Mitochondrion xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Mitochondrion

journal homepage: www.elsevier.com/locate/mito

Positive selection along the evolution of primate mitogenomes

Albert N. Menezes [a], Maria Carolina Viana [a], Carolina Furtado [a], Carlos G. Schrago [b], Héctor N. Seuánez [a,b,*]

[a] Genetics Division, Instituto Nacional de Câncer, Rio de Janeiro, Brazil
[b] Department of Genetics, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 12 March 2013
Received in revised form 16 May 2013
Accepted 3 June 2013
Available online xxxx

Keywords:
mtDNA
Primates
Positive selection

ABSTRACT

The mitochondrial genomes of four neotropical primates, *Aotus infulatus*, *Chiropotes israelita*, *Callimico goeldii* and *Callicebus lugens* were sequenced and annotated. Phylogenetic reconstructions with mitochondrial genes of other 66 primates showed a similar arrangement to a topology based on nuclear genes. Screening for positive selection identified 15 codons in 7 genes along 9 independent lineages, three with two or more genes and five in internal nodes, ruling out false positive estimates. Mitochondrial genes of the electron transport chain (ETC.) complexes evolved with high substitution rates. A study of nuclear ETC. genes might elucidate whether they co-evolved with their mitochondrial counterparts.

1. Introduction

The proposition that mitochondrial genes (mt-genes) have been subjected to non-neutral evolution has been gaining support in the last decades (Ballard and Kreitman, 1995; Dowling et al., 2008; Nachman et al., 1996; Singh and Hale, 1990). This was the case of the model experiment with six generations of modified mice carrying proofreading-deficient mitochondrial DNA (mtDNA) polymerase, in which most non-synonymous mutations in the mitochondrial genome were rapidly eliminated, indicating a strong purifying selection (Stewart et al., 2008). Moreover, positive selection, or the rapid fixation of advantageous mutations, in mtDNA has been reported for some mammalian lineages (Foote et al., 2011; Hassanin et al., 2009; Mishmar et al., 2003; Shen et al., 2010), a finding that, though rare to present, might eventually prove to be more frequent than previously assumed.

On the other hand, some studies indicating positive mtDNA selection have been questioned (Ingman and Gyllensten, 2007; Mishmar et al., 2003; Sun et al., 2007), mostly because sampling problems or low genetic distances between the taxa under study lead to false positive estimates. Similarly, the proposed diversity of mitochondrial haplotypes of some human populations presumably resulting from positive selection related to climate (Mishmar et al., 2003) has not been confirmed and was subsequently ruled out by large scale sequencing of complete human mitochondrial genomes in other populations (Gunnarsdottir et al., 2011; Schonberg et al., 2011).

Positive selection has been previously indicated for genes of the electron transport chain complexes, including mitochondrial subunits in primates (Hughes and Friedman, 2008; Wright, 1990) but this was not confirmed to present by site-by-site, codon screening (Zhao et al., 2012). In view of the scarcity of evidence in favor of non-neutral evolution along evolutionary mammalian lineages, we have analyzed mtDNA genomes of 70 primate species. We herein report strong evidence of positive selection in at least four mitochondrial genes and nine evolutionary lineages.

2. Materials and methods

2.1. Preparation of biological samples and libraries

Peripheral blood samples were obtained from one captive *Callimico goeldii* kept in the Centro de Primatologia do Rio de Janeiro (CPRJ-INEA). DNA sample of *Aotus infulatus* was provided by Dr. Artur Silva from the Genetics Department of Universidade Federal do Pará, Belém, Brazil. Liver tissue from one *Chiropotes israelita* and one *Callicebus lugens* captured in the field were provided by Dr. Cibele R. Bonvicino.

Blood samples used in this study were part of the blood samples regularly collected for checkups and control of captive animals at CPRJ-INEA. Field and sample collections were carried out following the national guidelines and provisions of IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Brazil; permanent license number 11375-1). The granting of this license by IBAMA followed approval by its Ethics Committee.

