Supplementary Material

Supplementary Figures

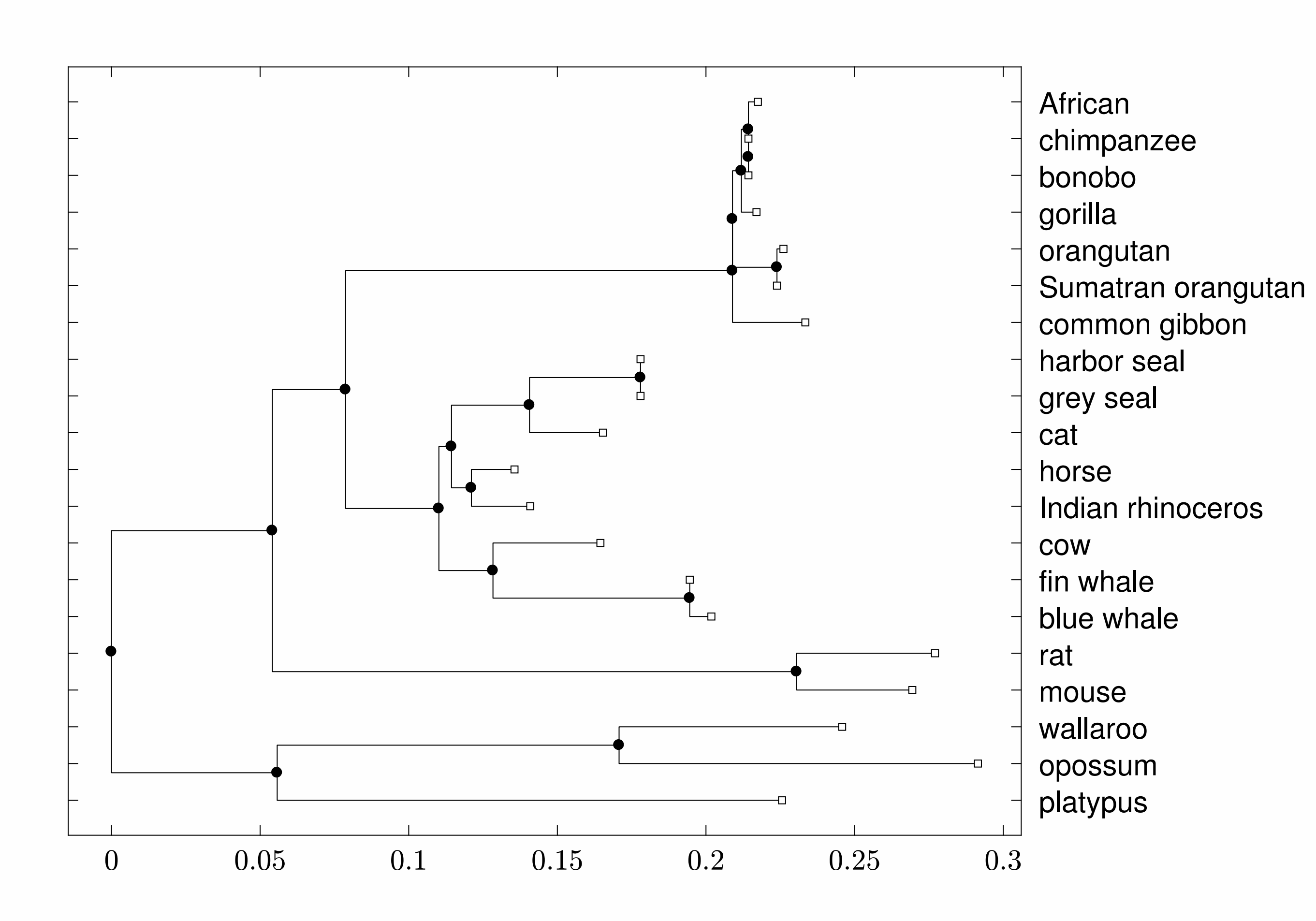

Figure 1: Each branch length in this tree is the differences between the branch length estimated under RaMoSS and the branch length estimated under RaMoSSwDT. Only differences greater than 0.002 in absolute magnitude are shown. The scale on the horizontal axis is the number of single nucleotide substitution per codon. RaMoSS produced larger branch lengths compared to RaMoSSwDT on most branches.

