

An investigation of the variation in the transition bias among various animal mitochondrial DNA

Elise M.S. Belle¹, Gwenael Piganeau², Mike Gardner³, Adam Eyre-Walker^{*}

Centre for the Study of Evolution, School of Biological Sciences, University of Sussex, BN1 9QG Brighton, United Kingdom

Received 20 May 2004; received in revised form 5 April 2005; accepted 17 May 2005

Available online 20 July 2005

Received by G. Pesole

Abstract

The transition:transversion ratio (ts/tv) is known to be very high in human mitochondrial DNA, but we have little information about this ratio in other species. Here we investigate the transition bias in animal mitochondrial DNA using single nucleotide polymorphism data at four-fold degenerate sites. We investigate this pattern of polymorphism in the cytochrome *b* gene (*cyt-b*) in 70 species using a total of 1823 mutations. We show that most species show a bias towards transitions but that the ratio varies significantly between species. There is little evidence for variation within orders or genera and between closely related species such as the great apes. The majority of the variation appears to be at a higher phylogenetic levels: between orders and classes. We test whether the variation in ts/tv ratio could be due to variation in the metabolic rate by considering whether the ratio is correlated to base composition. We find no evidence that the metabolic rate affects the ts/tv ratio. We also investigate the relative frequencies of C to T or T to C (C ↔ T) mutations and A to G or G to A (A ↔ G) mutations. We show that overall they occur at significantly different frequencies, and that there is significant variation in their relative frequency between species and between classes. We find no evidence in support of the hypothesis that this variation could be due to different metabolic rates. © 2005 Elsevier B.V. All rights reserved.

Keywords: Mutation pattern; Strand asymmetry; Metabolic rate

1. Introduction

Point mutations can be divided into transitions, changes between the purines A and G, or changes between the pyrimidines C and T, and transversions, changes between purines and pyrimidines. In mammalian nuclear DNA, transition mutations appear to be approx-

imately twice as frequent as transversions, this is evident in the substitution patterns of mammalian pseudogenes (Gojobori et al., 1982; Li et al., 1984), in synonymous and non-coding SNPs in humans (Cargill et al., 1999), in SNPs in mice (Lindblad-Toh et al., 2000) and in the divergence of coding and non-coding sequences in mammals (Rosenberg et al., 2003). However, transitions are about as common as transversions in synonymous and intron SNPs in *Drosophila* DNA (Moriyama and Powell, 1996).

In contrast to the modest transition bias observed in mammalian nuclear DNA, transitions appear to be about 15 times as frequent as transversions in human mitochondrial DNA (Brown et al., 1982; Tamura and Nei, 1993). However, it is unclear whether this high ts/tv ratio is unique to humans or whether it is a common feature of animal mitochondrial DNA (mtDNA). Yang and Yoder (1999) attempted to address this question using a dataset of cytochrome *b* sequences from 28 primate species. Unfortun-

Abbreviations: ts/tv, transition-transversion ratio; *cyt-b*, cytochrome b; *ND5*, NADH dehydrogenase subunit 5; *ND6*, NADH dehydrogenase subunit 6; TCR, transcription coupled repair.

* Corresponding author. Tel.: +44 1273 878 480; fax: +44 1273 877 351.

E-mail address: a.c.eyre-walker@sussex.ac.uk (A. Eyre-Walker).

¹ Present address: Dipartimento di Biologia, Università di Ferrara-Via Borsari, 46-44100 Ferrara-Italy.

² Present address: Modèles en biologie cellulaire et évolutive-UMR 7628, Laboratoire Arago BP 44-66651 Banyuls sur mer-France.

³ Present address: School of Biological Sciences-University of Bristol, Woodland Road-BS8 1UG Bristol-United Kingdom.

nately their analysis was inconclusive; there was some evidence that ts/tv varied across taxa, but the authors ultimately concluded that there was probably a problem with their method.

Estimating the ts/tv ratio is problematic in mitochondrial DNA because the rate of substitution is very high and mtDNA shows asymmetric base composition. This makes the correction for multiple substitutions complex. A solution to this problem is to use sequences from within a species; these sequences are so closely related that there should be no need to correct for multiple hits.

In this paper we examine the ts/tv ratio in animal mtDNA using data from 70 species in which multiple alleles from mtDNA have been sequenced. This method allowed us to test whether the transition bias is ubiquitous among animals and whether it has the same strength in all species. We also examined whether the frequencies of C ↔ T and G ↔ A mutations were the same, and if not, whether or not their relative rates varied across taxa. Intuitively one would expect the rate of C ↔ T mutation to equal that of G ↔ A, since a C ↔ T mutation on one strand is a G ↔ A on the other. However, base composition in mitochondrial DNA is often highly asymmetric; for example the heavy strand of human mitochondrial DNA, which is the coding strand of all but one of the protein coding genes, is 39% A, 13% T, 5% G and 42% C at third codon positions of four fold degenerate codons (Perna and Kocher, 1995). This suggests that the mutation pattern of the two strands differs, since there is no evidence of selection on synonymous codon use in mtDNA.

