Comparative sequence analysis of the non-protein-coding mitochondrial DNA of inbred rat strains.

Abhyankar, Avinash; Park, Hee-Bok; Tonolo, Giancarlo; Luthman, Holger

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0008148

2009

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Comparative Sequence Analysis of the Non-Protein-Coding Mitochondrial DNA of Inbred Rat Strains

Avinash Abhyankar1*, Hee-Bok Park1, Giancarlo Tonolo1,2, Holger Luthman1

1 Medical Genetics Unit, Department of Clinical Sciences-Malmö, Lund University, Malmö, Sweden, 2 Diabetology Unit, Azienda USL 2 Olbia, Olbia, Italy

Abstract

The proper function of mammalian mitochondria necessitates a coordinated expression of both nuclear and mitochondrial genes, most likely due to the co-evolution of nuclear and mitochondrial genomes. The non-protein coding regions of mitochondrial DNA (mtDNA) including the D-loop, tRNA and rRNA genes form a major component of this regulated expression unit. Here we present comparative analyses of the non-protein-coding regions from 27 Rattus norvegicus mtDNA sequences. There were two variable positions in 12S rRNA, 20 in 16S rRNA, eight within the tRNA genes and 13 in the D-loop. Z-test was performed to check whether the frequency of the polymorphisms was significantly different from random expectation. Only one of the three neutrality tests used demonstrated statistically significant evidence for selection in 16S rRNA and tRNA-Cys. Based on our analyses of conserved sequences, we propose that some of the variable nucleotide positions identified in 16S rRNA and tRNA-Cys, and the D-loop might be important for mitochondrial function and its regulation.


Introduction

Mitochondria are the major energy producers in eukaryotic cells. Over millions of years of coexistence and coevolution, mitochondria have lost a considerable amount of their genome to the eukaryotic nuclear DNA [1,2]. The mammalian mitochondrial DNA (mtDNA) encodes 37 genes, 13 of which form essential subunits of four mitochondrial respiratory chain complexes. The remaining genes for these complexes are encoded by the nuclear genome. Consequently, mitochondrial biogenesis, and hence function, needs an elaborate coordination of nuclear and mitochondrial gene expression [3,4]. Apart from several ultra-short intergenic non-coding regions, mtDNA possesses a large non-coding D-loop that harbors regulatory regions for transcription and replication. The D-loop regulates mitochondrial replication and transcription in accordance with the energy demands, while the mitochondrial rRNAs and tRNAs ensure fulfillment of this task. Having its own genetic code different from the nuclear genetic code, mitochondria need their own protein biosynthesis system in the form of the mitochondrial ribosome (mitoribosome) built around 12S rRNA and 16S rRNA. The mitoribosome is responsible for the biosynthesis of the 13 proteins coded by the mtDNA and translates them with the help of 22 tRNAs also encoded by mtDNA. The non-protein-coding regions of the mtDNA are indispensable for cellular energy homeostasis, and genetic variation in these regions could have metabolic and fitness consequences. Since the protein-coding and the non-protein-coding regions of mtDNA serve different purposes – function and regulation of function – the variation pattern and the evolutionary pressures are expected to be different. Furthermore, the relative significance of coding sequence variation compared to the regulatory sequence variation, from an evolutionary perspective, remains poorly understood [5]. For this reason we investigated the protein-coding and the non-protein-coding regions separately. Here we present a molecular evolutionary analysis of the RNA genes and the D-loop of the rat mitochondrial genome. Information from 27 complete Rattus norvegicus mtDNA sequences was used.