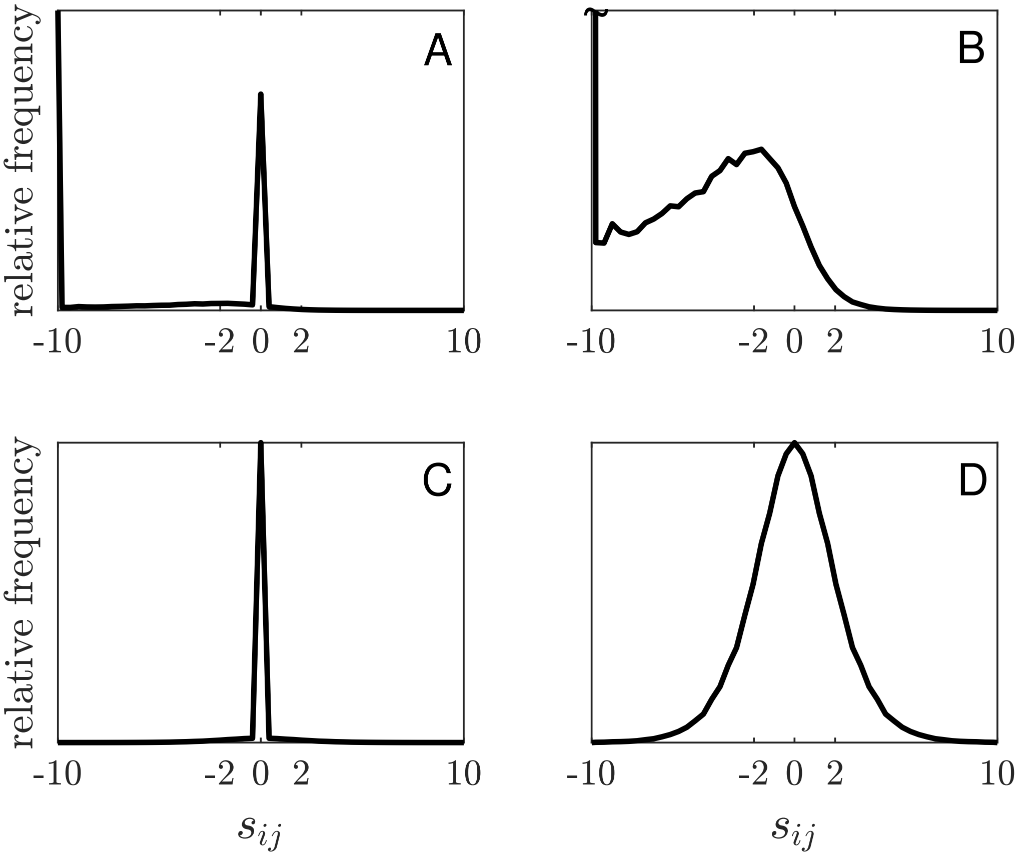

Figure 2: The distributions of scaled selection coefficients resulting from the proposed method of simulating vectors of site-specific fitness coefficients for mammal mtDNA. A: all mutations; B: nonsynonymous mutations; C: all substitutions; and D: nonsynonymous substitutions. These are very similar in shape to empirical distributions obtained by an analysis of 12 mitochondrial genes of 244 placental mammal species (Tamuri *et al.*, 2012).