2. Material and methods

Datasets were obtained by scanning databases and back issues of the journal of *Molecular Phylogenetics and Evolution*. Sequences were directly retrieved from the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>). Nucleotide sequences were aligned by hand in Sequence Navigator, the complete or longest sequence being used as reference. We extracted cytochrome *b* sequences from 70 different species (Fig. 1 and Supplementary Table 1) that we classified according to their class (9 categories) and order (22). For each species we had between 4 and 188 sequences from different individuals. We also compiled the complete non-overlapping protein coding sequences from 53 humans (Ingman et al., 2000), 3 gorillas, 4 orangutans and 2 chimpanzees. To each of these complete sequences we added additional sequences where available (Supplementary Table 2). For example, there are an additional 16 *ND5* sequences from gorillas available, so our total number of *ND5* sequences was 19. Except *ND6*, all these genes are encoded on the L-strand of the mitochondrial genome.

We extracted synonymous polymorphisms segregating at four-fold, two-fold and zero-fold degenerate sites. Using sequences from within a species should make the analysis of the ts/tv ratio simple by eliminating the need to correct for multiple hits. However, to check that this assumption was reasonable we calculated Watterson's (1975) estimate of $2N_e u$ for four-fold sites and the two types of two-fold sites (i.e. AG CT sites) as $\theta_w = \frac{s}{\sqrt{\sum_i^n -1}/i}$ where l is the number of sites, s the number of polymorphisms and n the number of sequences. If θ_w is less than 0.2 we can be fairly confident that each mutation is a unique event, i.e. the chance of a polymorphic site having been hit twice is only about 10% if $\theta_w = 0.2$. We excluded from our analyses any species in which θ_w values was greater than 0.2 for any of the three categories of site.

We used a χ^2 test of independence to test whether there was variation in the ts/tv ratio between taxa; this test is justified because the numbers of transitions and transversions are binomially distributed (in essence, our alignment of sequences can be thought of as a sample of some infinitely long alignment, so all discrete characters within the alignment are multi-nomially distributed). To assess whether the ratio of C ↔ T transitions to A ↔ G transitions varied significantly between species we calculated the log odds ratio $\alpha = \ln(x_{ct}/y_{ct})/(x_{ag}/y_{ag})$ for each species, where x_{ct} is the number of sites with a C ↔ T transition, y_{ct} is the number of sites which are fixed for C or T, x_{ag} is the number of sites with an A ↔ G transition, and y_{ag} is the number of sites fixed for A or G. α is expected to be normally distributed, since the numbers of each transition are binomially distributed, with a variance equal to $V_\alpha = 1/x_{ct} + 1/y_{ct} + 1/x_{ag} + 1/y_{ag}$ (Selvin, 1995). Thus the sum $\sum (\alpha - a)^2 / V_\alpha$ is χ^2 distributed with $(k-1)$ degrees of freedom where $a = \sum ((1/V_\alpha)\alpha) / \sum (1/V_\alpha)$ where k is the number of taxa compared.

To assess the phylogenetic level at which the differences in ts/tv ratio, or the relative rates of the two transition types, occurred, we performed a nested analysis of variance. The analysis was performed on $\arcsin \sqrt{(ts/(ts + tv))}$ for the ts/tv ratio, to make the data normally distributed, and α , for the relative rates of transitions.

We used the method of orthogonal contrasts to investigate the correlations between quantities in order to remove phylogenetic non-independence (Felsenstein, 1985). This method accounts for the bias that may be introduced by shared ancestry among our species, and thus allows direct comparative analyses of correlated evolution. For mammals, we used the phylogeny given in Liu et al. (2001); for the *Calomys* species, we used the one of Salazar-Bravo et al. (2001). For other species, we assumed the phylogeny given in the taxonomy browser of the National Center for Biotechnology Information (NCBI), and for a few species we used information from

