Mitochondrial–Nuclear Interactions and Accelerated Compensatory Evolution: Evidence from the Primate Cytochrome c Oxidase Complex

Naoki Osada*[,1,2] and Hiroshi Akashi[1,2]

[1]Division of Evolutionary Genetics, Department of Population Genetics, National Institute of Genetics, Mishima, Japan

[2]Department of Genetics, The Graduate University for Advanced Studies (SOKENDAI), Mishima, Japan

***Corresponding author:** E-mail: nosada@lab.nig.ac.jp.

Associate editor: John H McDonald

Abstract

Accelerated rates of mitochondrial protein evolution have been proposed to reflect Darwinian coadaptation for efficient energy production for mammalian flight and brain activity. However, several features of mammalian mtDNA (absence of recombination, small effective population size, and high mutation rate) promote genome degradation through the accumulation of weakly deleterious mutations. Here, we present evidence for "compensatory" adaptive substitutions in nuclear DNA- (nDNA) encoded mitochondrial proteins to prevent fitness decline in primate mitochondrial protein complexes. We show that high mutation rate and small effective population size, key features of primate mitochondrial genomes, can accelerate compensatory adaptive evolution in nDNA-encoded genes. We combine phylogenetic information and the 3D structure of the cytochrome c oxidase (COX) complex to test for accelerated compensatory changes among interacting sites. Physical interactions among mtDNA- and nDNA-encoded components are critical in COX evolution; amino acids in close physical proximity in the 3D structure show a strong tendency for correlated evolution among lineages. Only nuclear-encoded components of COX show evidence for positive selection and adaptive nDNA-encoded changes tend to follow mtDNA-encoded amino acid changes at nearby sites in the 3D structure. This bias in the temporal order of substitutions supports compensatory weak selection as a major factor in accelerated primate COX evolution.

Key words: compensatory evolution, adaptive evolution, mitochondria, mutation rate, primates.

Introduction

A fundamental role of mitochondria in energy metabolism, genetic features of mitochondria (maternal inheritance without recombination), and direct physical interactions between mtDNA- and nuclear DNA- (nDNA) encoded proteins provide a rich system for elucidating mechanisms of natural selection. Shen et al. (2010) found evidence for accelerated protein evolution for both mtDNA- and nDNA-encoded components of respiratory enzyme complexes in the ancestral lineage leading to bats. Evidence for adaptive evolution is biased toward genes directly involved in energy production suggesting that Darwinian evolution of these genes played a critical role in the origin of mammalian flight. Brain activity is also energetically costly (Aiello and Wheeler 1995) and accelerated evolution of both mtDNA- and nDNA-encoded mitochondrial proteins may reflect adaptive evolution for enhanced brain function in the lineage leading to humans (Goldberg et al. 2003; Grossman et al. 2004).

Although adaptive evolution of energy metabolism in mammals is consistent with lineage-specific accelerations of protein change, it is important to note that several features of mitochondrial genomes promote nonadaptive modes of evolution. Lack of recombination and a high mutation rate in mammals (Brown et al. 1979; Lynch et al. 2006) appear to have accelerated joint fixations of mutations that preserve Watson–Crick pairing in mitochondrial transfer RNA stem regions (Meer et al. 2010). Limited recombination can also promote fixations of slightly deleterious mutations (Lynch 2010; Weissman et al. 2010). Lynch (1996) proposed that mitochondrial genome degradation caused reduced thermal stabilities for mitochondrial tRNAs relative to their nuclear counterparts.

Figure 1 depicts three contrasting scenarios of evolution for mutations that interact in their effects on fitness. Figure 1*A* shows Darwinian adaptive evolution through stepwise fixations of mutations that increase organismal fitness. Although their fitness effects may not be strictly additive, each mutation confers a fitness benefit. Figure 1*B* shows a "compensatory neutral" (CN) scenario studied by Kimura (1985). Here, two mutations that are individually deleterious show little or no fitness effect in combination. Such sets of mutations can go to fixation jointly albeit at a much slower rate than for individual neutral mutations. The rate of CN evolution increases under high mutation rate and tight linkage between loci (Kimura 1985; Iizuka and Takefu 1996; Innan and Stephan 2001; Meer et al. 2010). This mode of evolution has been proposed to explain evolutionary patterns for both nDNA- and mtDNA-encoded RNA secondary structures (Chen and Stephan 2003; Meer et al. 2010).

Figure 1C shows a scenario of evolution under compensatory weak selection (CWS), proposed by Ohta (1973). In

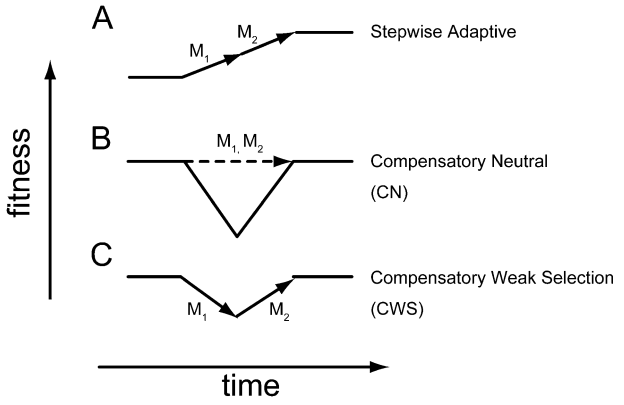

FIG. 1. Schematic models of multistep molecular evolution. M_1 and M_2 denote two successive mutations. (*A*) Stepwise adaptive model. Each adaptive mutation increases the fitness of organisms. (*B*) The CN model. Two compensatory mutations occur in the same haplotype and are simultaneously fixed in the population. (*C*) The CWS model. Fixation of a deleterious mutation is followed by a compensatory advantageous fixation.

this case, fixations of two mutations do not occur simultaneously; fixation of a slightly deleterious mutation is followed by an adaptive substitution that restores fitness. This scenario of slightly deleterious and compensatory adaptive evolution requires deleterious mutations of small effect because fixation probabilities become very small under even moderate selection (i.e., the "fitness valley" for strongly deleterious mutations prevents fixation without a linked compensatory change; Hartl and Taubes 1996; Innan and Stephan 2001). Biological function and organismal fitness are maintained at a roughly constant level under CWS as well as CN, but adaptive evolution (positive selection for compensatory mutations) is an important component of conservative evolution under weak selection. CWS has been proposed to account for global patterns of evolution at synonymous sites (reviewed in Hershberg and Petrov (2008)) as well as within noncoding regions (Ludwig 2002; Kenigsberg et al. 2010; Lawrie et al. 2011).