Results

Ribosomal RNA Genes

The mitoribosome is composed of a small subunit consisting of 12S rRNA and 29 proteins and a large subunit consisting of 16S rRNA and 58 proteins [6]. Comparison of the 27 rat mtDNA sequences (Table S1) revealed seven variable positions in 12S rRNA, five of them unique to the wild rats (Table S2). Excluding the five variant positions unique to the wild rats, only positions 935 and 942 were considered for further analysis. None of these two variable positions alter the predicted 12S rRNA secondary structure or the free energy estimates. Mapping of these two sites on the consensus secondary structure for mammalian mitochondrial 12S rRNA showed that they are located in the 3’ minor domain [7]. However, we could not find any conservation at these two positions when compared to nine different mammalian species (data not shown). In 16S rRNA there were 23 variable positions, 20 of those were found among the inbred strains, while three variable positions were unique to the wild rats (Table S2). Within 16S rRNA, we noted a poly-C tract starting at position 1131 varying between five and eleven cytosines. Six of the variant positions were located in this poly-C tract. Taken together in haplotypes, the variant positions within 16S rRNA affect the topology and free
energy estimates of the predicted secondary structures. We also assessed the conservation pattern for these variants using multiple alignments of nine different mammalian mitochondrial sequences. Of the variable positions in 16S rRNA only position 2170 was conserved among mammalian species; this C to T substitution is located in a 28-nucleotide long conserved sequence in close proximity to the L1-binding domain (Figure 1).

Transfer RNA Genes
The comparative analysis of the 22 tRNAs in mtDNA revealed a high degree of conservation. Only five of the 22 tRNAs had variable sites occurring in more than one strain (Table S2). All singletons were attributed to the wild rat sequences, except one at position 15350 that was unique to the WKY/NCrl strain. Three variable sites were observed in tRNA-Cys and two in tRNA-Pro, while tRNA-Tyr, tRNA-Aip and tRNA-Thr had one variable site each. There was a clear grouping pattern of the Wistar-derived and non-Wistar derived strains of the three variable positions in tRNA-Cys (positions 5200, 5202 and 5237). All strains originating from the Wistar rat (Table S2) shared the same allele at all these three positions indicating inheritance of an ancestral haplotype. At position 5202 the ‘Wistar’ allele was also shared by three wild rats – Wild/Cop, Wild/Tku and Wild/Mawi. A similar Wistar-specific grouping was seen for the remaining four variable tRNA genes (tRNA-Tyr, tRNA-Aip, tRNA-Thr and tRNA-Pro). To assess the structural implications of these variants, we modeled their secondary structures based on consensus secondary structures of the mammalian mitochondrial tRNAs [8]. Figure 2 shows our secondary structure models for all the five tRNA genes.

The D-Loop
The only major non-coding region of the mitochondrial DNA is the D-loop. A total of 13 variable sites were found in the D-loop of inbred strains – eleven substitutions and two insertion/deletions. We mapped the known D-loop functional sites to the rat mitochondrial sequence [9,10,11,12] (Table 1). Six substitutions were located in the termination associated sequences (TAS, ETAS), one substitution in the central block (CB), while the conserved sequence block 2 (MT-CSB2) had one insertion/deletion. Position 15460, located in ETAS1 deserves special attention since it is not only conserved between various Rattus species (Rattus, R. exulans, R. tiomanicus, R. hoffmanni, R. tanezumi, and R. nudicaudatus) but also in nine different mammalian species. The last nucleotide of the D-loop was also variable.

Tests for Selection
The Tajima’s D test and the Fu & Li’s D and F tests were performed on all the RNA genes and the D-loop to assess any deviation from neutrality. Since the results of these tests would also be influenced by population size changes, we estimated the F2 and R2 statistics but did not find support for such population changes (data not shown). We found evidence for selection in 16S rRNA and tRNA-Cys based on Tajima’s D test, whereas Fu and Li’s D and F tests did not provide any evidence for selection in the RNA genes or in the D-loop (Table 2).