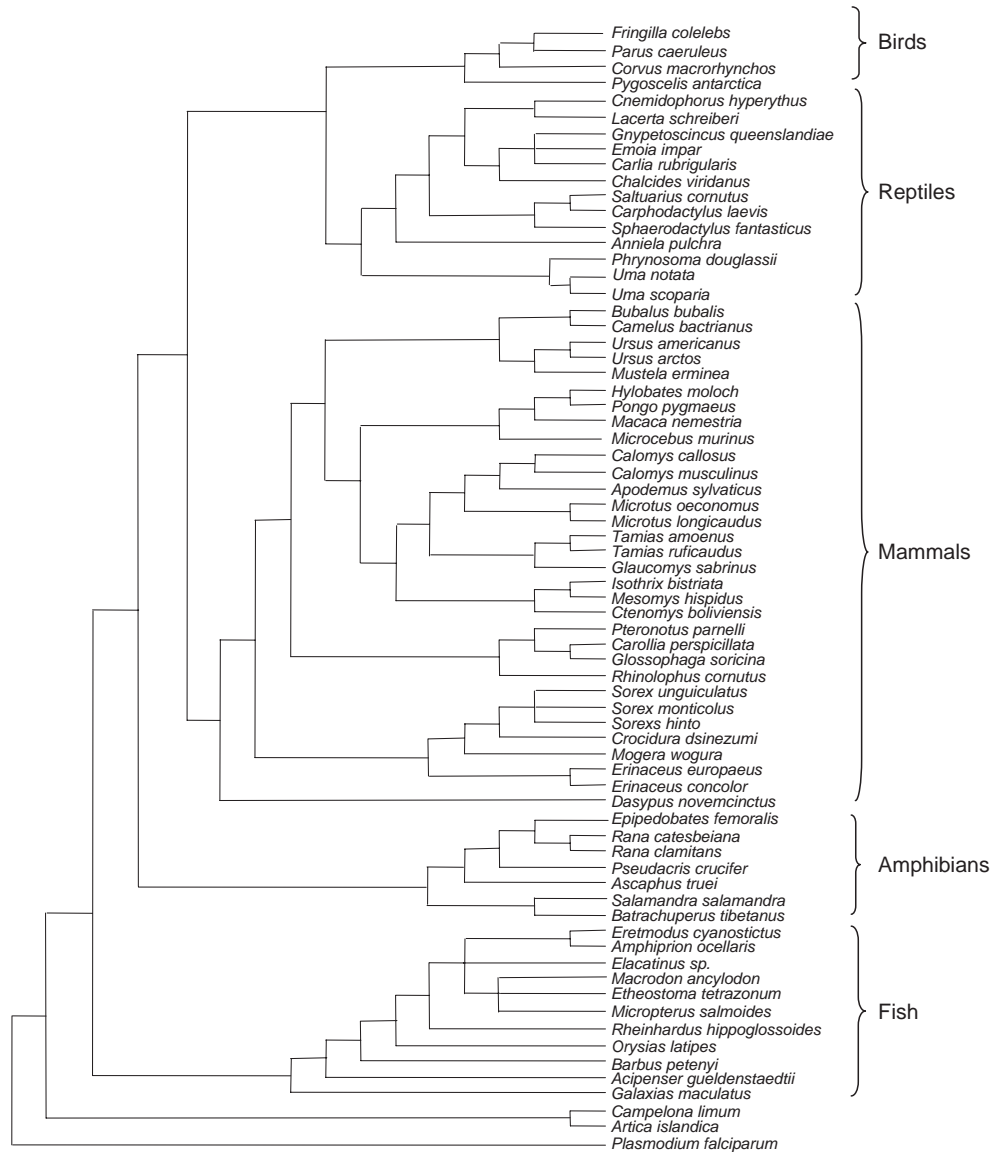


Fig. 1. Phylogeny of the species used in this study.

the Finnish Museum of Natural History (<http://www.fnnh.helsinki.fi/users/haaramo/>).

Data on metabolic rates was compiled for 15 of our mammalian species from Bromham et al. (1996) and White and Seymour (2003).

To estimate the time of divergence between various rodent species we used a molecular clock based on amino acid divergence, calibrated using the divergence between *cyt-b* genes of mouse and rat. The time of divergence between those species is still highly controversial, so we used a range of estimates from the recent literature from the figure of 16 MYR given by Springer et al. (2003) to that of 33 MYR given by Nei et al. (2001). We used amino acid divergence rather than the divergence at synonymous sites because there is evidence of changes in the pattern of mutation at synonymous sites which invalidates the models used to correct for multiple hits. We corrected the amino acid

divergence for multiple hits using the correction formula $d_n = -\ln(1-p)$ where p is the proportion of amino acids which differ between the two sequences being considered.

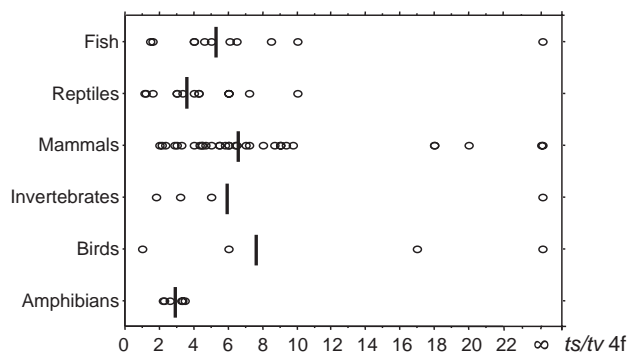


Fig. 2. The ts/tv ratio at four-fold degenerate sites. Each point represents one species, the average for each class is represented by the vertical bars.

3. Results

3.1. The transition bias across species

The numbers of synonymous transition and transversion polymorphisms segregating at four-fold degenerate sites in the cytochrome *b* gene of 70 species are given in Fig. 2 and Supplementary Table 1. We chose not to apply any corrections to our estimates and present here the raw ratios to simplify the presentation, they are therefore upwardly biased; however, none of the analyses were actually performed on the ts/tv ratio itself. The ts/tv ratio in *cyt-b* is always greater or equal to 1, it varies from 1.0 in the chinstrap penguin, *Pygoscelis antarctica*, to 20.0 in the long-clawed shrew, *Sorex unguiculatus*, (or infinity if we consider species for which no transversion was observed). There is significant variation in the ts/tv ratio at four-fold degenerate sites between species ($\chi^2=138.8$, $df=69$, $p<0.0001$). This conclusion is unaltered by the inclusion of the complete primate sequences (Table 1). It should be noted here that there are two different ways in which to define the ts/tv ratio. The method we have adopted here is to consider the transition rate over the sum of the two transversion rates; so if the ts/tv ratio is equal to one, then the rate at which each transition occurs is actually twice that at which each transversion pathway occurs. This is the generally used definition of this parameter.

3.2. Variation in transition bias

3.2.1. No variation in ts/tv within genera

Given that the ts/tv ratio varies between taxa, for example in Primates it varies from 4.7 in *Pongo pygmaeus* to 18.0 in *Phyllobates moloch*, it is of interest to investigate over what timescale this transition bias changes. In our data set, we have 8 groups of species from the same genus: two *Rana* (frogs), two *Uma* (fringe-toed lizards), two *Calomys* (vesper mice), two *Microtus* (meadow voles), two *Tamias* (chipmunks), two *Erinaceus* (hedgehogs), three *Sorex* (shrews) and two *Ursus* (bears). The ts/tv ratios are not significantly different within any of these genera and the summed χ^2 is also not significant ($\chi^2=10.2$, $df=10$, N.S.).

