

Comparative Sequence Analysis of the Non-Protein-Coding Mitochondrial DNA of Inbred Rat Strains

Avinash Abhyankar^{1*}, Hee-Bok Park¹, Giancarlo Tonolo^{1,2}, 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. 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.

Citation: Abhyankar A, Park H-B, Tonolo G, Luthman H (2009) Comparative Sequence Analysis of the Non-Protein-Coding Mitochondrial DNA of Inbred Rat Strains. PLoS ONE 4(12): e8148. doi:10.1371/journal.pone.0008148

Editor: Laszlo Orban, Temasek Life Sciences Laboratory, Singapore

Received: July 21, 2009; **Accepted:** November 8, 2009; **Published:** December 7, 2009

Copyright: © 2009 Abhyankar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from the Swedish Research Council (K2006-72X-09109-17-3 and Linné grant to the Lund University Diabetes Center), the Knut och Alice Wallenbergs Stiftelse, the Malmö University Hospital Research Foundations, the Medical Faculty of Lund University, the Novo Nordisk Foundation, the Swedish Diabetes Association, and Albert Pahlsson Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: avinash.abhyankar@med.lu.se

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

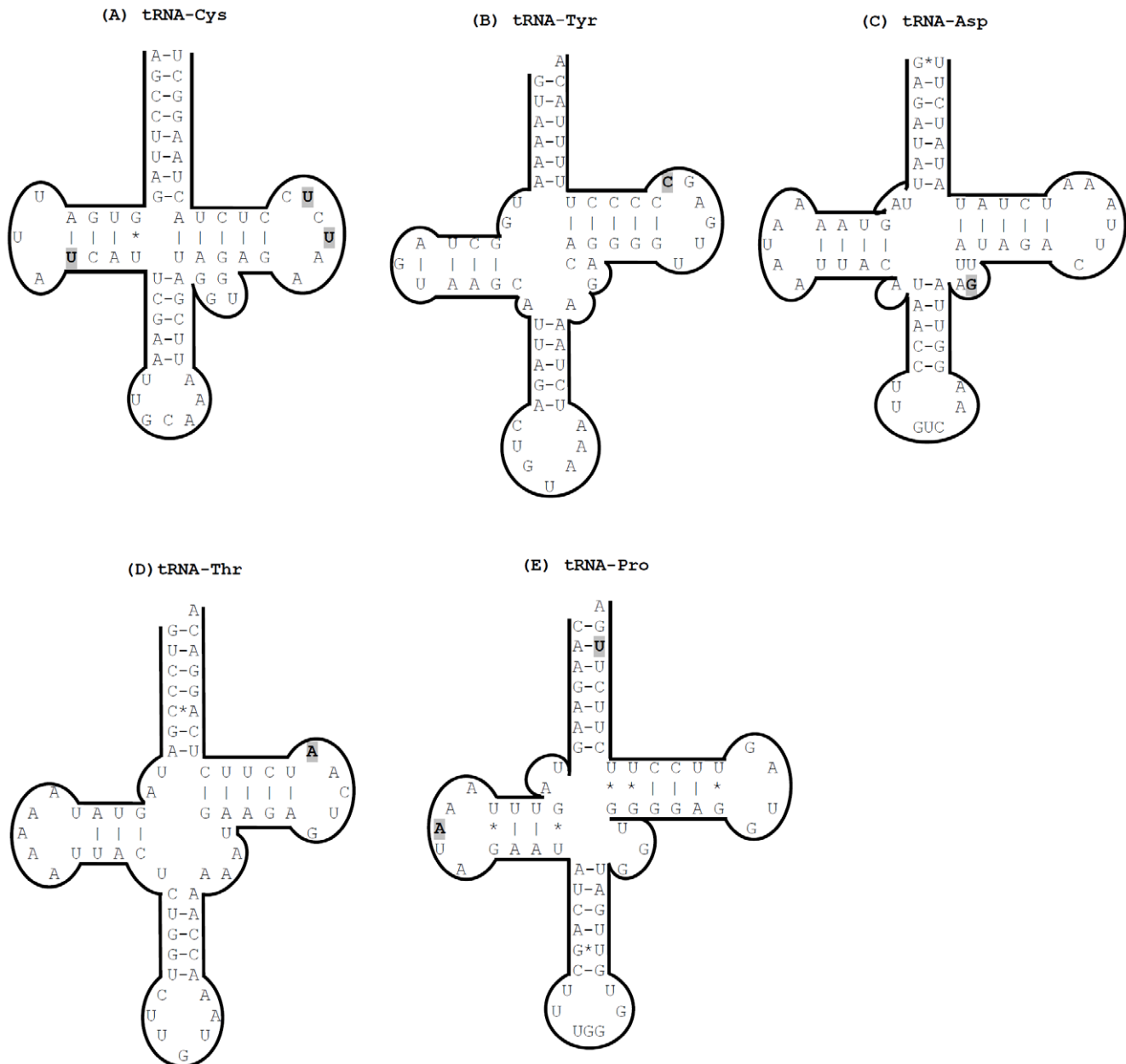


Figure 2. Inferred secondary structures for (A) *tRNA-Cys*, (B) *tRNA-Tyr*, (C) *tRNA-Asp*, (D) *tRNA-Thr* and (E) *tRNA-Pro*. Variable sites are highlighted in grey. The anticodon loop appears at the bottom of each model.
doi:10.1371/journal.pone.0008148.g002

function. In bacteria and archaea 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 cysteine 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

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

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

Nucleotide numbering represents position in the BN/SsNHsdMCW mtDNA sequence (Acc. No. NC_001665).
doi:10.1371/journal.pone.0008148.t001

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 *tRNA-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 2. Summary statistics for selection analyses in the non-protein-coding mtDNA.

