolution were estimated for the nonsynonymous (red) and synonymous positions (blue) of mitochondrial protein-coding genes and for (*B*) RNAs (tRNA + rRNA, brown) and D-loop (green). Error bars are based on the lower and upper limits of the divergence/coalescence times. Both *x* and *y* axes are on a logarithmic scale. The best fitting regression lines are shown.

could be due to the fact that only a small fraction of sites in this region are likely to be under selection.

The results of this investigation are supported by earlier studies on birds and primates. A previous study revealed higher rate of constrained site evolution within a penguin lineage compared with that estimated between the lineages (Subramanian et al. 2009). However, the neutral rates of within and between lineage comparisons were found to be similar. Furthermore, a study on hominids also found significantly higher rates of nonsynonymous site evolution using intraspecific comparisons than that obtained for interspecific comparison (Endicott and Ho 2008). In contrast, evolutionary rates at synonymous sites were not significantly different between the two comparisons. Based on the similarity of neutral evolutionary rates, Soares et al. (2009) devised a method to correct the time-dependent effect in estimating evolutionary rates using human mitochondrial genomes.

Despite using a very wide range of divergence/coalescence times, the rates obtained in this study are still influenced by the accuracy of the time estimates used. Therefore, we reexamined the temporal rate patterns without using any divergence/coalescence times. Our results showed a huge difference in the nonsynonymous rates (rN) and a broad similarity in the synonymous rates (rS) across all timescales. Therefore, the ratio of the former to the later will reveal the temporal pattern of nonsynonymous evolution. This ratio can be simplified as $rN/rS = (dN/T)/(dS/T) = dN/dS$, where T is divergence/coalescence time, dN and dS are divergences at nonsynonymous and synonymous sites. Now, after eliminating the time component (T), we estimated the dN/dS ratio for the five data sets. Figure 2 reveals a negative correlation between the dN/dS ratios and calibration times. The pattern observed in this result is possible only if the rate of nonsynonymous evolution declines with calibration times and the synonymous rate is constant across different timescales. Therefore, independent of the use of calibration times, this result provides substantial support for the results shown in figure 1.

McDonald and Kreitman (1991) introduced a neutrality test, which compares the ratio of nonsynonymous to synonymous diversity (pN/pS) within a population and the ratio of nonsynonymous to synonymous divergence (dN/dS) obtained from interspecies comparison. Higher pN/pS compared with dN/dS is suggestive of the presence of deleterious nonsynonymous polymorphisms and the reverse mean adaptive amino acid substitutions. Later, Rand and Kann (1996) proposed a measure called neutrality index (NI), which is the ratio of these two ratios,

$$\mathrm{NI} = \frac{pN}{pS} \Big/ \frac{dN}{dS}. \quad (1)$$

This equation can be written as

$$\mathrm{NI} = \frac{rN_1T_1}{rS_1T_1} \Big/ \frac{rN_2T_2}{rS_2T_2}, \quad (2)$$

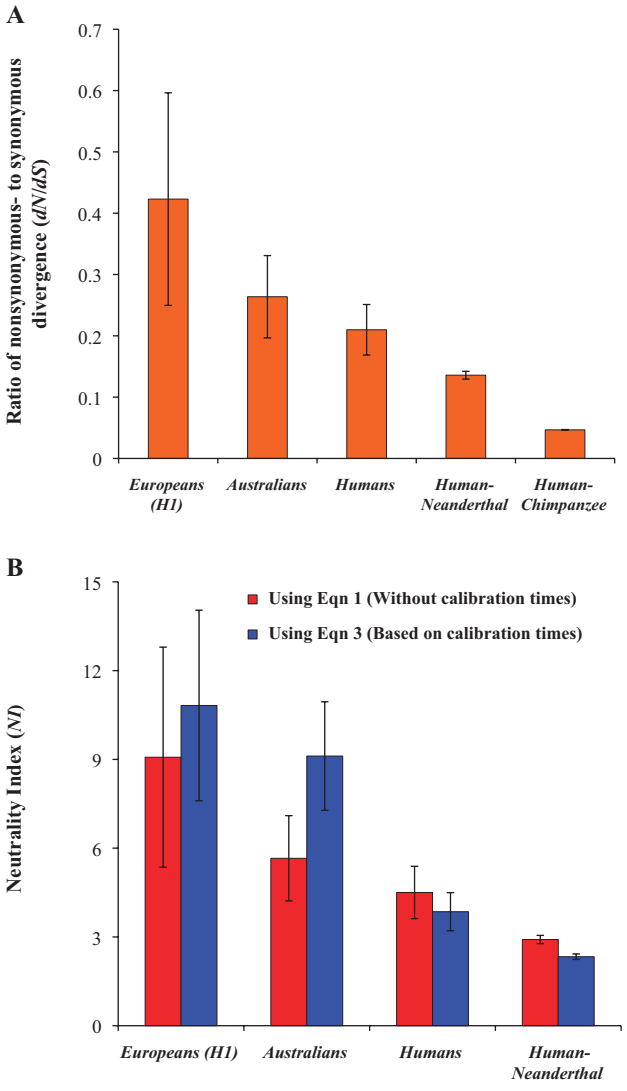

Fig. 2.—(*A*) Ratio of nonsynonymous to synonymous divergence (dN/dS) estimated for intra- and interspecific comparisons using the mitochondrial genes. Error bars are standard error of the mean. (*B*) Comparison of NI estimates obtained through two different methods. Red columns are the NI computed using equation (1) (without calibration times), in which the intraspecific dN/dS of different populations were divided by that obtained for the human–chimpanzee comparison. Blue columns are the NI estimated using equation (3), where the intraspecific nonsynonymous rates of evolution obtained for different populations were divided by the interspecific (human–chimpanzee) rate.