The cytochrome c oxidase (COX) complex is well suited for studying the evolution of within- and between-genome interactions. COX plays a critical role in the electron transport pathway in the mitochondrial membrane. The complex includes three mtDNA-encoded and ten nDNA-encoded subunits in mammals. Most importantly, the 3D structure of the COX complex shows a large number of potentially interacting residues between mtDNA- and nDNA-encoded subunits (Tsukihara et al. 1995). Previous studies have revealed three intriguing relationships between protein evolution and the 3D structures of mitochondrial complexes. First, mtDNA-encoded amino acids in close proximity in the COX1 structure show correlated substitution patterns among vertebrates (Wang and Pollock 2007). In addition, adaptive amino acid changes in primate COX complexes appeared to be clustered in regions of contact between protein subunits (Wu et al. 1997; Grossman et al. 2001; Schmidt et al. 2001; Uddin et al. 2008). Finally, several mutations in the *COX1* and *COX3* genes that are pathogenic in humans have been observed in naturally occurring genomes from nonhuman mammals. The deleterious effects of these mutations appear to be suppressed by amino acid substitutions at other sites in close proximity in the protein structures (Azevedo et al. 2009).

Rates of mammalian COX protein evolution are highly heterogeneous (Adkins et al. 1996; Wu et al. 2000; Grossman et al. 2001; Schmidt et al. 2001; Osada et al. 2002; Goldberg et al. 2003; Doan et al. 2004; Uddin et al. 2008). Brain-expressed genes generally evolve more slowly than other genes among mammals (Kuma et al. 1995; Duret and Mouchiroud 2000), but brain-expressed genes involved in electron transport have evolved more rapidly than the genome-wide average between human and macaque (Wang et al. 2007). Interestingly, accelerated COX evolution has been identified mainly in primate lineages. This pattern is consistent with coadaptation between nuclear and mitochondrial genomes to meet the aerobic demands of elevated brain activity in primate lineages, the "brain-energy" hypothesis (Grossman et al. 2004; Uddin et al. 2008). The brain-energy hypothesis is a form of Darwinian, stepwise adaptive evolution; amino acid changes in components of the COX complex that enhance metabolism confer a fitness advantage by allowing greater neural activity and were favored in the ancestors of primates.

The brain-energy hypothesis is a plausible and appealing explanation for accelerated COX evolution but alternative explanations remain relatively unexplored. The combination of high mutation rates and small effective population sizes, key features of many animal mitochondrial genomes (Brown et al. 1979), should accelerate fixations of slightly deleterious mutations (Lynch 1996). Elevated rates of protein evolution at nDNA-encoded COX subunit genes may reflect compensatory adaptive fixations that restore mitochondrial function. Over time, CWS can lead to genetic incompatibility among independently evolving lineages and this mode of evolution could explain significant reductions in energy production when mtDNA- and nDNA-encoded proteins from divergent species function in the same mitochondria (Kenyon and Moraes 1997).

Here, we present evidence for a role of CWS in the accelerated evolution of the primate COX complex. We first show that mammals, especially primates, have experienced especially high mitochondrial synonymous substitution rates. Tests of codon- and lineage-specific protein rate accelerations reveal evidence for adaptive protein evolution confined to nDNA-encoded subunits of primate COX. Theoretical analyses and computer simulations show that accelerated compensatory evolution is expected in nuclear genomes under high mitochondrial mutation rates to slightly deleterious alleles. We find that, among sites that show evidence for adaptive evolution, amino acid fixations at nDNA-encoded sites tend to be preceded by substitutions at mtDNA-encoded sites nearby in the COX 3D structure. This biased order of substitutions supports that natural selection to maintain mitochondrial function in the face of strong mutation pressure contributed to the rapid evolution of primate COX genes.

Materials and Methods

Estimation of Mitochondrial/Nuclear Mutation Rates

Sources of nucleotide sequences are shown in the supplementary table 1, Supplementary Material online. Although two macaque species, *Macaca mulatta* and *M. fascicularis*, met the criteria for choosing species pairs for the analysis, they were filtered because they appear to share a large number of polymorphisms only in the nuclear genomes (Osada et al. 2010). For the cow–bison comparison, we combined the data of *Bison bonasus* and *B. bison* because they are equally diverged from *Bos taurus*. Pairwise alignments were performed using CLUSTALW (Thompson et al. 1994) with default parameters. Synonymous divergence was estimated using the method of Nei and Gojobori (1986) and the maximum likelihood (ML) method (F3x4 model) implemented in CODEML (Yang 2007). Nuclear and mitochondrial synonymous divergences were estimated for concatenated data for all genes in each genome. The same method was applied to 1,000 resampled data sets to estimate bootstrap mean and confidence interval (CI) of the ratios of mtDNA to nDNA substitution rates (α). The unit for resampling was codons within the concatenated data.

Sequence Data of COX

The crystal structure of the COX complex determined from bovine proteins (PDB ID: 1OCC) was used for the analysis. For each pair of amino acid residues, the distance between the surfaces of two amino acid residues was measured using the structure information. Because the COX complex forms a homodimer, a given pair of amino acid residues has two physical distances in the 3D structure. The shorter distance was employed in this analysis. mtDNA-encoded *COX* sequences of orangutans, macaques, marmosets, and mice were downloaded from the public database and orthologous sequences of the nDNA-encoded *COX* subunits were extracted from their genome sequences using the University of California–San Cruz genome browser (supplementary table 1, Supplementary Material online). In *COX4I1* and *COX5B*, some N-terminal amino acid sequences were not available in the marmoset genome. Such regions were excluded from the analyses.

Detecting Sites Under Positive Selection

To infer sites under positive selection, the branch-site test of positive selection was performed (Zhang et al. 2005), setting the primate lineages (thick lines in fig. 3) as foreground lineages and the other branches leading to cows and mice as background lineages. In CODEML program, ML values between under Model 2a (model = 2, NSites = 2, and foreground $\omega > 1$) and Model 2b (model = 2 and NSites = 2, and foreground $\omega = 1$) were compared using the likelihood ratio test. Posterior probabilities by the Bayes empirical Bayes method were used to infer sites under positive selection.

Fixation Probability of Compensatory Changes

Consider two interacting loci, *A* and *B*, from mitochondrial and nuclear genomes, respectively. Let μ represent the nuclear mutation rate from *B* to *b* and $\alpha\mu$ represent the mitochondrial mutation rate from *A* to *a*. For simplicity, we do not consider back mutations. Effective population sizes are $2N_e$ for the nuclear locus, *B*, and $\frac{1}{2}N_e$ for the mitochondrial locus, *A*. Under CWS, wild-type fitness is maintained for genotypes *ABB* and *abb*, but *ABb* and *aBb*, and *Abb* and *aBB* genotypes suffer fitness loss (i.e., incompatibility). Fitness values for the three classes are 1, $1-hs$, and $1-s$, respectively, where *h* is a dominance parameter.