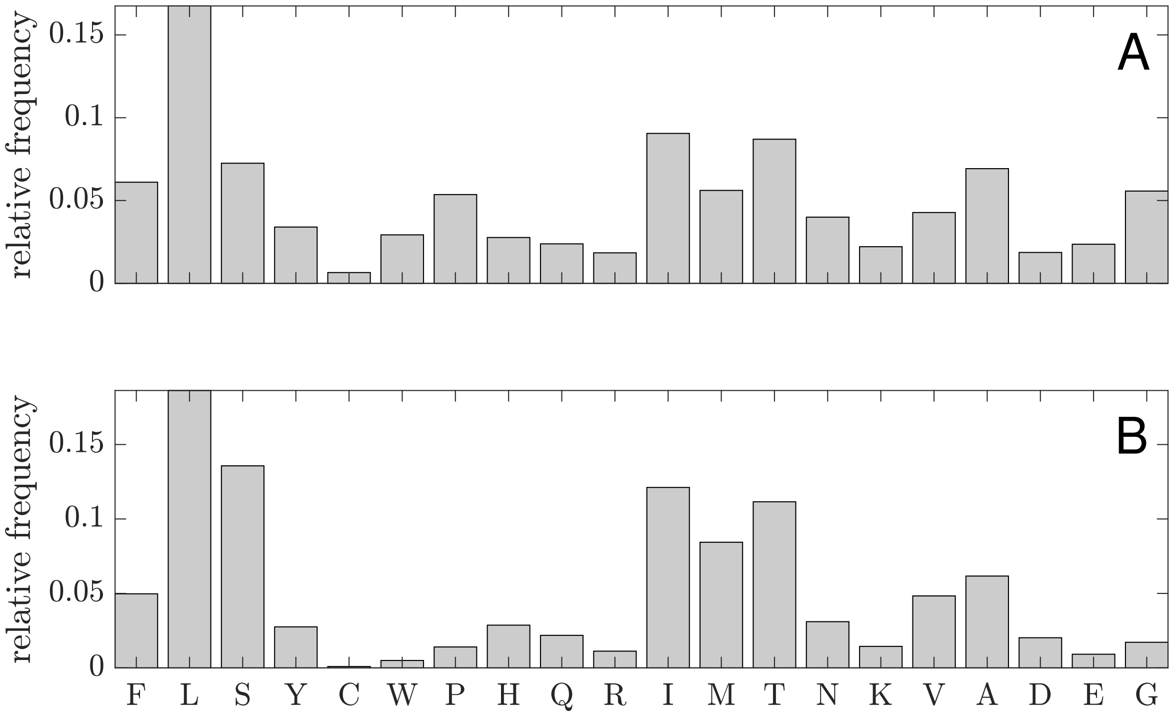

Figure 3: A comparison of the observed versus simulated amino acid frequencies. A: Frequencies obtained from the real data; B: the same for the simulated alignment (20 taxon, 3331 sites).