DNA from liver tissue was extracted with phenol chloroform (Sambrook et al., 1989) while DNA from fresh blood was extracted with QIAamp® DNA Mini and Blood Mini Kit (QIAGEN®). Sample

* Corresponding author at: Genetics Division, Instituto Nacional de Câncer, Rua André Cavalcanti, 37, 4th floor, 20231-050 Rio de Janeiro, R.J., Brazil. Tel.: +55 21 3207 6582; fax: +55 21 3207 6553.
E-mail addresses: albertmenezes@me.com (A.N. Menezes), mcarolviana@gmail.com (M.C. Viana), cfurtado@inca.gov.br (C. Furtado), guerra@biologia.ufrj.br (C.G. Schrago), hseuanez@inca.gov.br (H.N. Seuánez).

1567-7249/$ – see front matter © 2013 Elsevier B.V. and Mitochondria Research Society. All rights reserved.
http://dx.doi.org/10.1016/j.mito.2013.06.001

quality was checked by both electrophoresis in agarose 1% gels and NanoDrop® 1.000 Spectrophotometer (Thermo Scientific) and quantified using a Qubit® 2.0 Fluorometer (Life Technologies™). Library preparation followed the Illumina Sequencing Workflow protocols (Nextera Kit or Truseq DNA Sample Prep) for the HiSeq2000 platform. DNA containing 800–900 bp fragments (with Nextera kit) and 300–400 bp (with Truseq Kit), checked with an Agilent Bioanalyzer 2100 DNA with a high sensitivity DNA chip, was subsequently tested by qPCR using a Library Quantification Kit for library validation (KAPA — KK4824). DNA cluster generation was prepared for 3 lanes of a PE Flowcell v.3 for *C. lugens*, 4 lanes for *C. goeldii* and one lane for *C. israelita* and *A. infulatus* following the manufacturer's protocol. A 97×96 pairend run was carried out in an Illumina HiSeq2000 platform for *C. lugens* (with a minimum of 83.6% registered base calls with Q30 quality score) and a 99×93 pairend run for *C. goeldii*, *C. israelita* and *A. infulatus* (with a minimum of 87.5% registered base calls with Q30 quality score).

2.2. Data analyses

Output data was converted to Fastq files using CASAVA v1.8.2 software (Illumina®) and contigs were obtained with De Novo Assembly analysis with CLC Genomics Workbench software (CLC Bio). Contigs over 15 kb and below 18 kb were run against non-human sequences on Blast (http://blast.ncbi.nlm.nih.gov) to check for mitochondrial sequence hits. Genes were mapped using the Mitos website (http://mitos.bioinf.uni-leipzig.de/help.py) and, subsequently, checked manually with MEGA 5.1 (Tamura et al., 2011). Sequence data from other primate species were obtained from Genbank (Supplemental Table 1) and datasets containing alignments for each gene were manually constructed with MEGA 5.1 (Tamura et al., 2011). A dataset (Dat-Con) containing concatenated mt-genes was also created and the best model of evolution was inferred with ModelGenerator v. 0.85 (Keane et al., 2006). Phylogenetic reconstructions with Dat-Con were performed with PHYML 3.0 (Guindon et al., 2010) for maximum likelihood (ML) and with MrBayes 3.2.1 (Ronquist and Huelsenbeck, 2003) for Bayesian analysis (BA), with the general time reversible model (GTR) with gamma distributed rate heterogeneity with invariable sites (I+G). In MrBayes, the Markov chain Monte Carlo (MCMC) algorithm was implemented with two independent runs with four chains each. The cold chains were sampled every 100th generation until 20,000 trees were obtained. A burn-in of 2000 was used.

Analysis of codon usage was carried out by estimating: (i) The improved effective number of codons index (ENC) providing an estimate of the departure of codon usage from usage of synonymous codons. An ENC estimate of 20 corresponds to the highest codon bias resulting from usage of one codon per synonymous codon family while an estimate of 61 indicates that all codons of all synonymous codon families are equally used (Sun et al., 2013; Wright, 1990); (ii) The improved codon adaptation index (CAI2) showing the relative adaptiveness of codon usage (Sharp and Li, 1987; Xia, 2007) by comparing codon usage of a given sequence with the relative synonymous codon usage estimates of very highly expressed genes of a reference species, in this case, *Homo sapiens*. The higher the similarity between the pattern of codon usage of a given sequence and the reference pattern, the higher CAI2 estimates will result; and (iii) The codon bias index (CBI), measuring the frequency with which any coding region uses a subset of preferred codons (Bennetzen and Hall, 1982). ENC and CAI2 were calculated with DAMBE (Xia and Xie, 2001) while CBI and GC content at second (GC_2), third (GC_3) and all positions (CG) were calculated with DnaSP 5.10 (Librado and Rozas, 2009). The association between codon usage and phylogeny was tested with the Bayesian tip-significant test (Parker et al., 2008) implemented in BaTS software (http://evolve.zoo.ox.ac.uk/evolve/BaTS.html). This analysis tested against the null hypothesis postulating that codon usage estimates were randomly distributed among the tips of the phylogeny. If higher (or lower) values tended to be associated with monophyletic groups at a rate greater than pure chance, the null hypothesis was rejected at significance level equal to 0.05.