We derived general results assuming no cosegregation of deleterious and compensatory alleles. In the case of semidominance ($h = 1/2$), the rate of molecular evolution at nuclear loci (K_n) with selection coefficient *s* is given by the equation of (Kimura 1964),

$$K_n = \frac{2S}{1 - e^{-2S}}\mu, \quad (1)$$

where $S = N_e s$.

The rate of molecular evolution at mitochondrial loci (K_m) is given by,

$$K_m = \frac{S}{1 - e^{-S}}\alpha\mu. \quad (2)$$

When $h \neq 1/2$, the expected rate of molecular evolution is given in previous studies, for example, Kimura (1964). In mammals, mutation rates are higher in mtDNA than in nDNA, that is, $\alpha >> 1$. In addition, the efficacy of selection is weaker in mtDNA than in nDNA owing to the small effective size (small N_e hence small $N_e s$). Therefore, we assume that fixations of deleterious mutations in locus A are followed by compensatory adaptive substitutions at locus *B*. A key unknown parameter is the fraction of mutations that can compensate for deleterious fixations. Here, we assume that the rate of adaptive compensatory mutations equals the rate of spontaneous mutations (μ). This is unrealistic but allows us to make conclusions about the relative effects of selection coefficients and dominance on compensatory dynamics. In the scenario described above, the average waiting time for the fixation of compensatory mutations ($1/K_c$) is the sum of the waiting time for the fixation of mutations in mtDNA ($1/K_m$) and nDNA ($1/K_n$),

$$\frac{1}{K_c} = \frac{1}{K_m} + \frac{1}{K_n}. \quad (3)$$

Computer simulations assumed the Wright–Fisher model. In each generation, mutations were assigned with a probability of $\alpha\mu$ for a mitochondrial locus and μ for a nuclear locus, assuming the Poisson distribution. We set $N_e = 10{,}000$ and $\mu = 10^{-8}$, which are close to the estimated values in primate genomes. The number of gametes passed to the next generation was drawn from multinomial distributions, considering the relative fitness of each genotype. We counted the number of fixed alleles in the nuclear loci until

the sum of time exceeded 10^{12} generations. Three mtDNA/nDNA mutation ratios, $\alpha = 10$, 20, and 30, were used for both semidominant and dominant cases. Back mutation at the mitochondrial locus was also considered but is unlikely to have a large effect on the results because the rate of back mutation will be a small fraction of the rate to deleterious mutations. For example, if the back mutation rate is 1/5 of the deleterious rate, the expected nuclear gene substitution rate remains 2-fold faster the mutation rate in the nuclear locus (for $S \approx 4$ and $\alpha = 30$).

Evolutionary Interactions Between Subunits

We inferred amino acid substitutions in the primate COX complex using the ML method implemented in the PAML software package (Yang 2007). CODEML estimates the posterior probabilities of codon reconstructions at ancestral nodes in the tree. We used these reconstructions to infer branch-specific counts of amino acid substitutions at each codon. This allows us to incorporate uncertainty in ancestral reconstructions in our inferences of branch-specific counts of amino acid substitutions (details of the method are given in supplementary figure 1, Supplementary Material online and its legend). Only codons having a total of more than 0.05 substitutions among primate lineages (thick lines in fig. 3) were included in the analyses. CODEML generally assigned high probabilities to reconstructed ancestral states for these data. The sum of probabilities for inferred amino acid fixations (across sites and branches) were 409 and 114 for mtDNA and nDNA, respectively. Fixations with probabilities greater than 0.8 make up 60 and 82% of these totals, respectively.

For pairs of variable codons that encode sites in different subunits, the product of amino acid fixation counts for the two codons on a given branch of the gene tree was used as the "same-branch" count. Such counts were summed across primate lineages to give the total counts for each codon pair. We also estimated counts for "sequential" substitutions. For nuclear substitutions preceded by mitochondrial substitutions, we first identified codons that experienced amino acid changes in nDNA-encoded subunits in lineages downstream of the branches that connect to the ancestral primate node (asterisk in fig. 3). For each of these codons, we searched for mtDNA-encoded codons that experienced substitutions in branches closer to the ancestral primate node. Between-branch counts were calculated for each pair of codons as the product of the amino acid fixation counts for the codons on the relevant branches. Counts for mitochondrial substitutions preceded by nuclear substitutions were estimated similarly.

Codon pairs were divided into two classes according to their physical distances, <10 and >10 Å. Mann–Whitney U (MWU) statistics were calculated between distance classes for same-branch substitution counts and for sequential substitution bias. For statistical analysis, we randomized the codon position of variable sites and recalculated the MWU statistics. The process was repeated for 10,000 iterations to obtain null distributions. For same-branch tests, substitution counts were scaled to control for evolutionary rate differences among codons (i.e., at each codon, lineage-specific counts in the randomized data were scaled to maintain the total counts in the original data). The sequential substitution test does not require rescaling because we employ the difference in counts of substitutions in two directions.

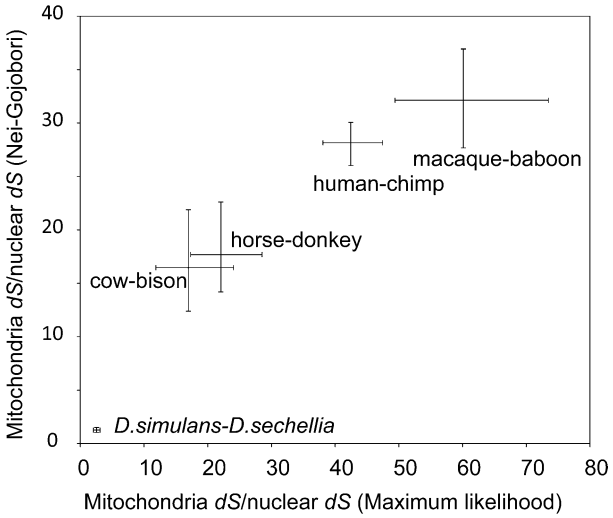

FIG. 2. Mitochondrial dS/nuclear dS between closely related species. The ratios were estimated by the Nei–Gojobori method (ordinate) and the ML method (abscissa). The ratios of human–chimpanzee, macaque–baboon, cow–bison, horse–donkey, and *D. simulans–D. sechellia* were calculated for concatenated sequences for each genome. Error bars represent 95% CI by bootstrap resampling of codons within each concatenated sequence.