We then tested whether there is a significant variation at the larger timescale of orders. We have 10 groups of species from the following orders: Anura (4 species), Carnivora (3), Chiroptera (4), Insectivora (7), Passeriformes (3), Perciformes (6), Primates (4), Rodentia (10), Ruminantia (2) and

Squamata (14). The ts/tv ratios are not significantly different within any of these orders, except in Perciformes ($\chi^2=12.2$, $df=5$, $p=0.03$). If we sum the χ^2 values, there is no evidence, overall, of heterogeneity in the ts/tv ratio within orders ($\chi^2=63.36$, $df=48$, N.S.).

3.2.2. No variation within great apes

Although the great apes do not belong to the same genus, they are much more closely related, in terms of time, than many species belonging to the same genus; indeed all great apes are thought to have diverged from one another within the last 18 MYR (Goodman et al., 1998). We therefore decided to investigate the evolution of the ts/tv ratio within the great apes. This is also of interest given the study of Yang and Yoder (1999) in which they found some evidence of a difference in the ts/tv ratio within primates. In this analysis, we can take advantage of the fact that the complete mtDNA sequences are available in several primates. The numbers of transitions and transversions at 4-fold degenerate sites, summed across all genes, in humans, chimpanzees, gorillas and orangutans are given in Table 1; there were too few transversions to test whether there was any variation in the ts/tv ratio between genes. There is no evidence of variation in the ts/tv ratio: overall the χ^2 test is not significant ($\chi^2=5.8$, $df=3$, $p=0.12$). Furthermore, after applying Bonferroni correction for multiple comparisons, the ts/tv ratio is not significantly different, either between the two species for which we have more than 150 polymorphisms, humans and orangutans, or between humans and its closest relative, the chimpanzee. These results therefore suggest that the ts/tv ratio has not changed much during the evolution of the great apes.

3.2.3. Variation at higher phylogenetic levels

Since we found no evidence of a significant change in the ts/tv ratio over a short period of time we performed a nested analysis of variance nesting the data as orders within classes to test whether there is variation at larger timescales. Using this analysis we can ask whether there is significant variation in the ts/tv ratio between classes, given the variation between orders. We found a significant variation in the ts/tv ratio (actually the arcsine transformation of the data: see Material and Methods) between classes, given the variation between orders ($F=2.97$, $df=8$, $p=0.009$), and between orders within a class, knowing the variation within orders ($F=2.18$, $df=13$, $p=0.025$). We therefore have evidence of heterogeneity in the ts/tv ratio in animal mitochondrial DNA between orders and classes given the variation at lower phylogenetic levels.

One potential problem with these kind of analyses is that different phylogenetic levels (genus, orders, classes) are not necessarily at the same phylogenetic depth (i.e. equally divergent) in different groups. However, these analyses serve the purpose to stress that, overall, the ts/tv ratio seems to be rather homogeneous when one looks at low

Table 1

The numbers of transition (ts) and transversion (tv) mutations at four-fold degenerate sites in protein coding genes of selected primates

Species	# sequences	ts	tv	ts/tv
<i>Homo sapiens</i>	53	165	12	13.75
<i>Pan troglodytes</i>	2 to 20	33	3	11.00
<i>Gorilla gorilla</i>	3 to 19	17	3	5.67
<i>Pongo pygmaeus</i>	2 to 15	220	32	6.87

phylogenetic levels and varies only when one reaches higher levels like classes (which are better defined).

3.3. Strand asymmetry in transition rates

Transitions can be divided up into those involving purines and those involving pyrimidines. Overall the $T \leftrightarrow C$ transitions ($Ts_{T \leftrightarrow C}$), appear to occur more frequently than the $A \leftrightarrow G$ transitions ($Ts_{A \leftrightarrow G}$) since 27.4% of two-fold and four-fold CT sites (i.e. sites which are either C or T or both) show a transition against 18.5% of AG sites. To assess whether $T \leftrightarrow C$ and $A \leftrightarrow G$ transitions occur at significantly different frequencies, we performed a two-by-two χ^2 heterogeneity test for each species. Individually, the χ^2 values were significant in one about a quarter of the species (17/70), and overall the summed χ^2 is highly significant ($\chi^2=488.2$, $df=69$, $p<0.0001$), indicating that the two transitions do not occur at similar rates. Moreover, there is evidence that the relative frequencies of $T \leftrightarrow C$ and $A \leftrightarrow G$ transitions differ between species ($\chi^2=292.5$, $df=69$, $p<0.0001$), suggesting that not all species show the same asymmetric pattern of mutation. About 75% of the taxa show a higher rate of $T \leftrightarrow C$ transitions, with the relative rates of transition varying from *Etheostoma tetrazonum* in which $A \leftrightarrow G$ transitions occur twice as frequently as $T \leftrightarrow C$ transitions, to *Erinaceus concolor* in which $T \leftrightarrow C$ transitions outnumber $A \leftrightarrow G$ transitions almost 5-fold.