Discussion
Mitochondrial DNA encodes few but essential components of the respiratory chain complexes I, III, IV and V. The two ribosomal RNAs provide a scaffold for the mitochondrial ribosomal proteins (MRPs). The mammalian mitoribosome has significantly reduced RNA content as compared to its bacterial counterpart; this reduction is accounted for by an increase in the number of MRPs [13,14]. This reduction exerts strict structural constraints on the ribosomal RNAs for efficient and accurate

Figure 1. Location of variable position 2170 in the predicted secondary structure of the mammalian mitochondrial 16S rRNA. In the enlarged L1 binding domain, position 2170 is encircled and highlighted by a red arrow. L1-BD denotes the L1 binding domain. I, II, III, IV, V and VI represent the rRNA domains, while black arrows represent the predicted tertiary interactions. Blue color represents regions predicted from comparative sequence analysis, orange colour represents predictions by Mfold software, while green colour in domain III represents alternative secondary structure predicted using Alifold software. The figure has been modified from reference [6].
doi:10.1371/journal.pone.0008148.g001
function. In bacteria and archa the ribosomal protein L1 has a dual function as a ribosomal protein binding 23S rRNA and as a translational repressor by binding mRNA [15,16]. The L1 binding domain in the mammalian mitochondrial 16S rRNA was found to be highly conserved [17]. According to our analysis, only one of the observed variable positions in the rat 16S rRNA (position 2170) is highly conserved and might be of functional importance due to its close proximity to the L1 binding domain (Figure 1).

Out of the 22 tRNA genes only five had variant positions among the 27 investigated rat sequences. According to our prediction, tRNA-Cys variant A5202G could potentially have a destabilizing effect on its secondary structure and compromise the efficiency of cystein incorporation in a growing peptide chain. Stem-loop structures in the vicinity of the L-strand origin are also important for accurate and efficient replication of mtDNA [18,19,20]. Two of the three tRNA-Cys variants (positions 5200 and 5202) are located in these loop structures. Taken together, the observed variation in the rat mitochondrial tRNA-Cys might not only affect the role as a tRNA but also affect priming of L-strand replication.

Mitochondria have an unusually high capacity for initiation of DNA replication, higher than needed for maintenance of mtDNA copy number. However, almost 95 percent of the replication events terminate prematurely resulting in formation of the 7S DNA [21]. Specific conserved short sequences have been identified that are associated with this premature termination event and are referred to as TAS and ETAS (extended TAS) elements [11,22]. It has been shown that this replication termination might regulate the mtDNA copy number [23,24]. The levels of mtDNA within a cell change

---

**Figure 2. Inferred secondary structures for (A) tRNA-Cys, (B) tRNA-Tyr, (C) tRNA-Asp, (D) tRNA-Thr and (E) tRNA-Pro.** Variable sites are highlighted in grey. The anticodon loop appears at the bottom of each model.

doi:10.1371/journal.pone.0008148.g002
according to the oxidative needs and, coupled with transcription, defines the oxidative capacity of the cell. Eight variant nucleotide positions within the D-loop were located in known functional sites. However, analysis of mitochondrial D-loop sequences from 27 mammalian species revealed a length variation in the ETAS sequences [25]. Moreover, in the human mtDNA two regions, HV1 and HV2, have been shown to be hypervariable [26,27]. HV1 in human mtDNA corresponds to positions 15284–15643 in the rat and hence, the variant nucleotide positions at these locations might not lead to major functional changes (Table 1). However, the D-loop variations located within the central block (CB) and conserved sequence block 2 (MT-CSB2) might affect mitochondrial biogenesis, since they are located outside the two hypervariable regions.

The results of the neutrality tests did not provide obvious evidence for selection in any of the non-protein-coding regions. Only Tajima’s D test provided evidence for selection in 16S rRNA and 12S rRNA-Cys. The disagreement between the tests is likely caused by the different approaches employed to identify deviation from neutrality. The two Fu & Li’s tests consider the genealogy of the sequences used to estimate the statistics, while the Tajima’s test is genealogy independent. Considering the different sensitivities of the neutrality tests to the number of variable sites, these results must be interpreted with caution. Moreover, due to high mutation rate in mtDNA, especially in the D-loop, it is not possible to account for reverse-mutations, and hence we cannot completely rule out selection with the methods used. It should also be considered that the results presented here are based on analysis of 23 inbred strains and only four sequences from wild rats. In conclusion, we have identified a few sites in the RNA genes and the D-loop that might play a role in mitochondrial biogenesis and maintenance.