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

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/Mcwi (Milwaukie, 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 (segregating 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 genes Mfold web server was used to compute the minimum free 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)

References

- Dyall SD, Brown MT, Johnson PJ (2004) Ancient invasions: from endosymbionts to organelles. *Science* 304: 253–257.
- Blier PU, Dufresne F, Burton RS (2001) Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends Genet* 17: 400–406.
- Wu Z, Puigserver P, Andersson U, Zhang C, Adelmant G, et al. (1999) Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* 98: 115–124.
- Scarpulla RC (2006) Nuclear control of respiratory gene expression in mammalian cells. *J Cell Biochem* 97: 673–683.
- Castillo-Davis CI, Hartl DL, Achaz G (2004) Cis-regulatory and protein evolution in orthologous and duplicate genes. *Genome Res* 14: 1530–1536.
- Mears JA, Sharma MR, Gutell RR, McCook AS, Richardson PE, et al. (2006) A structural model for the large subunit of the mammalian mitochondrial ribosome. *J Mol Biol* 358: 193–212.
- Springer MS, Douzery E (1996) Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. *J Mol Evol* 43: 357–373.
- Helm M, Brule H, Friede D, Giege R, Putz D, et al. (2000) Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA* 6: 1356–1379.
- Cantatore P, Daddabbo L, Fracasso F, Gadaleta MN (1995) Identification by *in organello* footprinting of protein contact sites and of single-stranded DNA sequences in the regulatory region of rat mitochondrial DNA. Protein binding sites and single-stranded DNA regions in isolated rat liver mitochondria. *J Biol Chem* 270: 25020–25027.
- Roberti M, Musicco C, Polosa PL, Milella F, Gadaleta MN, et al. (1998) Multiple protein-binding sites in the TAS-region of human and rat mitochondrial DNA. *Biochem Biophys Res Commun* 243: 36–40.
- Sbisa E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved

sequences and their functional and evolutionary implications. *Gene* 205: 125–140.