Results

Ratio of Mitochondrial to Nuclear Synonymous Substitution Rates

We estimated rates of synonymous substitutions in mtDNA and nDNA of several mammalian taxa. We restricted the analyses to closely related species to avoid saturation and to reduce model-dependence in divergence estimation. Pair-wise synonymous divergence was estimated between *Homo–Pan* (human–chimpanzee), between Cercopithecinae (macaque–baboon), between *Equus* (horse–donkey), and between *Bovinae* (cow–bison). In addition to the mammals, *Drosophila* (*D. simulans–D. sechellia*) was also analyzed. These species pairs have completely sequenced mtDNA sequences from both members, a sequenced nuclear genome from at least one member, at least ten complete nuclear gene sequences from the other member available in the public databases, and Nei–Gojobori synonymous divergence (Nei and Gojobori 1986) in mitochondrial genes smaller than one. Average Nei–Gojobori synonymous divergence values in nDNA were 0.014, 0.021, 0.021, 0.015, and 0.044 between *Homo–Pan*, Cercopithecinae, *Equus*, *Bovinae*, and *Drosophila*, respectively.

Figure 2 shows large differences in ratios of mtDNA to nDNA substitution rates (α) among taxa; primates show

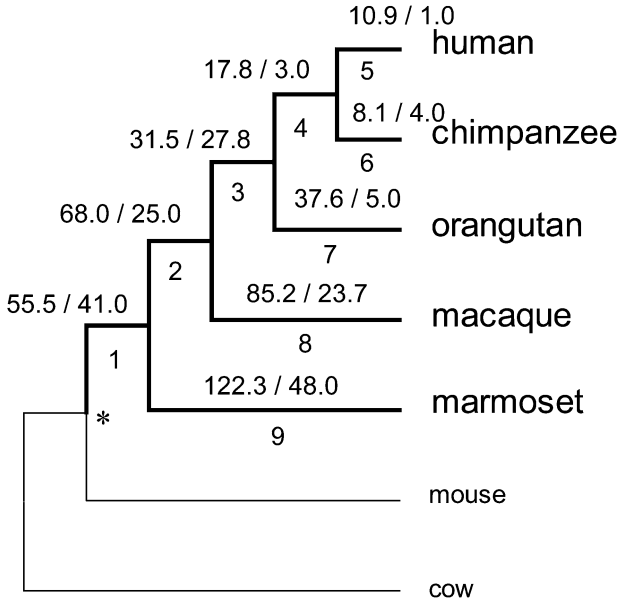

FIG. 3. Patterns of amino acid substitutions in COX subunits. Phylogenetic tree of five primates and two out-group species. The numbers of amino acid substitutions in mtDNA- (left) and nDNA-encoded (right) subunits are shown on the branches. The asterisk shows the ancestral node of primates. The numbers shown under the branches are labels used in supplementarytable 2, Supplementary Material online.

the highest α, followed by other mammals and Drosophila. The ratio in the human–chimpanzee comparisons was 28.2 by the Nei–Gojobori method and 42.5 by the ML method (Yang 2007). The ratio in the macaque–baboon comparisons was also high: 32.1 by the Nei–Gojobori method and 60.1 by the ML method. In contrast, synonymous divergence in mtDNA is only 1.2–2.5 times higher than in nuclear genes for *Drosophila*. *Equus* and *Bovinae* had α values of 15–20, which is also higher than *Drosophila*, but roughly 2-fold lower than in primates. Relatively high α values in mammals, especially primates, are likely to reflect mutation rate differences between mtDNA and nDNA because the impact of selection at synonymous sites appear to be weak in mammals (Duret et al. 2002). Selection may reduce synonymous divergence at nuclear genes in *Drosophila* but α remains similar for nuclear genes showing low codon bias. For genes in the upper 50% of effective number of codon values (Wright 1990), $\alpha = 1.13$ with 95% CI 0.93–1.33, Nei–Gojobori; $\alpha = 2.38$ with 95% CI 1.88–2.90, ML.

Inference of Fast Evolving Sites in COX Genes

We first tested for adaptive protein evolution in *COX* complex genes among primate lineages. COX gene sequences from five primate species (humans, chimpanzees, orangutans, macaques, and marmosets) and two out-groups (mouse and bovine) were employed in the analyses. Orthologous sequences for three mitochondrial encoded *COX* subunits (*COX1*, *COX2*, and *COX3*) and seven nuclear encoded subunits (*COX4I1*, *COX5A*, *COX5B*, *COX6A2*, *COX6C*, *COX7A1*, and *COX7B*) were available for this set of species. A total of 1,001 and 591 codons from mtDNA- and nDNA-encoded *COX* genes were included in the analyses, respectively. The phylogenetic tree for these data and the number of substitutions in mtDNA- and nDNA-encoded subunits are shown in figure 3.

Table 1. Candidates of sites under natural selection with $\omega > 1$.

Gene	−2LnL (Model 2a; $\omega > 1$ vs. $\omega = 1$)	ω[a]	Selected sites ($\omega > 1$)[b]
COX5A	7.269	6.39	5Q (0.996), 38T (0.973), 70V (0.946)
COX6A2	7.436	3.85	36Y (0.948), 43P (0.903), 48R (0.996), 51Q (0901), 73S (0.980)
COX6C	7.922	4.72	23V (0.991), 42Q (0.951), 55D (0.904), 56V (0.989), 57M (0.948), 70Q (0.903)

[a] ω values for adaptive sites.
[b] Posterior probabilities of $\omega > 1$ are shown in the parentheses. Sequence positions and amino acid residues are based on the bovine protein structure (PDB ID: 1OCC).

COX sequences were analyzed using the ML framework implemented in PAML (Yang 2007). The branch-site test of positive selection allows detection of codon-specific adaptive protein evolution at some codons (ratio of nonsynonymous to synonymous substitution (ω) greater than one) on prespecified lineages (Zhang et al. 2005). We labeled primate lineages (thick lines in fig. 3) as "foreground" and out-group lineages as "background". In the three mtDNA-encoded COX genes, branch-site tests did not detect evidence for codons undergoing adaptive amino acid changes. However, three of the seven nDNA-encoded *COX* genes (*COX5A*, *COX6A2*, and *COX6C*) showed evidence for sites under positive selection in primates ($P = 0.035$, 0.0032, and 0.0025, respectively). Among these genes, seven sites showed a signature of positive selection with a posterior probability greater than 0.95 and seven more sites had posterior probabilities between 0.90 and 0.95 (table 1 and supplementary fig. 2, Supplementary Material online).