### Table 1. Functional sites in the D-loop of rat mtDNA.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Description</th>
<th>Start (bp)</th>
<th>End (bp)</th>
<th>Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETAS1</td>
<td>Termination-associated sequence</td>
<td>15446</td>
<td>15503</td>
<td>1 substitution</td>
</tr>
<tr>
<td>TAS-D</td>
<td>Termination associated sequence</td>
<td>15497</td>
<td>15511</td>
<td>1 substitution</td>
</tr>
<tr>
<td>TAS-C</td>
<td>Termination associated sequence</td>
<td>15520</td>
<td>15531</td>
<td>0</td>
</tr>
<tr>
<td>TAS-B</td>
<td>Termination associated sequence</td>
<td>15541</td>
<td>15554</td>
<td>1 substitution</td>
</tr>
<tr>
<td>TAS-A</td>
<td>Termination associated sequence</td>
<td>15571</td>
<td>15584</td>
<td>1 substitution</td>
</tr>
<tr>
<td>ETAS2</td>
<td>Termination-associated sequence</td>
<td>15511</td>
<td>15572</td>
<td>2 substitutions</td>
</tr>
<tr>
<td>CB</td>
<td>Central Block</td>
<td>15673</td>
<td>15979</td>
<td>1 substitution</td>
</tr>
<tr>
<td>MT-OHR</td>
<td>H-strand origin</td>
<td>16026</td>
<td>16026</td>
<td>0</td>
</tr>
<tr>
<td>MT-CSB1</td>
<td>Conserved sequence block 1</td>
<td>16027</td>
<td>16052</td>
<td>0</td>
</tr>
<tr>
<td>MT-CSB2</td>
<td>Conserved sequence block 2</td>
<td>16083</td>
<td>16099</td>
<td>1 insertion/deletion</td>
</tr>
<tr>
<td>MT-CSB3</td>
<td>Conserved sequence block 3</td>
<td>16116</td>
<td>16133</td>
<td>0</td>
</tr>
<tr>
<td>MT-LSP</td>
<td>L-strand transcript initiation site</td>
<td>16193</td>
<td>16193</td>
<td>0</td>
</tr>
<tr>
<td>MT-TFL</td>
<td>Tfam binding site</td>
<td>16212</td>
<td>16226</td>
<td>0</td>
</tr>
<tr>
<td>MT-TFH</td>
<td>Tfam binding site</td>
<td>16267</td>
<td>16286</td>
<td>0</td>
</tr>
<tr>
<td>MT-HSP1</td>
<td>H-strand transcript initiation site</td>
<td>16298</td>
<td>16298</td>
<td>0</td>
</tr>
</tbody>
</table>

Nucleotide numbering represents position in the BN/SsNHsdMCW mtDNA sequence (Acc. No. NC_001665).

doi:10.1371/journal.pone.0008148.t001

### Table 2. Summary statistics for selection analyses in the non-protein-coding mtDNA.

<table>
<thead>
<tr>
<th>Feature</th>
<th>No. variable sites</th>
<th>No. haplotypes</th>
<th>Diversity per site</th>
<th>Tajima’s D</th>
<th>Fu &amp; Li’s D</th>
<th>Fu &amp; Li’s F</th>
</tr>
</thead>
<tbody>
<tr>
<td>12S rRNA</td>
<td>3</td>
<td>2</td>
<td>0.001</td>
<td>2.037</td>
<td>0.858</td>
<td>1.35</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>20</td>
<td>8</td>
<td>0.003</td>
<td>2.56**</td>
<td>0.765</td>
<td>1.48</td>
</tr>
<tr>
<td>tRNA-Cys</td>
<td>3</td>
<td>3</td>
<td>0.022</td>
<td>2.323*</td>
<td>0.62</td>
<td>1.00</td>
</tr>
<tr>
<td>tRNA-Tyr</td>
<td>1</td>
<td>2</td>
<td>0.007</td>
<td>1.566</td>
<td>0.642</td>
<td>1.01</td>
</tr>
<tr>
<td>tRNA-Asp</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1.566</td>
<td>0.62</td>
<td>1.00</td>
</tr>
<tr>
<td>tRNA-Thr</td>
<td>1</td>
<td>2</td>
<td>0.007</td>
<td>1.505</td>
<td>0.642</td>
<td>0.99</td>
</tr>
<tr>
<td>tRNA-Pro</td>
<td>2</td>
<td>3</td>
<td>0.003</td>
<td>-1.164</td>
<td>-1.558</td>
<td>-1.66</td>
</tr>
<tr>
<td>D-loop</td>
<td>12</td>
<td>6</td>
<td>0.003</td>
<td>1.334</td>
<td>1.105</td>
<td>1.10</td>
</tr>
</tbody>
</table>