3.3.1. Variation of the strand asymmetry within genera

Surprisingly, the relative rates of the two transitions appear to vary over quite short timescales. In our 8 groups of species which come from the same genus, two showed significant variation in the relative rate of transitions: *Microtus* ($\chi^2=7.2$, $df=1$, $p<0.01$) and *Tamias* species ($\chi^2=4.1$, $df=1$, $p<0.05$), both of which are rodents. The summed χ^2 is also significant ($\chi^2=19.6$, $df=9$, $p<0.05$) which suggests that these significant results are not a consequence of multiple tests. Without rodents, the summed χ^2 was no longer significant ($\chi^2=6.51$, $df=6$, N.S.) suggesting that they accounted for most of the variation. Using a divergence time for mouse and rat between 16 MYR (Springer et al., 2001) and 33 MYR (Nei et al., 2003) to calibrate a molecular clock, we estimated that the two *Microtus* species diverged between 11.5 and 23.8 MYR ago and the two *Tamias* species between 2.0 and 4.1 MYR ago. Although the *Microtus* species diverged within a similar time frame as great apes (18 MYR), and the *Tamias* species over a much shorter time frame, there is no evidence of variation in the relative rate of transitions within primates.

3.3.2. Variation of the strand asymmetry at higher levels

To investigate whether there is also variation at larger timescales, we performed a nested analysis of variance on α , nesting the data as genera within orders and orders within classes. We found significant variation in the relative frequencies of the two types of transitions between classes,

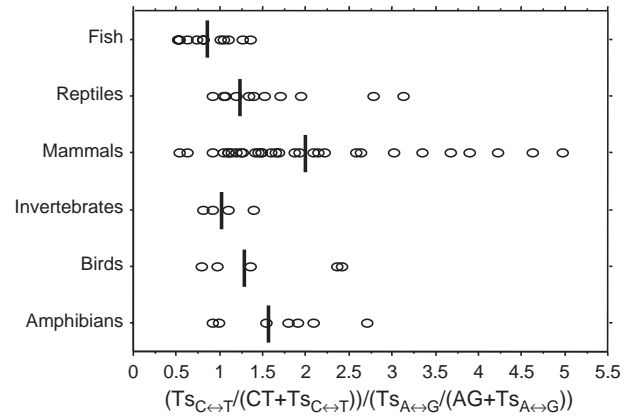


Fig. 3. The relative rates of $C \leftrightarrow T$ and $A \leftrightarrow G$ transitions. The quantity plotted is the proportion of CT sites with a CT transition at two-fold and four-fold sites, divided by the proportion of AG sites with an AG transition (i.e. $(X_{ct}/(X_{ct}+Y_{ct})) / (X_{ag}/(X_{ag}+Y_{ag}))$). Each point represents one species, the average for each class is represented by the vertical bars.

given the variation between orders ($F=7.25$, $df=7$, $p=0.043$), but we did not find any evidence of variation between orders or genera. However, for most genera we had only one species, which may restrict the power of the analysis, so we repeated the analysis nesting only orders within classes; the result was unaltered. The variation between classes does not appear to be due to any particular class (Fig. 3). We therefore have a somewhat odd situation in which we have evidence for heterogeneity in the relative rates of transition between closely related species, although this appears to be due to one order, the rodents, and between classes, but no evidence of heterogeneity at intermediate phylogenetic levels.

4. Discussion

4.1. The transition bias

Our results show that synonymous transition polymorphisms appear to be generally more common than transversions at four-fold degenerate sites in numerous animal mitochondrial DNAs, in addition to humans as already shown by Brown et al. (1982) and Tamura and Nei (1993). This is in sharp contrast with plant mitochondrial DNA where a transversion bias has been observed (Wolfe et al., 1987). This result therefore shows that this bias is widespread in animal mitochondrial DNA and suggests that the mitochondrial genomes of plants and animals follow very different patterns of evolution.

As with many observations in molecular evolution, we can explain the excess of transitions either as a consequence of selection or as a consequence of mutation. Selection is known to act upon synonymous codon use in many organisms, but it seems unlikely that selection is responsible for the transition bias for several reasons. First, there is a significant correlation between the ts/tv ratio at four-fold

and zero-fold degenerate sites (Spearman: $r_s=0.300$, $p=0.002$); if selection on synonymous codon use was the cause of the transition bias it would not be expected to cause a similar bias at zero-fold sites. Moreover, only very specific models of selection on synonymous codon use are likely to affect the ts/tv ratio greatly; for example if C and T ending codons are equally fit but are more or less fit than A and G ending codons. This conjecture, that selection on synonymous codon bias would not affect the ts/tv ratio, is supported by data from *Drosophila simulans*. In this species, there is selection on synonymous codon bias (Akashi and Schaeffer, 1997; Kliman, 1999; Begun, 2001) but the ts/tv ratio for synonymous polymorphisms at four-fold degenerate sites is very similar to that in introns, which we presume are not subject to selection, or at least not to the same selection (Moriyama and Powell, 1996); hence selection on synonymous codon use seems not to have affected the ts/tv ratio in this species. Moreover, selection on synonymous codon use has never been reported in mitochondrial DNA. Similarly, compositional selection is unlikely to affect the ratio, since selection for high GC content would not modify the ts/tv ratio unless the ratio differs for GC and AT base pairs. Although we cannot rule out selection, it seems most likely that the transition bias is mutational in origin.