Theoretical Predictions

The COX complex is composed of interacting sites encoded by genomes that experience different rates of mutation and recombination as well as effective population sizes. We derived theoretical expectation of rates of compensatory evolution for such a scenario (see Materials and Methods) and conducted computer simulations that allow simultaneous segregation of epistatic mutations and adaptive fixations in mtDNA (these factors were not considered in the theoretical formulation above). Simulations were conducted under a standard Wright–Fisher model, using parameter values close to those estimated in primate genomes ($N_e = 10{,}000$ and $\mu = 10^{-8}$). Theoretical predictions and simulation results are concordant for both semidominant (fig. 4A) and dominant (fig. 4B) cases. When $\alpha = 30$, the rate of

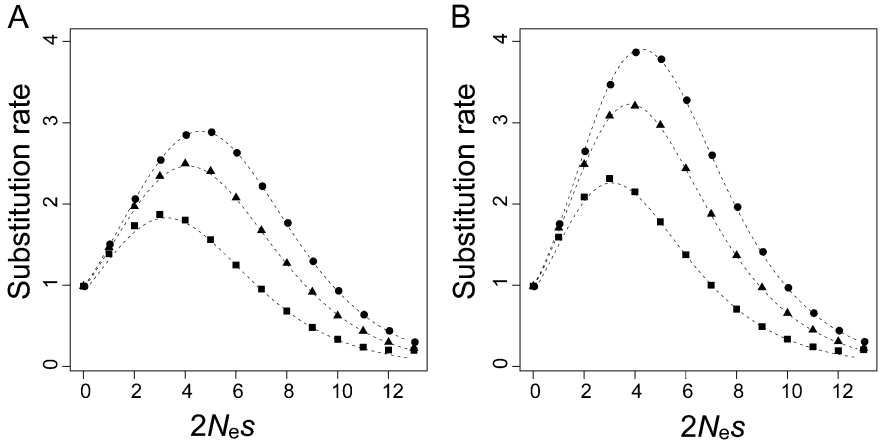

FIG. 4. Predicted rates of compensatory evolution. Rates of evolution at nuclear genes are shown as a function of the fitness effects of compensatory changes. Results for semidominant mutations ($h = ½$) are shown in panel A and dominant mutations ($h = 0$) in panel B. Dashed lines show theoretical expectations. The squares, triangles, and circles show computer simulation results for $\alpha = 10$, 20, and 30, respectively. The y axis is in units of the mutation rate, $u = 10^{-8}$.

molecular evolution at the nuclear locus is nearly three and four times higher than the mutation rate for semidominant and dominant advantageous alleles, respectively. The substitution rate can exceed the mutation rate when the fitness valley is shallow, that is, $2N_e s \approx 5$. These findings are similar to predictions for weak selection models of codon usage evolution (Li 1987) under strongly biased forward and backward mutation rates between translationally preferred and unpreferred states at a single site.

Evolutionary Interactions Between Subunits

Knowledge of the COX 3D structure allowed us to study coevolution among sites from different subunits that are likely to physically interact. We inferred lineage-specific amino acid changes using reconstructions of ancestral codon states. Uncertainty in ancestral sequence reconstruction was taken into account in estimating lineage-specific nonsynonymous divergence. Probabilities for ancestral codons were estimated at each node in the phylogeny using PAML (see Materials and Methods). We examined all pairs of variable sites (codons) from different COX subunits. For each pair, we obtained counts of nonsynonymous substitutions that occurred at both codons in the same branch of the primate phylogeny (see Materials and Methods for details). We also estimated the physical distance between the amino acid residues in the COX 3D structure. The upper limit for electrostatic interactions between two amino acid residues is approximately 10 Å (Elcock et al. 1999; Johnson and Parson 2002). We classified pairs of sites as either potentially interacting (PI, $<$10 Å) or distant (non-PI, $>$10 Å) pairs.

We employed a modified MWU test to compare the distributions of counts for same-branch amino acid changes at PI pairs and non-PI pairs. Because measures among pairs of codons are nonindependent (i.e., the same codon can contribute to a large number of pairs in the data), the null distribution of the U statistic was obtained by randomizing codon positions among variable sites. In addition, we controlled for position-specific substitution patterns by scaling permuted data to tree lengths at each codon (see Materials and Methods). Figure 5A shows the observed MWU statistics and the null distributions from 10,000 randomized data sets. Same-branch amino acid changes are significantly elevated at PI pairs where the two codons are encoded by mtDNA and nDNA (MWU statistics = 5.32, $P = 0.0004$, two-tailed test) but not for PI pairs where both codons are encoded in the same genome (nDNA–nDNA: MWU statistics = 0.54, $P = 0.39$, two-tailed test; mtDNA–mtDNA: MWU statistics = −0.71, $P = 0.85$, two-tailed test). Results were similar for analyses restricted to amino acid substitutions that were assigned high probabilities ($\geq$0.8) by CODEML ($P < 0.001$, mtDNA–nDNA; $P = 0.46$, mtDNA–mtDNA; $P > 0.99$, nDNA–nDNA).

Coevolution among interacting sites is consistent with epistatic selection but does not distinguish between adaptive and compensatory models. We test a further prediction of directionality in protein evolution that is specific to CWS. Under CWS, slightly deleterious fixations are followed by compensatory adaptive changes. Because selection coefficients are small, the time lag between a deleterious fixation and a compensatory substitution can be considerable. The predicted excess of deleterious fixations at mtDNA-encoded sites and compensatory changes at nDNA-encoded sites of the COX complex allows a test of the directionality of amino acid substitutions; compensatory changes in the nDNA should often follow changes in the mtDNA. Adaptive scenarios, including the brain-energy hypothesis, do not predict such a directional bias.