All estimates exclude singletons observed in the sequences from wild rats.

Tajima’s D = Results of Tajima’s D test.

Fu & Li’s D = Results of Fu & Li’s D test with outgroup.

Fu & Li’s F = Results of Fu & Li’s F test with outgroup.

*=P<0.05.

**=P<0.01.

doi:10.1371/journal.pone.0008148.t002
Materials and Methods

Sequences and Analysis

Twenty seven complete Rattus norvegicus mtDNA sequences available in public databases were used, 13 of which have been sequenced in our laboratory (Table S1). The wild rats included in the study were caught at different geographical locations – Wild/Swe (Malmö, Sweden), Wild/Mcw (Milwaukee, USA), Wild/Cop (Copenhagen, Denmark), and Wild/Tku (Tokyo, Japan). Total genomic DNA extracted from rat tail was used to PCR amplify mtDNA with 32 overlapping primer pairs (Table S3). PCR products were cleaned with ExoSAP-IT [USB Corporation]. Cycle sequencing was performed using BigDye (Applied Biosystems) followed by ethanol-EDTA precipitation and separation on ABI3730 DNA Sequencer (RSKC-Malmö core facility). The sequences were processed with Phred [28,29] to assign quality values to each base call and assembled with the STADEN software [30].

Comparative Sequence Analysis of the Non-Coding Mitochondrial DNA

Multiple sequence alignments were computed using ClustalX [31] and visually inspected. DnaSP v. 4.50.3 [32] was used to estimate the nucleotide statistics (segmenting sites, haplotypes, nucleotide diversity). Since no crystal structure data are available for mammalian mitochondrial ribosomal RNAs, we referred to the predicted models for the mammalian mitoribosome [6,33,34]. Selected rat mitochondrial tRNA secondary structures were modeled on the predicted mammalian mitochondrial tRNA structures [8]. To assess the impact of variations in the RNA energy structures [35].

Tests to Identify Selection

Tajima's D test [36], Fu & Li's D and F tests with outgroup [37] were performed using DnaSP v. 4.50.3 [32]. For the Fu & Li's tests, we used the mouse reference mtDNA sequence (NC_005089) as outgroup. All these tests assess whether the DNA sequence is evolving randomly (neutrally) or by a non-random process. Non-random processes stand for either directional selection or balancing selection. However, non-random events might also be due to changes in the population size. To assess the effect of population changes we estimated the Fb [30] and R2 [39] statistics using the DnaSP program. Fu's Fs test estimates population changes by considering the number of different haplotypes in the sample, while the R2 compares the difference between the number of singleton mutations and the average number of nucleotide differences. In lieu of coalescence based permutation tests, the selection tests were assessed for sensitivity to the number of segregating sites. Both the Fu & Li's tests were considerably less sensitive to the number of segregating sites modeled on the dataset. Since the wild rat sequences were quiet divergent from the inbred population we did not include these wild rat sequences in the neutrality tests.

Author Contributions

Conceived and designed the experiments: AA GT HL. Performed the experiments: AA. Analyzed the data: AA. Contributed reagents/materials/analysis tools: AA HBP GT HL. Wrote the paper: AA HBP GT HL.

References