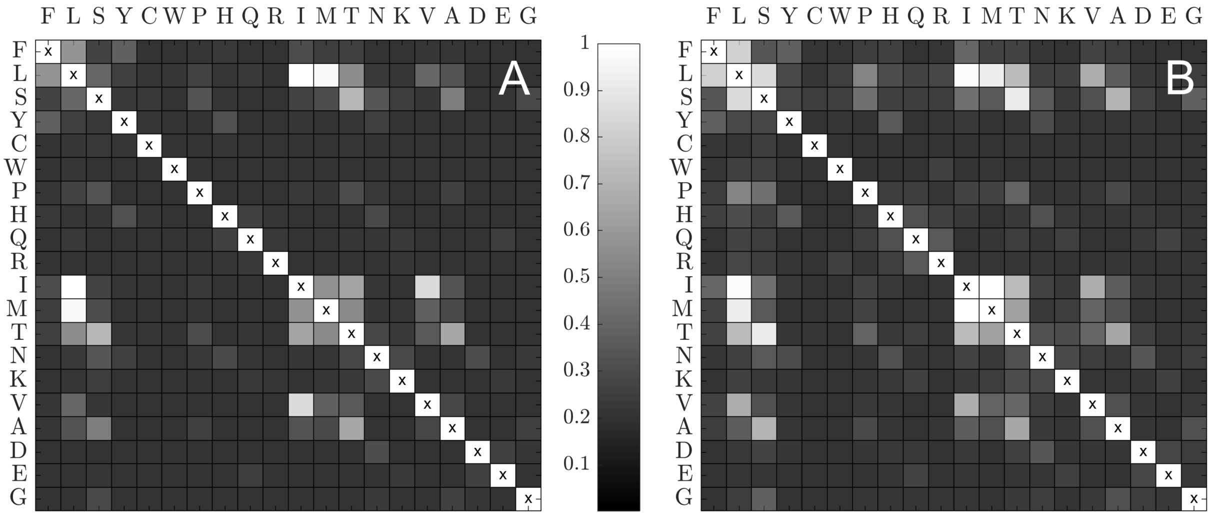

Figure 4: A comparison of the observed versus simulated relative pairwise amino acid frequencies. For any cell, the value is the proportion of sites where the amino acids indicated were both present. A: Values obtained from the real mtDNA alignment; B: values obtained from a simulated alignment (20 taxon, 3331 sites). The same grayscale applies to both panels.

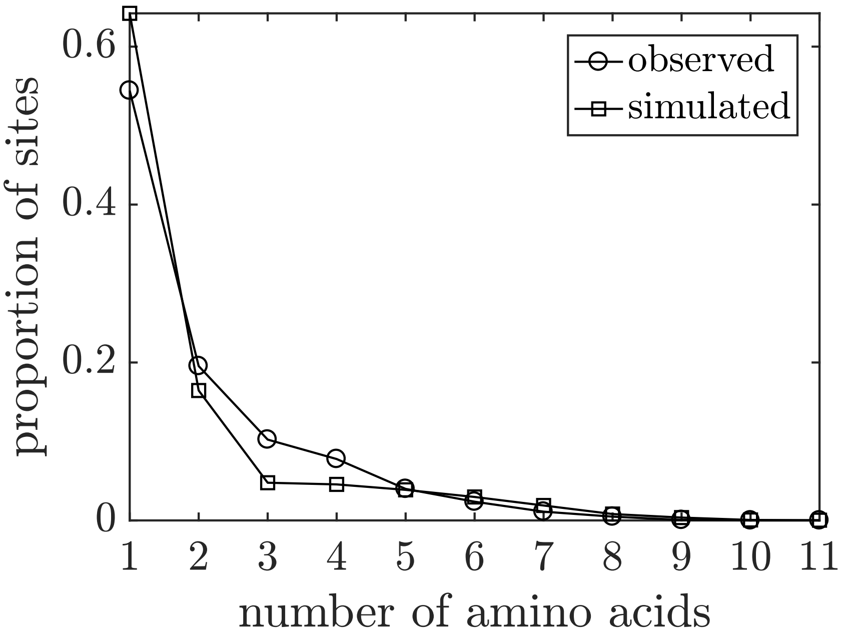

Figure 5: A comparison of the observed versus simulated distribution of the number of amino acids realized at a site for mammal mtDNA (20 taxon, 3331 sites).