Following a deleterious fixation, compensatory substitutions may occur within the same lineage or after splitting events in the phylogeny. To test for directional biases, we examined substitutions that occur on different branches of the phylogeny (after the split from the common ancestors of primates and rodents). For each pair of variable codons, we obtained counts of nDNA substitutions in branches downstream of mtDNA substitutions in the gene tree. We also obtained counts for changes in the reverse direction:

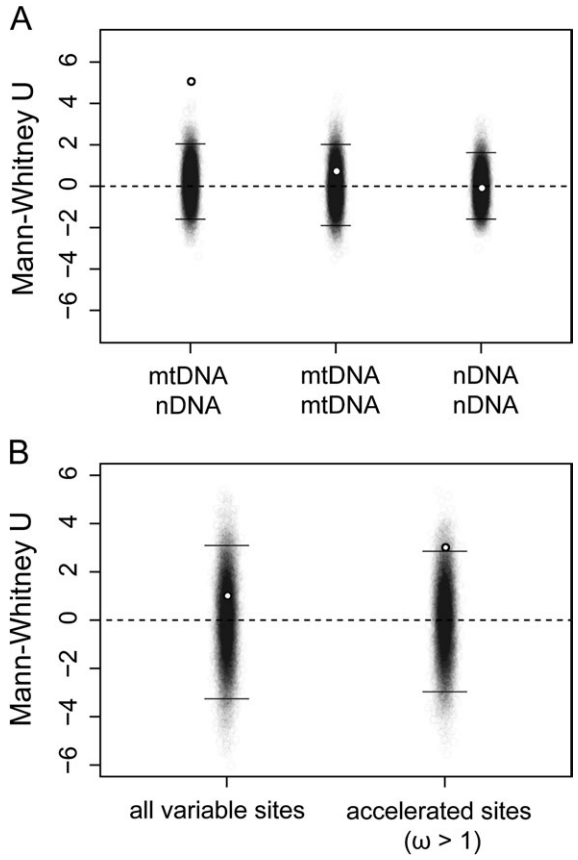

Fig. 5. Difference between same-branch substitutions for PI and non-PI pairs (*A*). Positive *U* values (*y* axis) indicate higher same-branch substitutions for PI pairs. The observed values are shown in red ($P < 0.05$) and white ($P > 0.05$) circles. Blue transparent circles show the null distribution of the statistic from 10,000 iterations. Random noise was added to the *x* axis values from the null distribution to allow visualization of the density of points. The horizontal red lines represent 95% CI of the null distributions. Values between mtDNA- and nDNA-encoded subunits, between mtDNA-encoded subunits, and between nDNA-encoded subunits are plotted. Differences in sequential amino acid changes between PI and non-PI pairs (*B*). Positive values on *y* axis indicate an excess of nDNA substitutions following mtDNA substitutions compared with the substitutions in the opposite direction in PI pairs. The data labels are the same as those in the panel B. The numbers for all variable sites and accelerated sites-only (posterior probability of $\omega > 1$ is >0.90) were plotted.

mtDNA substitutions in branches downstream of nDNA substitutions. The difference between these counts was employed as a measure of bias in the direction of substitutions. The difference values for PI and non-PI pairs were contrasted using the modified MWU test. Among all sites, we did not observe a significant difference in the directionality of substitutions between PI and non-PI pairs ($P = 0.54$; fig. 5*B*). However, among the 14 nDNA-encoded sites, which showed evidence for positive selection (PAML posterior probability of positive selection >0.90), PI pairs are biased toward nDNA substitutions preceded by mtDNA substitutions compared with non-PI pairs ($P = 0.018$; fig. 5*B*). High probability (≥ 0.8) amino acid substitutions show a similar, somewhat stronger association, $P = 0.0038$.

In the analyses above, pairs of sites were classified as potentially interacting based on a cutoff distance of 10 Å. Some experimental and computational studies suggest longer range compensatory changes (Pollock et al. 1999; Yeang and Haussler 2007; Lunzer et al. 2010). Our findings are similar using a 20 Å distance cutoff: same-branch tests ($P = 0.014$, mtDNA–nDNA; $P = 0.21$, mtDNA–mtDNA; $P = 0.58$, nDNA–nDNA), sequential tests ($P = 0.016$, positively selected sites).

Discussion

Our analyses support compensatory evolution between nDNA- and mtDNA-encoded subunits of COX, a mitochondrial protein complex. Rapid evolution in primate *COX* genes appears to be driven, at least in part, by adaptive evolution at nDNA-encoded mitochondrial proteins to counteract slightly deleterious mtDNA substitutions. Among great apes, excess rare amino acid polymorphisms in mtDNA-encoded genes support that a considerable fraction of protein changes have slightly deleterious fitness effects (Hasegawa et al. 1998; Nachman et al. 1998; Green et al. 2008). The lack of recombination and small effective population size of the mitochondrial genome allows a larger range of deleterious mutations to go to fixation with appreciable probabilities relative to the nuclear genome. Lynch (1996) has argued that deleterious fixations have resulted in destabilized and functionally compromised tRNAs encoded in mammalian mtDNA. A dramatically elevated mutation rate in primate mtDNA will exacerbate fitness decline through slightly deleterious fixations.

CWS may be important in protein evolution (Ohta 1973; Depristo et al. 2005), but this model has proven difficult to test. Dean and coworkers (Lunzer et al. 2010) combined biochemical analyses and phylogenetic inference to argue that successive amino acid changes reduced and then restored performance in the recent history of a bacterial enzyme. We attempt to test for recurring CWS in a system that should favor this mode of evolution. Our theoretical model assumes that protein changes in nDNA-encoded components of mitochondrial complexes can compensate for fitness loss in the mitochondrial genome. We find that rates of protein evolution in nDNA-encoded genes can be elevated considerably when weakly deleterious mutations are common and when mutation rates are high in mtDNA-encoded genes. This combination of factors in primate mitochondrial genomes should accelerate compensatory protein evolution in nDNA-encoded mitochondrial genes.

Both Darwinian evolution (possibly to enhance brain function) and compensatory fixations (CN and CWS) are consistent with elevated rates of protein evolution in nDNA-encoded mitochondrial proteins in primates. We attempt to distinguish between these scenarios by localizing amino acid changes both within the 3D structure of the COX complex and within the primate phylogenetic tree. We observed an excess of same-branch changes at potentially interacting sites as predicted by CWS, but it could be argued that such sites are also critical in COX kinetics

and are thus likely targets of selection for enhanced metabolism. We tested a further, and more specific, prediction of compensatory evolution: fitness-restoring amino acid substitutions in nDNA-encoded COX genes should tend to follow deleterious fixations in mtDNA-encoded genes. Some such changes will occur in the same branch of the gene tree, but the dense sampling within primates allows us to test the temporal order of changes at candidate sites for CWS.