4.2. Reasons for variation in ts/tv ratio

We found no variation in the ts/tv ratio within orders, genera or between closely related species such as the great apes. This result appears to contradict those of Yang and Yoder (1999), and of Adachi and Hasegawa (1996). Yang and Yoder found evidence of variation in the ts/tv ratio within primates but they suspected that there might be a problem with their method. Adachi and Hasegawa estimated that the transition rate in humans was twice that in other primates, however they did not confirm this statistically. If we compare the number of transitions and transversions in humans (165 and 12 respectively) to the sum across the other great apes (278 and 38), the ts/tv ratio in humans is approximately twice that in the great apes, as Adachi and Hasegawa found; however, this is not significant ($p=0.06$ Fishers exact test) particularly if we correct for multiple comparisons.

Most of the variation in the transition bias appears to be at a higher level. For the reasons given above it seems likely that this variation is mutational in origin, rather than caused by selection. A number of different factors could yield variation in the ts/tv ratio. For example, Martin (1995) has shown that the AT content of mammalian mtDNA is correlated to metabolic rate. It is therefore possible that metabolic rate also affects the ts/tv ratio. He suggested that the metabolic rate affects AT content through the mutagenic effects of oxygen free radicals which are released in the mitochondria during aerobic respiration. Oxygen free radicals have at least two mutagenic effects: (i) they generate abasic sites (Schaaper et al., 1983; Randall et al.,

1987), i.e. sites at which the sugar backbone of the DNA remains intact, but the base has been removed; and (ii) they lead to spontaneous deamination (Lindahl, 1993).

The first pathway-generation of abasic sites-is expected to affect the ts/tv ratio because DNA polymerases tend to incorporate A opposite abasic sites; if the site which has become abasic was G, then we get a C to A transversion; if it was C, we get a G to A transition; and if it was A, we get a T to A transversion. Hence, an increase in metabolic rate is expected to decrease the ts/tv ratio, unless oxygen free radicals tend to make G sites abasic more readily than C or T sites. It is also expected to increase AT content and lead to a temporary excess of GC→AT (C→T or G→A) over AT→GC transitions; the increase is only temporary because at equilibrium we expect the number of GC→AT mutations to be approximately equal to the number of AT→GC mutations, unless selection is acting upon synonymous codon use (Eyre-Walker, 1997).

The second pathway – spontaneous deamination – is expected to increase the ts/tv ratio because the base most likely to undergo spontaneous deamination is C, which becomes T. Hence an increase in metabolic rate is expected to increase the ts/tv ratio, increase AT content, and to temporarily increase the proportion of transitions which are GC→AT.

Is variation in metabolic rate responsible for the variation in the ts/tv ratio? We can test this hypothesis by investigating whether or not the proportion of transitions which are GC→AT is correlated to AT content since we expect both of these to be positively correlated to metabolic rate and to each other under both biochemical pathways. To remove phylogenetic non-independence we used the method of orthogonal contrasts (Felstenstein, 1985) assuming the phylogeny given in Fig. 1. There is a significant negative correlation between the proportion of transitions which are GC to AT ($ts_{GC\rightarrow AT}/ts$) and AT content ($r=-0.300$, $p=0.015$) (Fig. 4), which is opposite to our prediction. This negative correlation remains if we restrict our analysis to mammals only, the group for which we have the most

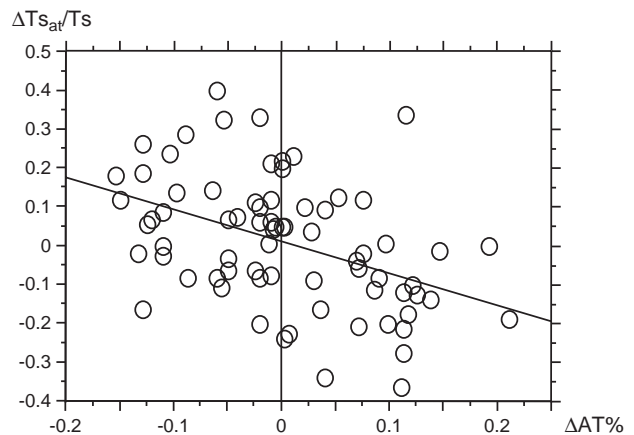


Fig. 4. The correlation between the change in the proportion of transitions which are GC→AT ($\Delta Ts_{at}/Ts$) and the change in AT content ($\Delta AT\%$) over the phylogeny given in Fig. 1 ($r=-0.300$, $p<0.015$) $\Delta Ts_{at}/Ts$.

species; the effect is therefore not due to differences between classes. This negative correlation is more easily explained in terms of selection acting upon synonymous codon use: if selection favours A and T ending codons then most mutations segregating in the population will tend to be AT→GC mutations (Akashi, 1995).

An alternative analysis is to investigate the correlation between GC content and the proportion of mutations which are transitions. We do not have a clear prediction since it depends on whether or not the first or second pathway, by which oxygen free radicals cause mutation, dominates. If the pathway generating abasic sites predominates we would expect a negative correlation between the AT content and the ts/tv ratio; but if the pathway leading to spontaneous deamination is more important we would then expect a positive correlation between these two variables. Recent studies have described the existence of a repair system for abasic sites (Pinz and Bogenhagen, 1998, Bogenhagen et al., 2001), while no repair mechanisms have yet been found for spontaneous deaminations. Therefore if metabolic rates have an effect on the ts/tv ratio, we would expect a positive correlation between the AT content and the ts/tv ratio. We found no significant correlation, therefore suggesting that variation in metabolic rate does not affect the ts/tv ratio greatly.