Supplementary Tables

Model	LL	rate ratios	proportions	switching	mean S,D,T
M0	-6972	$\hat{\omega}_0 = 0.02$			
M3	-6890	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.08)$	$\hat{p}_0 = 0.84$		
CLM3	-6866	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.21)$	$\hat{p}_0 = 0.93$	$\hat{\delta} = 0.06$	
RaMoSS	-6859	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.03)$	$\hat{p}_{\text{M3}} = 0.72, \hat{p}_0 = 0.86$		
		$(\hat{\omega}'_0, \hat{\omega}'_1) = (0.01, 0.44)$	$\hat{p}'_0 = 0.88$	$\hat{\delta} = 0.17$	
M0wDT	-6960	$\hat{\omega}_0 = 0.01$			90.5%, 7.1%,2.4%
M3wDT	-6888	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.07)$	$\hat{p}_0 = 0.85$		94.1%, 4.6%,1.3%
CLM3wDT	-6866	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.19)$	$\hat{p}_0 = 0.92$	$\hat{\delta} = 0.06$	98.6%, 1.1%, 0.3%
RaMoSSwDT	-6859	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.03)$	$\hat{p}_{\text{M3}} = 0.73, \hat{p}_0 = 0.85$		99.6%, 0.4%, 0.0%
		$(\hat{\omega}'_0, \hat{\omega}'_1) = (0.01, 0.43)$	$\hat{p}'_0 = 0.88$	$\hat{\delta} = 0.18$	

Table 1: Median values for parameter estimates derived from 100 alignments generated under RaMoSS with $\alpha = \beta = 0$.

Model	LL	rate ratios	proportions	switching	mean S,D,T
M0	-10,023	$\hat{\omega}_0 = 0.13$			
M3	-9589	$(\hat{\omega}_0, \hat{\omega}_1) = (0.01, 0.42)$	$\hat{p}_0 = 0.70$		
CLM3	-9585	$(\hat{\omega}_0, \hat{\omega}_1) = (0.01, 0.45)$	$\hat{p}_0 = 0.71$	$\hat{\delta} = 0.02$	
RaMoSS	-9551	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.31)$ $(\hat{\omega}'_0, \hat{\omega}'_1) = (0.05, 0.69)$	$\hat{p}_{\text{M3}} = 0.69, \hat{p}_0 = 0.74$ $\hat{p}'_0 = 0.65$	$\hat{\delta} = 0.00$	
M0wDT	-9985	$\hat{\omega}_0 = 0.11$			88.7%, 4.5%,6.8%
M3wDT	-9588	$(\hat{\omega}_0, \hat{\omega}_1) = (0.01, 0.41)$	$\hat{p}_0 = 0.70$		98.5%, 0.7%, 0.7%
CLM3wDT	-9583	$(\hat{\omega}_0, \hat{\omega}_1) = (0.01, 0.44)$	$\hat{p}_0 = 0.71$	$\hat{\delta} = 0.02$	99.1%, 0.4%, 0.5%
RaMoSSwDT	-9550	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.31)$ $(\hat{\omega}'_0, \hat{\omega}'_1) = (0.05, 0.70)$	$\hat{p}_{\text{M3}} = 0.69, \hat{p}_0 = 0.74$ $\hat{p}'_0 = 0.65$	$\hat{\delta} = 0.00$	99.5%, 0.2%, 0.5%

Table 2: Median values for parameter estimates derived from 100 alignments generated under M3(k=n) with $\alpha = \beta = 0$.

Model	LL	rate ratios	proportions	switching	mean S,D,T
M0	-8056	$\hat{\omega}_0 = 0.05$			
M3	-7713	$(\hat{\omega}_0, \hat{\omega}_1) = (0.01, 0.25)$	$\hat{p}_0 = 0.76$		
CLM3	-7698	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.30)$	$\hat{p}_0 = 0.78$	$\hat{\delta} = 0.05$	
RaMoSS	-7670	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.12)$ $(\hat{\omega}'_0, \hat{\omega}'_1) = (0.00, 0.62)$	$\hat{p}_{\text{M3}} = 0.79, \hat{p}_0 = 0.82$ $\hat{p}'_0 = 0.57$	$\hat{\delta} = 0.22$	
M0wDT	-8018	$\hat{\omega}_0 = 0.04$			80.9%, 13.7%,5.4%
M3wDT	-7702	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.21)$	$\hat{p}_0 = 0.76$		90.3%, 7.6%,2.1%
CLM3wDT	-7691	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.25)$	$\hat{p}_0 = 0.78$	$\hat{\delta} = 0.05$	92.8%, 5.7%, 1.5%
RaMoSSwDT	-7666	$(\hat{\omega}_0, \hat{\omega}_1) = (0.00, 0.12)$ $(\hat{\omega}'_0, \hat{\omega}'_1) = (0.00, 0.56)$	$\hat{p}_{\text{M3}} = 0.80, \hat{p}_0 = 0.82$ $\hat{p}'_0 = 0.60$	$\hat{\delta} = 0.20$	96.3%, 2.9%, 0.8%