Among 14 amino acid positions of nDNA-encoded COX components that are strong candidates for adaptive protein evolution, our analyses revealed a tendency for accelerated substitutions in branches downstream (more recent) of mtDNA-encoded changes nearby in the COX 3D structure. Our theoretical model and computer simulations showed CWS is prevalent when selection is relatively weak: that is, $1 < S < 10$. In this range of natural selection, the assumption of the time lag between mtDNA and nDNA substitutions is reasonable. Consider compensatory mutations at an nDNA codon following a deleterious fixation at a physically interacting site. If we assume a nonsynonymous mutation rate of 7.5×10^{-10} (three-fourths of the neutral mutation rate per year per site, 1.0×10^{-9}) and $S = 4$, the probability of a compensatory fixation at this codon within 8×10^{7} years (roughly the primate divergence time) is <0.4.

In order to account for the patterns observed among primates, our theoretical work focused on a scenario of fixations at nDNA-encoded subunits that compensate for deleterious substitutions in mtDNA-encoded subunits. Our analyses support this scenario but raise a further issue: compensatory evolution between mtDNA-encoded subunits and between nDNA-encoded subunits should result in excesses of same-branch and sequential changes for pairs of potentially interacting sites encoded in the same genome. For nDNA-encoded subunits, a lower mutation rate/site and the small number of PI-paired sites predict CWS at a rate that may be too slow to detect on the timescale of primate molecular evolution. However, for mtDNA-encoded subunits, the number of PI pairs is higher than for the other two comparisons (nDNA–nDNA, nDNA–mtDNA) and mutations rates are sufficiently high to detect CWS. We propose two potential explanations for the lack of evidence for compensatory protein evolution between mtDNA-encoded COX subunits. First, the three mtDNA-encoded subunits form a large hydrophobic core in the protein complex; interacting sites among core proteins may be under higher selective constraint than sites in the interface region. In addition, our models assumed a fixed mtDNA to nDNA ratio of effective population sizes of 1:4. Complete linkage in the mtDNA genome will promote greater reductions in this ratio due to both positively or negatively selected mutations and consequent declines in the efficacy of selection (Charlesworth 1994).

The strongest evidence for CWS has so far come from studies of RNA secondary structure and synonymous codon usage. RNA secondary structure studies in particular have been able to take advantage of the ability to identify compensatory substitutions from their biophysical properties (i.e., reestablishing Watson–Crick pairing; Chen and Stephan 2003). Identifying compensatory interactions are more difficult in protein structures, but we attempted to find biophysical support for compensation between PI pairs in mtDNA- and nDNA-encoded COX proteins. Among the 14 sites in nDNA with $\omega > 1$, we identified nine high-probability substitutions that radically changed the volume on interacting surface between mtDNA- and nDNA-encoded subunits in the primate lineages (Zhang 2000). Four of them had high-probability amino acid changes at nearby mtDNA-encoded sites either in the same branches or in ancestral branches. In all cases, the amino acid changes in nDNA and mtDNA caused volume changes in opposing directions. Although these pairs may be candidates for sites under functional compensation (supplementary table 2, Supplementary Material online), this approach is crude since the context of amino acid changes is not considered in predicting fitness consequences. Other sites with $\omega > 1$ experienced only "conservative" amino acid changes (similar volumes, polarities, and charges; Zhang 2000).

Although our results confirm a specific prediction of the CWS model, it is important to note that the CWS and brain-energy models are not mutually exclusive. Our analyses support that CWS has contributed to accelerated protein evolution of nDNA-encoded COX proteins, but functional adaptation for higher, or more efficient, metabolism may have also played a role in fast primate COX evolution. In addition, it is possible that mitochondrial proteins have evolved rapidly to reduce somatic mutation through oxidative damage to the mtDNA of long-lived animals such as primates (Rottenberg 2006).

Although our analyses focused on the COX complex, other sites of physical and functional interaction between mtDNA- and nDNA-encoded genes should show evidence of CWS among primates. For example, several genes in the mitochondrial electron transport chain complex I, III, and V, all of which are composed with mtDNA- and nDNA-encoded proteins, were suggested to be under accelerated evolution in primate lineages (Grossman et al. 2004). In addition, CWS predicts high rates of compensatory nuclear changes in other lineages where mtDNA mutation rates are high and population sizes are small, regardless of metabolic rates. Distinguishing between CN and CWS will require dense sampling of lineages within a taxa to separate same-branch and sequential changes. For example, a bias toward same-branch substitutions in the mammalian phylogeny strongly supports CN evolution to preserve Watson–Crick pairings in mitochondrial tRNA stem structures (Meer et al. 2010).

CWS allows the maintenance of fitness in the face of slightly deleterious mutations. Endosymbiont genomes often show high mutation rates and may have limited opportunities for genetic exchange (Moran and Plague 2004; Foster et al. 2005). Compensatory evolution in nuclear genomes may be critical for maintaining fitness in such cases and may help to avoid "mutational meltdown" in which small population size enhances the accumulation of

deleterious mutations which limits the carrying capacity of the population and ultimately results in extinction (Lynch 1996). nDNA-encoded genes that interact with endosymbiont genes are strong candidates for CWS. More generally, biological systems that involve interactions between genetic regions undergoing high rates of mutation to weakly deleterious alleles and/or reduced efficacy of selection and regions with a higher efficacy of selection (e.g., higher recombination rates, hemizygous inheritance) are likely to undergo CWS. Long-term evolution under CWS can lead to incompatibilities between populations despite conservative evolution (Kondrashov 1995) and it will be of interest to determine whether genes involved in such incompatibilities fall into these two categories.

Supplementary Material

Supplementary figures S1 and S2 and tables 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Tomoko Ohta and two anonymous reviewers for a number of valuable suggestions. We thank the Marmoset Genome Sequencing Consortium for making their data publicly available. We also thank The Broad Institute for generation and public release of the horse genome sequence. This work was supported by KAKENHI, Grand-in-Aid for Young Scientists (A) (22687021).

References

Adkins R, Honeycutt R, Disotell T. 1996. Evolution of eutherian cytochrome c oxidase subunit II: heterogeneous rates of protein evolution and altered interaction with cytochrome c. *Mol Biol Evol*. 13:1393–1404.

Aiello LC, Wheeler P. 1995. The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Curr Anthropol*. 36:199–221.

Azevedo L, Carneiro J, Van Asch B, Moleirinho A, Pereira F, Amorim A. 2009. Epistatic interactions modulate the evolution of mammalian mitochondrial respiratory complex components. *BMC Genomics*. 10:266.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A*. 76:1967–1971.

Charlesworth B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet Res*. 63:213–227.

Chen Y, Stephan W. 2003. Compensatory evolution of a precursor messenger RNA secondary structure in the *Drosophila melanogaster Adh* gene. *Proc Natl Acad Sci U S A*. 100:11499–11504.

Depristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet*. 6:678–687.