More generally, we can test whether the ts/tv ratio is correlated to metabolic rate by using direct estimates of the metabolic rates. We obtained estimates for 15 of the species we surveyed, all of them mammals; there was no significant correlation either with the ts/tv ratio or with the proportion of transitions which were GC to AT.

4.3. Strand asymmetry in the transition rates

We have shown that overall there is a strand asymmetry in the frequency of transitions in the gene considered (cytochrome *b*), C↔T transitions occur approximately 1.5 times more frequently than A↔G transitions, and the relative frequency of $T_{S_{T\rightarrow C}}$ and $T_{S_{A\rightarrow G}}$ transitions vary significantly between taxa.

There are at least two possible ways in which a strand asymmetry could be produced: it could be due to the asymmetric pattern of DNA replication in mitochondria, or to the transcription process. There is some debate about how replication in mitochondria proceeds. The traditional view is that DNA replication starts on the L-strand, in mammals at least, and proceeds about two thirds of the way round the mitochondrial genome, until the origin of replication on the H-strand is exposed, at which point replication along the H-strand begins. The H-strand was therefore thought to be single stranded for some considerable time during which period it is potentially exposed to hydrolytic and oxidative damage. More recent work has suggested that in mammals the replication of mitochondrial DNA may be more conventional with synchronous leading and lagging strand replication (Holt et al., 2000) so that neither of the strands

stays single stranded for long (Yang et al., 2002). However, we do not know if this mechanism applies to other vertebrate species, therefore, we have tested whether, according to the traditional model, the time of single stranded exposure could possibly lead to a strand asymmetry. If some regions stay single stranded for prolonged periods, then the two processes, which might occur on the H-strand during this phase, are the deamination of cytosine to uracil, and hence thymine (after two rounds of DNA replication), and the deamination of adenine to hypoxanthine, and hence to guanine. If these pathways occur at different rates, as we might expect, then there will be strand asymmetry in the relative rates of the two transitions.

Reyes et al. (1998) estimated that the deamination of cytosine is about twice as frequent as that of adenine in mammalian mtDNA, and estimates from nuclear DNA suggest that it is much more frequent (Lindahl, 1993). However, data from Faith and Pollock (2003) suggest a primary role of mutations from adenine to hypoxanthine. If the deamination of cytosine is more frequent on the H-strand, then this will manifest itself as a higher rate of G→A mutation on the coding L-strand, which is the opposite of the pattern generally found in the taxa studied here. Note that mitochondrial geneticists often refer to the H-strand as the coding strand for most protein coding genes, because it is the strand which is transcribed for these genes, but this is unconventional, because genes are usually given 5' to 3' which means the coding strand is the L-strand for most protein coding genes, including *cyt-b*, i.e. the mitochondrial sequence in GenBank is a copy of the L-strand. This does not necessarily mean that replication cannot explain the asymmetry: the relative rates of C↔T and A↔G mutation will depend as much on the rate of C→T and A→G mutation as on T→C and G→A, and there may be other processes which depend on the time the DNA is single stranded.

To test whether the relative frequencies of C↔T and A↔G transitions could be affected by the DNA replication we looked at the correlation between α and the time for which a gene is single stranded in humans (times for which genes are single stranded came from Reyes et al. (1998) ignoring *ND6* which is coded on the H-strand): there was no significant correlation. However, humans are one of the minority of species which actually show a higher rate of A↔G mutation, this test may therefore not be applicable to the majority of species studied here.

Transcription could also potentially explain the differences in the relative rates of the two transitions since the two strands of the mitochondrial genome are likely to be transcribed to different extents. In mammals there are 3 transcription units, two which are transcribed from the H-strand and one from the L-strand. The net effect in mammals is that the L-strand is transcribed more often than most of the H-strand (the exception being the two rRNA genes), and this means that the H-strand is single stranded more often, just as it is during DNA replication, and would therefore be

more prone to cytosine deaminations; however this is not what we observe, so we think that this aspect of transcription is unlikely to explain the bias.

Certain types of repair are targeted towards the strand being transcribed through the process of transcription coupled repair (TCR). TCR is a repair mechanism targeted towards pyrimidine dimers on the transcribed strand. The principle mutation from pyrimidine dimers are C→T transitions, so TCR is expected to yield a lower frequency of this mutation on the protein coding L-strand (Oller et al., 1992).

To test whether transcription could be responsible for the differences in the rates of the two transitions we compared the relative rates in the human *ND6* gene, which is transcribed many times more frequently than the other protein coding genes, to that in the other genes; there was no significant difference ($\chi^2=0.30$, $df=1$, N.S). Therefore the strand asymmetry in the relative rates of the two types of transition does not appear to be due to the transcription mechanism.

There is also the question of why the relative rates of the two types of transition differ between taxa. For the reasons discussed in relation to the ts/tv ratio we might expect metabolic rate to affect the relative rates of the two transitions. We therefore investigated whether α was correlated to either metabolic rate, or AT content using orthogonal contrasts; in neither case did we find a correlation, suggesting that metabolic rate has no effect on the relative rates of the two transition types.

5. Conclusion

We have shown that the ts/tv ratio, and the relative rates of the two transition types, vary significantly between species. However, the timescale over which these quantities vary is different. The relative rates of the two transition types seems to be more labile, at least in rodents, with significant variation even between fairly closely related species; however the reason for the variation in either quantity has to be further investigated, as this seems unrelated to metabolic rates.