Table 3: Median values for parameter estimates derived from 100 alignments generated under MutSel-mmtDNA with $\alpha = \beta = 0$.

Generating Alignments using MutSel-mmtDNA

The most direct way to simulate alignments consistent with real data is to estimate site-specific fitness coefficients from the real data and then use the MutSel framework of Halpern and Bruno (1998) to construct site-specific subsitution rate matrices. The Pyvolve software package (Spielman and Wilke, 2015) includes modules for this purpose. In the course of our study, it was necessary to generate data consistent with the 20-taxon concatenated alignment of H-strand mammalian mitochondrial DNA sequences provided by the PAML software package (Yang, 2007). This alignment is too small for the direct approach. It was therefore necessary to devise a more approximate generating procedure, which is described in this section.

The degree to which an alignment generated under MutSel can be said to be realistic is in large part dependent on how site-specific amino acid fitnesses are constructed. One method is to draw vectors of fitness coefficients from a normal distribution (e.g, Jones *et al.*, 2017). But since fitnesses are random, it is possible to draw a vector that assigns nearly the same fitness to a pair of amino acids with very different physicochemical properties. Hence, a site pattern might easily contain both isoleucine (hydrophobic) and serine (hydrophylic). This is unlikely to occur at a real site evolving under stringent selection without a drastic change in the physicochemcial requirements for that site. It would be more realistic to see isoleucine occur together with the similarly hydrophobic aliphatic amino acids leucine and valine. Furthermore, the stringency of selection, determined by the standard deviation of the normal distribution, must have a realistic level of variance across sites.

Taking these requirements into account, the following method was used to generate a vector of fitness coefficients for a codon site under what we call the MutSel-mmtDNA generating model:

1. A codon for the h^{th} site was randomly drawn using a multinomial distribution with probabilities equal to the empircal codon frequence for the real mtDNA.

2. The amino acid X corresponding to the chosen codon was assigned a provisional fitness of 0.25.

3. A provisonal vector of fitnesses for the remaining amino acids was constructed by dividing $\langle \upsilon_{Y_1}, ..., \upsilon_{Y_{19}} \rangle$ by its largest element, where υ_Y is the number of site patterns in the real mtDNA that included both amino acids X and Y (Figure 4). This gave the amino acid that paired most frequently with X (call it Z) a fitness of one and all other amino acids Y $\neq$ X a fitness less than one.

4. Each element of $\langle \upsilon_{Y_1}, ..., \upsilon_{Y_{19}} \rangle$ was then reduced by a random draw from a half-normal distribution with mean zero and standard deviation one. This was meant to increase the variation in stationary frequencies across sites. The expected value of the half-normal distribution is $\sqrt{2/\pi} \approx 0.80$. The expected fitness of Z was therefore 0.20, slightly less than the fitness of X. Other amino acids tended to have lower fitness.

5. A scaling factor $\sigma^h \sim 0.001 + (0.01 - 0.001) \times B$ was drawn to determine the stringency of selection at the site, where $B \in [0, 1]$ is a beta random variable with shape parameters $u, v > 0$. Values of $\sigma^h \in [0.001, 0.01]$ closer to the upper bound correspond to greater stringency, whereas values closer to the lower bounded correspond to a balance between selection and drift that typically results in heterotachy (Jones *et al.*, 2017). Parameters u and v for the beta distribution were chosen to make the distributions of scaled selection coefficients s_{ij} match those reported by Tamuri *et al.* (2012) as closely as possible, as described below.

