Sequence and structure of the mitochondrial control region of the Cuban rodent Capromys pilorides (Rodentia: Capromyidae)

Alexander Silva1, Adriana Artiles2,3, William Suárez4, Gilberto Silva4

1Grupo de Tecnología, Empresa de Gestión del Conocimiento y la Tecnología, GECYT Calle 20 e/ 41 y 47 #4110, Playa, La Habana, Cuba
2Laboratorio de Genética Molecular, Hospital Hermanos Ameijeiras San Lázaro 701 esq. Belascoain, Centro Habana, CP 10 300, La Habana, Cuba
3Laboratorio de Sanidad Acuícola, Centro de Investigaciones Pesqueras, CIP 5ta. Avenida y 246, Barlovento, Santa Fe, Playa, CP 19100, La Habana, Cuba
4Departamento de Paleogeografía y Paleobiología, Museo Nacional de Historia Natural de Cuba, MNHN Cu Obispo 61, Plaza de Armas, Habana Vieja, CP 10100, La Habana, Cuba
E-mail: alejo@gecyt.cu

ABSTRACT

The complete mitochondrial DNA (mtDNA) control region from Capromys pilorides, an autochthon Cuban rodent, was sequenced and compared to two other species of hystricognath caviomorph rodents, in order to know patterns of variation and to explore the existence of previously described domains and other elements in rodents. The results revealed that the complete D-loop region of this species is 1336 base pairs long. Our data were compatible with the proposal of three domains [extended terminal associated sequences (ETAS), central (CD), and conserved sequence blocks (CSB)] within the control region, as well as the subsequences ETAS1, ETAS2, CSB1, CSB2, and CSB3. Likewise, a repetitive DNA region between the subsequences CSB1 and CSB2 was observed. The most conserved domain in the mitochondrial control region was the CD domain followed by ETAS and CSB domains in that order. The comparative analysis on base composition and genetic distance support the rationale of using the mitochondrial control region as a source of useful markers for population genetic studies with application to the conservation of this and other related Cuban rodent species, some of them under severe extinction risk.

Keywords: Capromys pilorides, D-loop structure, rodents

Introduction

The maternal inheritance pattern of vertebrate mitochondrial DNA, together with the presence of orthologous genes in single copies, an extremely low recombination rate [1], high mutation rates [2] and number of copies that facilitates its amplification, have made this biomolecule an essential tool for studies in genetics, taxonomy, systematics and evolution, as well as the ideal target for genetic research on biodiversity conservation. Mitochondrial DNA has been the most recurrent source of molecular markers during the last three decades [3]. Mammalian mitochondrial genomes are closed double-stranded circular molecules containing 13 protein-coding genes, 2 ribosomal RNA genes and 22 tRNA genes. Non-coding regions are circumscribed to two areas, called the control region or D-loop, involved in the replication and transcription of these molecules, and the OL region, involved in replication initiation [4]. Studies have revealed that the most rapidly evolving part of the mitochondrial genome is the control region or D-loop [5]. Research on the mammalian D-loop [6] show that can be divided into 3 domains: Extended Termination-Associated Sequences (ETAS; from the proline tRNA to the central domain), the central domain (CD), and Conserved Sequence Blocks (CSB) (from the CD to the phenylalanine...
tRNA). Comparative studies of the mitochondrial control region (MCR) of mammals have demonstrated that each domain has a different pattern of variation, as ETAS and CSB evolve rapidly, whereas CD is strongly conserved between species [6, 7].

The analysis of 25 full-length MCR sequences from 23 species of the Sciurongnathi and Hystricognathi suborders of the Rodentia order, plus one of Lagomorph order [8], suggested that the only sequence elements of this region that is conserved across all rodent species is the central domain (CD), a conserved region of this region that is conserved across all rodent species [8], suggested that the only sequence elements (66%) are extinct. There are 7 living species in Cuba (Solenodon cubanus) [12].