Acknowledgements

This study was funded by the BBSRC (A.E.W., M.G., E.B. and G.P.) the Royal Society (A.E.W.), and the School of Life Sciences at the University of Sussex (E.B.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2005.05.019.

References

- Adachi, J., Hasegawa, M., 1996. Tempo and mode of synonymous substitutions in mitochondrial DNA of primates. *Mol. Biol. Evol.* 13, 200–208.
- Akashi, H., 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* 139, 1067–1076.
- Akashi, H., Schaeffer, S.W., 1997. Natural selection and the frequency distributions of “silent” DNA polymorphism in *Drosophila*. *Genetics* 146, 295–307.
- Begun, D.J., 2001. The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol. Biol. Evol.* 18, 1343–1352.
- Bogenhagen, D.F., Pinz, K.G., Perez-Jannotti, R.M., 2001. Enzymology of mitochondrial base excision repair. *Prog. Nucleic Acid Res. Mol. Biol.* 68, 257–271.
- Bromham, L., Rambaut, A., Harvey, P.H., 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J. Mol. Evol.* 43, 610–621.
- Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C., 1982. Mitochondrial DNA sequences of primates, tempo and mode of evolution. *J. Mol. Evol.* 18, 225–239.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., et al., 1999. Characterization of single nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22, 231–238.
- Eyre-Walker, A., 1997. Differentiating between selection and mutation bias. *Genetics* 147, 1983–1987.
- Faith, J.J., Pollock, D.D., 2003. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics* 165, 735–745.
- Felstenstein, J., 1985. Phylogenies and the comparative method. *Am. Nat.* 125, 1–15.
- Gojobori, T., Li, W.H., Graur, D., 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.* 18, 360–369.
- Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., et al., 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* 9, 585–598.
- Holt, I.J., Lorimer, H.E., Jacobs, H.T., 2000. Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* 100, 515–524.
- Ingman, M., Kaessmann, H., Paabo, S., Gyllensten, U., 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708–713.
- Kliman, R.M., 1999. Recent selection on synonymous codon usage in *Drosophila*. *J. Mol. Evol.* 49, 343–351.
- Li, W.-H., Wu, C.-I., Luo, C.-C., 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* 21, 58–71.
- Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature* 362, 709–715.
- Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., et al., 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.* 24, 381–386.
- Liu, F.G.-R., Miyamoto, M.M., Freire, N.P., Ong, P.Q., Tennant, M.R., et al., 2001. Molecular and morphological supertrees for eutherian (placental) mammals. *Science* 291, 1786–1789.
- Martin, A., 1995. Metabolic rate and directional substitution in animal mitochondrial DNA. *Mol. Biol. Evol.* 12, 1124–1131.
- Moriyama, E.N., Powell, J.R., 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* 13, 261–277.
- Nei, M., Xu, P., Glazko, G., 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. U. S. A.* 98, 2497–2502.
- Oller, A.R., Fijalkowska, I.J., Dunn, R.L., Schaaper, R.M., 1992. Transcription-repair coupling determines the strandedness of ultraviolet

- mutagenesis in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 89, 11036–11040.
- Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–358.
- Pinz, K.G., Bogenhagen, D.F., 1998. Efficient repair of abasic sites in DNA by mitochondrial enzymes. Mol. Cell Biol. 18, 1257–1265.
- Randall, S.K., Eritja, R., Kapland, B.E., Petruska, J., Goodman, M.F., 1987. Nucleotide insertion kinetics opposite abasic lesions in DNA. J. Biol. Chem. 262, 6864–6870.
- Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15, 957–966.
- Rosenberg, M.S., Subramanian, S., Kumar, S., 2003. Patterns of transitional mutation biases within and among mammalian genomes. Mol. Biol. Evol. 20, 988–993.
- Salazar-Bravo, J., Dragoo, J.W., Tinnin, D.S., Yates, T.L., 2001. Phylogeny and evolution of the neotropical rodent genus *Calomys*: inferences from mitochondrial DNA sequence data. Mol. Phylogenet. Evol. 20, 173–184.
- Schaaper, R.M., Kunkel, T.A., Loeb, L.A., 1983. Infidelity of DNA synthesis associated with bypass of apurinic sites. Proc. Natl. Acad. Sci. U. S. A. 80, 487–491.
- Selvin, S., 1995. Practical Biostatistical Methods. Duxbery Press, pp. 266–267.
- Springer, M.S., Murphy, W.J., Eizirik, E., O'Brien, S.J., 2003. Placental mammal diversification and the Cretaceous–Tertiary boundary. Proc. Natl. Acad. Sci. U. S. A. 100, 1056–1061.
- Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.
- Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276.
- White, C.R., Seymour, R.S., 2003. Mammalian basal metabolic rate is proportional to body mass. Proc. Natl. Acad. Sci. U. S. A. 100, 4046–4049.
- Wolfe, K.H., Li, W.-H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. U. S. A. 84, 9054–9058.
- Yang, Z., Yoder, A.D., 1999. Estimation of the transition/transversion rate bias and species sampling. J. Mol. Evol. 48, 274–283.
- Yang, M.Y., et al., 2002. Biased incorporation of ribonucleotides on the mitochondrial L-Strand accounts for apparent strand-asymmetric DNA replication. Cell 111, 495–505.