Doan JW, Schmidt TR, Wildman DE, Uddin M, Goldberg A, Hüttemann M, Goodman M, Weiss ML, Grossman LI. 2004. Coadaptive evolution in cytochrome c oxidase: 9 of 13 subunits show accelerated rates of nonsynonymous substitution in anthropoid primates. *Mol Phylogenet Evol*. 33:944–950.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol*. 17:68–74.

Duret L, Semon M, Piganeau G, Mouchiroud D, Galtier N. 2002. Vanishing GC-rich isochores in mammalian genomes. *Genetics* 162:1837–1847.

Elcock AH, Gabdoulline RR, Wade RC, Mccammon JA. 1999. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. *J Mol Biol*. 291:149–162.

Foster J, Ganatra M, Kamal I, et al. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol*. 3:e121.

Goldberg A, Wildman DE, Schmidt TR, Hüttemann M, Goodman M, Weiss ML, Grossman LI. 2003. Adaptive evolution of cytochrome c oxidase subunit VIII in anthropoid primates. *Proc Natl Acad Sci U S A*. 100:5873–5878.

Green RE, Malaspinas A-S, Krause J, et al. 2008. A complete neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134:416–426.

Grossman LI, Schmidt TR, Wildman DE, Goodman M. 2001. Molecular evolution of aerobic energy metabolism in primates. *Mol Phylogenet Evol*. 18:26–36.

Grossman LI, Wildman DE, Schmidt TR, Goodman M. 2004. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends Genet*. 20:578–585.

Hartl DL, Taubes CH. 1996. Compensatory nearly neutral mutations: selection without adaptation. *J Theor Biol*. 182:303–309.

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.

Hershberg R, Petrov DA. 2008. Selection on codon bias. *Annu Rev Genet*. 42:287–299.

Iizuka M, Takefu M. 1996. Average time until fixation of mutants with compensatory fitness interaction. *Genes Genet Syst*. 71:167–173.

Innan H, Stephan W. 2001. Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. *Genetics* 159:389–399.

Johnson ET, Parson WW. 2002. Electrostatic interactions in an integral membrane protein. *Biochemistry (Mosc)*. 41:6483–6494.

Kenigsberg E, Bar A, Segal E, Tanay A. 2010. Widespread compensatory evolution conserves DNA-encoded nucleosome organization in yeast. *PLoS Comput Biol*. 6:e1001039.

Kenyon L, Moraes CT. 1997. Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proc Natl Acad Sci U S A*. 94: 9131–9135.

Kimura M. 1964. Diffusion models in population genetics. *J Appl Probab*. 1:177–232.

Kimura M. 1985. The role of compensatory neutral mutations in molecular evolution. *J Genet*. 64:7–19.

Kondrashov AS. 1995. Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *J Theor Biol*. 175:583–594.

Kuma K, Iwabe N, Miyata T. Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. *Mol Biol Evol*. 12:123–130.

Lawrie D, Petrov DA, Messer PW. 2011. Faster than neutral evolution of constrained sequences: the complex interplay of mutational biases and weak selection. *Genome Biol Evol*. 3:383–395.

Li W-H. 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J Mol Evol*. 24:337–345.

Ludwig MZ. 2002. Functional evolution of noncoding DNA. *Curr Opin Genet Dev*. 12:634–639.

Lunzer M, Golding GB, Dean AM. 2010. Pervasive cryptic epistasis in molecular evolution. *PLoS Genet*. 6:e1001162.
Lynch M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol Biol Evol*. 13:209–220.
Lynch M. 2010. Scaling expectations for the time to establishment of complex adaptations. *Proc Natl Acad Sci U S A*. 107:16577–16582.
Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* 311:1727–1730.
Meer MV, Kondrashov AS, Artzy-Randrup Y, Kondrashov FA. 2010. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature* 464:279–282.
Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev*. 14:627–633.
Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* 150:1133–1141.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 3:418–426.
Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246:96–98.
Osada N, Kusuda J, Hirata M, Tanuma R, Hida M, Sugano S, Hirai M, Hashimoto K. 2002. Search for genes positively selected during primate evolution by 5'-end-sequence screening of cynomolgus monkey cDNAs. *Genomics* 79:657–662.
Osada N, Uno Y, Mineta K, Kameoka Y, Takahashi I, Terao K. 2010. Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*. *Mol Ecol*. 19:2884–2895.
Pollock DD, Taylor WR, Goldman N. 1999. Coevolving protein residues: maximum likelihood identification and relationship to structure. *J Mol Biol*. 287:187–198.
Rottenberg H. 2006. Longevity and the evolution of the mitochondrial DNA-coded proteins in mammals. *Mech Ageing Dev*. 127:748–760.
Schmidt TR, Wu W, Goodman M, Grossman LI. 2001. Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase. *Mol Biol Evol*. 18:563–569.
Shen Y-Y, Liang L, Zhu Z-H, Zhou W-P, Irwin DM, Zhang Y-P. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proc Natl Acad Sci U S A*. 107:8666–8671.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.
Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. 1995. Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 A. *Science* 269:1069–1074.
Uddin M, Opazo J, Wildman D, Sherwood C, Hof P, Goodman M, Grossman L. 2008. Molecular evolution of the cytochrome c oxidase subunit 5A gene in primates. *BMC Evol Biol*. 8:8.
Wang HY, Chien HC, Osada N, Hashimoto K, Sugano S, Gojobori T, Chou CK, Tsai SF, Wu CI, Shen CK. 2007. Rate of evolution in brain-expressed genes in humans and other primates. *PLoS Biol*. 5:e13.
Wang Z, Pollock D. 2007. Coevolutionary patterns in cytochrome c oxidase subunit I depend on structural and functional context. *J Mol Evol*. 65:485–495.
Weissman DB, Feldman MW, Fisher DS. 2010. The rate of fitness-valley crossing in sexual populations. *Genetics*. 110:123240.
Wright F. 1990. The 'effective number of codons' used in a gene. *Gene* 87:23–29.
Wu W, Goodman M, Lomax MI, Grossman LI. 1997. Molecular evolution of cytochrome c oxidase subunit IV: evidence for positive selection in simian primates. *J Mol Evol*. 44:477–491.
Wu W, Schmidt TR, Goodman M, Grossman LI. 2000. Molecular evolution of cytochrome c oxidase subunit I in primates: is there coevolution between mitochondrial and nuclear genomes? *Mol Phylogenet Evol*. 17:294–304.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
Yeang C-H, Haussler D. 2007. Detecting coevolution in and among protein domains. *PLoS Comput Biol*. 3:e211.
Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol*. 50:56–68.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.