Analysis of differential selection at codon sites was performed separately for each gene with the ML topology obtained with Dat-Con, using three algorithms available in the DataMonkey server (Delport et al., 2010), namely, SLAC, REL and FEL (Kosakovsky Pond and Frost, 2005). SLAC was used for inferring positive selection based on an ancestral state reconstruction of codon sites (Kosakovsky Pond and Frost, 2005). REL was used as a likelihood-based test for estimating the distribution of dN/dS ratios and subsequently assigning each codon site to a previously estimated dN/dS class, and empirical Bayes analysis for every codon site (Kosakovsky Pond and Frost, 2005). Finally, FEL is also likelihood-based for obtaining branch lengths and substitution rates, but the distribution of dN/dS ratios is not estimated and a likelihood ratio test is used site by site (Kosakovsky Pond and Frost, 2005). Simulation studies (Kosakovsky Pond and Frost, 2005) reported SLAC as the most conservative test, with REL and FEL showing a higher sensitivity for inferring positively selected codons. The codon model for each gene was also estimated with the DataMonkey server. Branch Site REL (Kosakovsky Pond et al., 2011) analysis was also run in DataMonkey to estimate branch-specific episodes of positive selection. A dataset containing only one species per genus was composed for verifying the robustness of findings and avoiding small tree branches that would increase the rate of false positive estimates.

3. Results and discussion

The mitochondrial genome of the species sequenced in our laboratory ranged from 16,523 bp to 16,689 bp (Supplemental Table 1). On average, coverage of mitochondrial genomes was 3307× for *C. lugens*, 269× for *C. goeldii*, 777× for *A. infulatus* and 264× for *C. israelita*. Sequence and annotation data were deposited in Genbank with accession numbers KC592390–KC592393. Q3

Bayesian and ML analyses produced an identical topology (Fig. 1). The topology herein described differed from a previous one based on nuclear genes (Perelman et al., 2011) although most differences occurred at low supported nodes or at nodes preceded by short branches. Our topology showed one incongruent grouping of taxa belonging to two different families, as was the case of *Perodicticus* (a member of the Lorisidae) grouping with the branch leading to *Galago* and *Otolemur* (members of the Galagidae) rather than with other Lorisidae (*Loris* and *Nycticebus*). Similarly, *Lepilemur* (family Lepilemuridae) was more basal with respect to *Propithecus* (family Indriidae). Within the neotropical clade, *Aotus* grouped with *Saimiri*/*Cebus* rather than with the callitrichids and, in the Hylobatidae, *Hylobates* grouped with *Symphalangus* rather than with *Nomascus*. In the Cercopithecidae, *Trachypithecus* grouped with *Presbytis* rather than with *Semnopithecus* and *Pygathrix* grouped with *Rhinopithecus* rather than *Nasalis*.

Differences between our mtDNA tree and the nuclear gene tree (Perelman et al., 2011) might have resulted from the different number of taxa included in analyses, changes in mutation rates, or selective forces. Interestingly, the most discordant arrangements were

Fig. 1. Maximum likelihood topology of primate mt-genes. Bold branches (numbered) represent lineages under positive selection. All nodes were strongly supported (LRT = 1) except for nodes shown in black circles (LRT between 1 and 0.9) and white circles (below 0.9). Codon Usage statistics (see Supplemental Table 2) relationship to mt DNA phylogeny. Statistics include effective number of codons (ENC), codon adaptation index (CAI2), codon bias index (CBI) and GC content at second (G+C2), third (G+C3) and all positions (G+Cc). ENC values were divided by 100 to fit the graph scale.

Please cite this article as: Menezes, A.N., et al., Positive selection along the evolution of primate mitogenomes, Mitochondrion (2013), http://dx.doi.org/10.1016/j.mito.2013.06.001