where rN_1 and rS_1 are intraspecific rates of evolution at synonymous and nonsynonymous sites, respectively, and rN_2 and rS_2 are interspecific rates of evolution at these corresponding sites. T_1 and T_2 are the population coalescence/divergence times within population and between species, respectively. Assuming that the rate of synonymous site evolution is similar between different timescales ($rS_1 = rS_2$), equation (2) can be simplified to

$$NI = \frac{rN_1}{rN_2}. \quad (3)$$

This suggests that NI is simply the ratio of the rates of nonsynonymous site evolution estimated within a species and between species. Therefore, the time-dependent variation in the rate of evolution at nonsynonymous sites could be determined using the divergences at synonymous and amino acid replacement sites alone, without using any calibration times. We computed NI by comparing the d*N*/d*S* ratios obtained for Europeans (H1), Australians, humans, and human/Neanderthal with the ratio estimated for the human–chimpanzee comparison using equation (1). We also estimated NIs by comparing the nonsynonymous rates (table 1 and fig. 1*A*) obtained for the human populations with that of human–chimpanzee using equation (3). Figure 2*B* shows that the NIs estimated using calibration times (2.3–10.8) are largely similar to those (2.9–9.1) estimated without using these times, and both clearly show time-dependent patterns of nonsynonymous rates of evolution. These estimates are very similar to the NIs obtained using the intra- and interspecies data from primates (Hasegawa et al. 1998), rodents (Rand and Kann 1996), and fruit flies (Rand et al. 1994). Therefore, the present and the previous studies provide solid evidence for the time dependency of molecular rates at amino acid replacement positions independent of calibration times.

Supplementary Material

Supplementary figure S1 and tables S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/). Alignments and input files for the programs MCMCcoal and BEAST can be obtained from http://www.mediafire.com/?udqdd25a9734jyi.

Acknowledgments

We are grateful to Australian Research Council discovery grant (2141900) and Griffith University for financial support.

Literature Cited

Achilli A, et al. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet. 75:910–918.
Bandelt HJ. 2008. Clock debate: when times are a-changin': time dependency of molecular rate estimates: tempest in a teacup. Heredity 100:1–2.
Bowler JM, et al. 2003. New ages for human occupation and climatic change at Lake Mungo, Australia. Nature 421:837–840.
Burridge CP, Craw D, Fletcher D, Waters JM. 2008. Geological dates and molecular rates: fish DNA sheds light on time dependency. Mol Biol Evol. 25:624–633.
Debruyne R, Poinar HN. 2009. Time dependency of molecular rates in ancient DNA data sets, a sampling artifact? Syst Biol. 58:348–360.
Emerson BC. 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. Syst Biol. 56:337–345.
Endicott P, Ho SY. 2008. A Bayesian evaluation of human mitochondrial substitution rates. Am J Hum Genet. 82:895–902.
Garcia-Moreno J. 2004. Is there a universal mtDNA clock for birds? J Avian Biol. 35:465–468.
Gratton P, Konopinski MK, Sbordoni V. 2008. Pleistocene evolutionary history of the Clouded Apollo (Parnassius mnemosyne): genetic signatures of climate cycles and a 'time-dependent' mitochondrial substitution rate. Mol Ecol. 17:4248–4262.
Green RE, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. Nature 444:330–336.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5: e1000695.
Hasegawa M, Cao Y, Yang Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol Biol Evol. 15:1499–1505.
Henn BM, Gignoux CR, Feldman MW, Mountain JL. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. Mol Biol Evol. 26:217–230.
Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol. 22:1561–1568.
Howell N, Howell C, Elson JL. 2008. Time dependency of molecular rate estimates for mtDNA: this is not the time for wishful thinking. Heredity 101:107–108.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.
Navascues M, Emerson BC. 2009. Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. Mol Ecol. 18:4390–4397.
Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press.
Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.
Rand DM, Dorfsman M, Kann LM. 1994. Neutral and nonneutral evolution of Drosophila mitochondrial-DNA. Genetics 138:741–756.
Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol. 13:735–748.
Rannala B, Yang ZH. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.
Soares P, et al. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet. 84:740–759.
Subramanian S, et al. 2009. High mitogenomic evolutionary rates and time dependency. Trends Genet. 25:482–486.
Swofford DL. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates.
Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Associate editor: B. Venkatesh