International Congress Series 1261 (2004) 109–111

www.ics-elsevier.com

Is selection at mtDNA really a major concern?

L. Pereira[a,*], A. Amorim[a,b]

[a] *IPATIMUP (Instituto de Patologia e Imunologia Molecular da Universidade do Porto), Population Genetics, Rua Dr. Roberto Frias s/n 4200-465 Porto, Portugal*
[b] *Faculdade de Ciências da Universidade do Porto, Portugal*

Abstract. Evidence for selection on human mtDNA has been obtained from the results of some widely used tests: (a) higher replacement (*R*) to synonymous (*S*) substitution ratios in human genes ND2, CO1, ATP6, CO3 and CytB (and overall) than expected by human/chimpanzee comparisons; (b) distributions of absolute ($K/(K+S)$) and corrected (for coding degeneracy; Ka/(Ka + Ks)) frequencies for R are heterogeneous for human mtDNA genes (higher for ATP6, ATP8 and CytB); and (c) no correlation between gene length-*R* (in opposition to gene length-*S*). However, a mere 8.3% of sequences show $R/(R+S) \geq 0.5$ and, furthermore, *R*/*S* ratio is strongly correlated with haplogroup classification based on HVR variability. Demographic history seems therefore enough to explain the observed diversity in human mtDNA coding region, with older haplogroups showing lower *R*/*S* ratios than younger ones.

Keywords: mtDNA; Selection; Replacement substitution (*R*); Synonymous substitution (*S*)

1. Introduction

Some mtDNA genetic features have caused major general concerns: mutation rates, heteroplasmy, recombination and selection. Concerning the last one, the finding of an excess *R* inside human mtDNA genes [1], the association of certain mtDNA lineages with human longevity [2] or with male fertility [3], seemed to support the inference that natural selection (such as driven by climatic conditions [4]) has shaped regional mtDNA variation in humans. In most of these works, the ratios *R*/*S* within haplogroups were analysed. Since haplogroups are just collections of sequences that can vary profoundly but are assumed to share a common origin, we thought that, since individual sequences are indeed the true units of selection, the *R*/*S* analysis should be therefore performed using the individual sequences' information. Then, if selection forces had really played a major role in shaping human mtDNA diversity, we should expect that *R*/*S* values would be uncorrelated with haplogroup classification (mainly based upon control region variation). The aim of this paper was just to test this expectation.

* Corresponding author. Tel.: +351-22-5570700; fax: +351-22-5570799.
E-mail address: lpereira@ipatimup.pt (L. Pereira).

0531-5131/
doi:10.1016/S0531-5131(03)01838-7

2. Materials and methods

We analysed the complete mtDNA sequences of 613 humans [5,6] and 3 chimpanzees (GeneBank accession numbers: D38113, NC_001643, X93335). Hand made alignment of sequences allowed the counting of *R* and *S*. Chi-squared tests and Bonferroni's correction were calculated for *R* vs. *S* contents in each gene, in humans. DnaSP [7] program was used to measure the statistical significance (*G*-test with Williams correction for continuity) of the contingence tables of McDonald–Kreitman test [8], and to calculate the total possible *R* and *S* per gene.

3. Results

McDonald–Kreitman test results showed that the ratio *R*/*S* is significantly higher in humans than what would be expected by human/chimpanzee comparisons, for the following genes: ND2 ($G=9.994$, $P=0.00157$**), CO1 ($G=16.800$, $P=0.00004$***), ATP6 ($G=9.944$, $P=0.00161$**), CO3 ($G=8.389$, $P=0.00378$**) and CytB ($G=10.146$, $P=0.00145$**). Considering the mtDNA coding region as a whole, the value was also significantly ($G=57.461$, $P=0.00000$***), the ratio of *R*/*S* being 1:5 (or 17% of *R*).

The frequency distribution of *R* ($R/(R+S)$) by gene in humans is significantly heterogeneous ($\chi^2=52.7202$, $P=0.00000$***), and the major genes responsible for this were (decreasingly): ATP6, ATP8 and CytB. We further assayed the influence of two parameters in this measure: length and degeneracy content between genes. For the first, there was a good correlation between gene length-total number of substitutions ($Y=7.9298X$, $R^2=0.8717$) and gene length-*S* ($Y=11.551X$, $R^2=0.9241$), but almost no correlation was observed for gene length-*R* ($Y=23.168X$, $R^2=0.4079$). Since the