6. A vector $\mathbf{f}^h$ of fitness coefficients for the 60 codons was then constructed from the amino acid fitnesses assuming synonymous codons to be equally fit. This vector was scaled to make its standard deviation equal to σ^h.

7. $\mathbf{f}^h$ was then used to construct the site-specific rate matrix A^h as described below.

For a given (u, v), 1000 draws of $\mathbf{f}^h$ were used to approximate the PDFs of the s_{ij} for all mutations, nonsynonymous mutations, all substitution and nonsynonymous substitutions (as detailed in the next section). Probabilities $p(s_{ij} < -2)$, $p(-2 < s_{ij} < 2)$ and $p(s_{ij} > 2)$ were calculated and compared with empirical values reported by Tamuri *et al.* (2012). This process was repeated over a grid of (u, v) coordinate pairs. The coordinate corresponding to the smallest sum of squared differences between simulated and empirical probabilities was found to be $(u, v) = (0.08, 0.02)$. These values give σ^h a U-shaped density function with most of its mass near the upper and lower bounds of its domain $[0.001, 0.01]$.

Site-specific fitness coefficients $\mathbf{f}^h = \langle f_1^h, ..., f_{60}^h \rangle$ for the 60 codons were converted into scaled selection coefficients $s_{ij}^h = 2N_e(f_j^h - f_i^h)$ assuming an effective population size of $N_e = 1000$ and a ploidy of two. These were used to construct a site-specific matrix of fixation probabilities:

$$W_{ij}^h \propto \begin{cases} 1 & \text{if } s_{ij}^h = 0 \\ \frac{s_{ij}^h}{1-\exp\left(-s_{ij}^h\right)} & \text{otherwise} \end{cases} \tag{1}$$

The corresponding site-specific rate matrix A^h was then made to be proportional to the element-wise product of the mutation-rate matrix M, where the mutation rate from codon $i = i_1 i_2 i_3$ to codon $j = j_1 j_2 j_3$ was specified as:

$$M_{ij} \propto \begin{cases} \kappa^{n_t} \Pi_{i_k \neq j_k} \pi_{j_k}^* & \text{if } n = 1 \\ \alpha \kappa^{n_t} \Pi_{i_k \neq j_k} \pi_{j_k}^* & \text{if } n = 2 \\ \beta \kappa^{n_t} \Pi_{i_k \neq j_k} \pi_{j_k}^* & \text{if } n = 3 \end{cases} \tag{2}$$

(see main article for details) and W in (1): $A^h \propto M \circ W$, where M was constructed using model parameters estimated from the real mtDNA alignment. Each A^h has its own vector of stationary frequencies $\boldsymbol{\pi}^h = \langle \pi_1^h, ..., \pi_{60}^h \rangle$ and its own expected rate r^h. All A^h were divided by r, the mean of the r^h, so that branch length could be interpreted as the expected number of single nucleotide substitution per codon.

References

Halpern, A. L. and Bruno, W. J. 1998. Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol. Biol. Evol.*, 15: 910–917.

Jones, C. T., Youssef, N., Susko, E., and Bielawski, J. P. 2017. Shifting balance on a static mutation-selection landscape: a novel scenario of positive selection. *Mol. Biol. Evol.*, 34: 391–407.

Spielman, S. and Wilke, C. O. 2015. Pyvolve: A flexible Python module for simulating sequences along phylogenies. *PLoS ONE*, 10: 1–7.

Tamuri, A. U., dos Reis, M., and Goldstein, R. A. 2012. Estimating the distribution of selection coefficients from phylogenetic data using sitewise mutation-selection models. *Genetics*, 190: 1101–1115.

Yang, Z. H. 2007. PAML4: phylogentic anallysis by maximum likelihood. *Mol. Biol. Evol.*, 24: 1586–1591.