- Larizza A, Pesole G, Reyes A, Sbisa E, Saccone C (2002) Lineage specificity of the evolutionary dynamics of the mtDNA D-loop region in rodents. *J Mol Evol* 54: 145–155.
- Patel VB, Cunningham CC, Hantgan RR (2001) Physicochemical properties of rat liver mitochondrial ribosomes. *J Biol Chem* 276: 6739–6746.
- Smits P, Smeitink JAM, van den Heuvel LP, Huynen MA, Ettema TJG (2007) Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucl Acids Res* 35: 4686–4703.
- Kraft A, Lutz C, Lingenhel A, Grobner P, Pienl W (1999) Control of ribosomal protein L1 synthesis in mesophilic and thermophilic archaea. *Genetics* 152: 1363–1372.
- Mayer C, Kohrer C, Grobner P, Pienl W (1998) MvaL1 autoregulates the synthesis of the three ribosomal proteins encoded on the MvaL1 operon of the archaeon *Methanococcus vannielii* by inhibiting its own translation before or at the formation of the first peptide bond. *Mol Microbiol* 27: 455–468.
- Branlant C, Krol A, Machatt A, Ebel J-P (1981) The secondary structure of the protein L1 binding region of ribosomal 23S RNA. Homologies with putative secondary structures of the L11 mRNA and of a region of mitochondrial 16S rRNA. *Nucleic Acids Res* 9: 293–307.
- Hixson JE, Wong TW, Clayton DA (1986) Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J Biol Chem* 261: 2384–2390.
- Tapper DP, Clayton DA (1981) Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands. *J Biol Chem* 256: 5109–5115.
- Martens PA, Clayton DA (1979) Mechanism of mitochondrial DNA replication in mouse L-cells: localization and sequence of the light-strand origin of replication. *J Mol Biol* 135: 327–351.

Supporting Information

Table S1 Rat strains and mtDNA sequences

Found at: doi:10.1371/journal.pone.0008148.s001 (0.04 MB DOC)

Table S2 Variable positions in the rat non-protein-coding mtDNA

Found at: doi:10.1371/journal.pone.0008148.s002 (0.06 MB XLS)

Table S3 Primer pairs used for PCR and DNA sequence analysis

Found at: doi:10.1371/journal.pone.0008148.s003 (0.05 MB DOC)

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.

21. Bogenhagen D, Clayton DA (1978) Mechanism of mitochondrial DNA replication in mouse L-cells: kinetics of synthesis and turnover of the initiation sequence. *J Mol Biol* 119: 49–68.
22. Doda JN, Wright CT, Clayton DA (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci U S A* 78: 6116–6120.
23. Brown TA, Clayton DA (2002) Release of replication termination controls mitochondrial DNA copy number after depletion with 2',3'-dideoxycytidine. *Nucleic Acids Res* 30: 2004–2010.
24. Kai Y, Miyako K, Muta T, Umeda S, Irie T, et al. (1999) Mitochondrial DNA replication in human T lymphocytes is regulated primarily at the H-strand termination site. *Biochim Biophys Acta* 1446: 126–134.
25. Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* 205: 125–140.
26. Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, et al. (2007) An enhanced mitomap with a global mtDNA mutational phylogeny. *Nucleic Acids Res* 35: D823–828.
27. Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet* 67: 1029–1032.
28. Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res* 8: 186–194.
29. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res* 8: 175–185.
30. Staden R (1996) The staden sequence analysis package. *Mol Biotechnol* 5: 233–241.
31. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
32. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
33. Cavdar Koc E, Burkhart W, Blackburn K, Moseley A, Spremulli LL (2001) The small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of ribosomal proteins present. *J Biol Chem* 276: 19363–19374.
34. Burk A, Douzery EJP, Springer MS (2002) The secondary structure of mammalian mitochondrial 16S rRNA molecules: refinements based on a comparative phylogenetic approach. *J Mamm Evol* 9: 225–252.
35. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res* 31: 3406–3415.
36. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
37. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
38. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
39. Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19: 2092–2100.
40. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, et al. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428: 493–521.
41. Pak JW, Vang F, Johnson C, McKenzie D, Aiken JM (2005) MtDNA point mutations are associated with deletion mutations in aged rat. *Exp Gerontol* 40: 209–218.
42. Schlick NE, Jensen-Seaman MI, Orlebeke K, Kwitek AE, Jacob HJ, et al. (2006) Sequence analysis of the complete mitochondrial DNA in 10 commonly used inbred rat strains. *Am J Physiol Cell Physiol* 291: C1183–1192.
43. Nilsson MA, Gullberg A, Spotorno AE, Arnason U, Janke A (2003) Radiation of extant marsupials after the K/T boundary: evidence from complete mitochondrial genomes. *J Mol Evol* 57 Suppl 1: S3–12.
44. Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisà E, et al. (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *J Mol Evol* 28: 497–516.