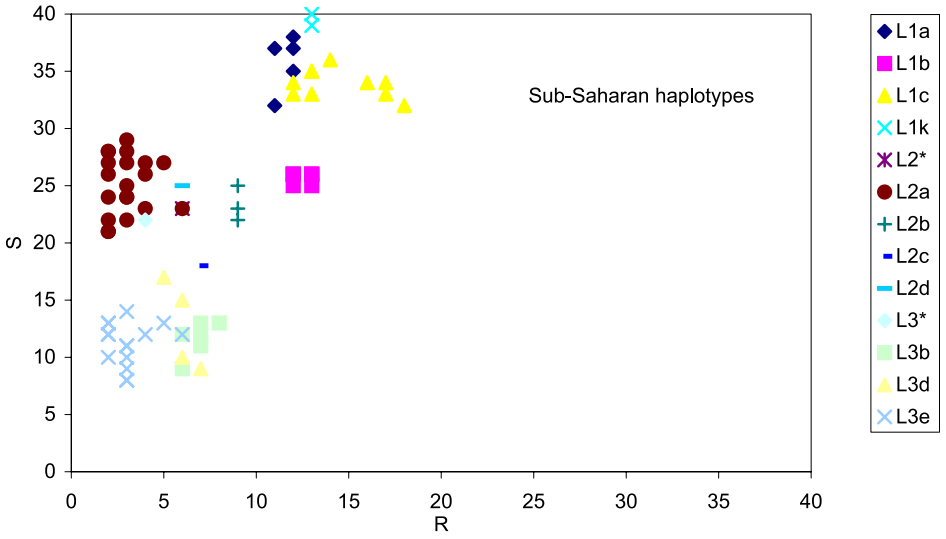

Fig. 1. Distribution of *R* vs. *S* in individual sequences from sub-Saharan haplogroups.

proportion of *S* vs. *R* sites is highly conserved between mtDNA genes, the frequency of *R* corrected for coding degeneracy (Ka/(Ka + Ks), where Ka is the number of *R* per total possible non-synonymous sites and Ks the number of *S* per total possible synonymous sites) is parallel to the absolute frequency distribution of *R*.

Despite these signs for differential accumulation of *R* between human mtDNA genes, when analysing *R* vs. *S* at individual sequences, a mere 8.3% of the sample present a frequency $R/(R+S)$ equal or higher than 0.5. The distribution of $R/(R+S)$ by haplotype is very heterogeneous, but when haplogroup classification is added (e.g. sub-Saharans in Fig. 1), we can observe that there is a high correlation between them. The proportion inside each haplogroup of sequences that showed frequencies Ka/(Ka + Ks) equal or higher than 0.5 by haplogroup were: 19% for haplogroup A1 ($n = 26$), 4% for H ($n = 217$), 12.5% for HV ($n = 8$), 11% for V ($n = 9$), 20% for U5 ($n = 25$), 67% for J2b ($n = 3$) and 93% for J1 ($n = 28$). And the distribution of $K/(K+S)$ by gene inside haplogroup *J* was highly heterogeneous.

4. Conclusions

Selection (as detected through *R*/*S* ratios) does not seem to play a major role in shaping human mtDNA diversity since the observed variation of these ratios reflects control region (HVR) defined haplogroups. Selection on coding region would necessarily act in a different way (if any) of that in the control region, and necessarily haplogroup classification should be erased/confounded when considering *R*/*S* ratios. Thus, demographic history seems enough to explain diversity in human mtDNA coding region. With time, haplogroups seem to accumulate more *S* comparatively to *R*, as it is already happening with L haplogroups. Even for haplogroup *J* (relatively young and showing no-star like phylogeny), the one with highest ratios, no specific coding region associated pattern could be discerned.

Acknowledgements

LP has a research grant (SFRH/BPD/7121/2001) from Fundação para a Ciência e a Tecnologia (FCT). IPATIMUP is supported by POCTI (QCA III). This study is included in the Project POCTI/ANT/45139/2002 financed by FCT (Eixo 2, Medida 2.3 do POCTI, QCA III).

References

[1] M.W. Nachman, et al., Genetics 142 (1996) 953–963.
[2] G. Rose, et al., Eur. J. Hum. Genet. 9 (2001) 701–707.
[3] E. Ruiz-Pesini, et al., Am. J. Hum. Genet. 67 (2000) 682–696.
[4] D. Mishmar, et al., Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 171–176.
[5] M. Ingman, et al., Nature 408 (2000) 708–713.
[6] C. Herrnstadt, et al., Am. J. Hum. Genet. 70 (2002) 1152–1171.
[7] J. Rozas, R. Rozas, Bioinformatics 15 (1999) 174–175.
[8] J.H. McDonald, M. Kreitman, Nature 351 (1991) 652–654.