Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 355 (2005) 58 – 66

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

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

Elise M.S. Belle [1], Gwenael Piganeau [2], Mike Gardner [3], 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.

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 approximately 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. Unfortu-

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.

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.05.019

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 $2N_eu$ for four-fold sites and the two types of two-fold sites (i.e. AG CT sites) as $\theta_w = \frac{S}{l\sum_i^{n-1} 1/i}$ 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 $\alpha = \mathrm{Ln}(x_{ct}/y_{ct})/(x_{ag}/y_{ag})$ 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_{at} 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 X^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 $\sqrt{(\mathrm{ts}/(\mathrm{ts}+\mathrm{tv}))}$ 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 (Felstenstein, 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