Capromyidae family, endemic to the Antilles, belongs to the hystricognath caviomorph rodents of the New World, and represents the only endemic family exclusively composed of island dwellers [11].

Capromyinae, one of the subfamilies grouped into the Capromyidae family, contains all living and extinct species of hutia. Five genera with 26 species are listed, classified into 127 genera and 10 families. The Pecomysidae family, endemic to the Antilles, belongs to the hystricognath caviomorph rodents of the New World, and represents the only endemic family exclusively composed of island dwellers [11].

Efforts to map world biodiversity have uncovered around 2000 species of rodents; of which, more than 40 species and 12 genera have been discovered in neotropical zones alone since 1992 [9]. This mammalian suborder, the only endemic family excluded from the ETAS and CSB1 from do-etc., is the central domain (CD), a conserved region of this region that is conserved across all rodent species [8], suggested that the only sequence elements (66%) are extinct. There are 7 living species in Cuba (Solenodon cubanus) [12].

World, and represents the only endemic family exclusively composed of island dwellers [11].

Materials and methods

Species included in the study

Table 1 contains relevant data on the species of this study, including their taxonomic classification at the family and suborder levels within the Rodentia order, as well as the GenBank accession number for the sequences used in the comparisons.

Extraction and amplification of DNA

Total DNA from two CP specimens belonging to the collection of frozen biological materials of the National Museum of Natural History of Cuba was obtained from liver samples, using the DNeasy Tissue system (Qiagen, USA) and the protocol recommended by the manufacturer. This material was used to amplify a mitochondrial DNA fragment of approximately 2.3 kb long, containing the sequences for the 3’ end of the cytochrome b gene, threonine and proline tRNA, the MCR, phenylalanine tRNA, and a portion of the 12s gene, using primers O-009 (5'-GGCTATGCTCAC-CTGCCTC-3') and O-012 (5'-GGTTGCTTTGGA- TACCCGTC-3') (Figure 1). Both primers were designed based on published sequences of mitochondrial cytochrome b and 12s genes from CP [18, 19], using the FastPCR software application [20] (Figure 1).

The amplification reactions (PCR) were set up in a volume of 50 μL, using the components of the GoTaq Core system and 2.5 U of Taq polymerase, both obtained from Promega (USA). The amplification used an initial denaturation step at 94 °C for 5 min, followed by 35 cycles of a denaturation step at 94 °C for 45 s, an annealing step at 58 °C for 45 s, and an extension step at 72 °C for 2.5 min, followed by a final single extension step at 72 °C for 10 min.

Amplification products were examined in 8% agarose gel in TBE buffer (Tris base 54 g/L, boric acid 27.5 g/L, 20 mL of 0.5 M EDTA pH 0.8), and the 2.3 kb product was purified with the Wizard SV Gel and PCR Clean-Up System from Promega (USA).

Cloning and sequencing

The purified fragments were ligated into pGEMT-easy, using the conditions and components of the pGEM-T and pGEM-T Easy Vector Systems (Promega, USA). XL-1 Blue competent cells [21], obtained from the Center for Genetic Engineering and Biotechnology of Cuba, were transformed with the ligation mixture and the positive clones were selected on LBA plates (tryptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L, pH 7.2, agar 15 g/L, ampicillin 100 μg/mL) to which 40 μL of both 100 mM IPTG and X-gal at 20 μg/mL were added.


Figure 1. Sketch of the amplified region of the mitochondrial genome of Capromys pilorides, displaying the approximate position of primers O-009 and O-012, used for initial amplification and later sequencing, and primers O-048 and O-049, used only during sequencing.
were added to facilitate the identification of recombinant clones.

Four white colonies and one blue colony obtained from the amplification of DNA from each CP specimen were submitted to colony PCR [22] to corroborate the presence of the 2.3 kb insert. Positive plasmids were purified with the Wizard Plus SV Miniprep DNA Purification System (Promega, USA), following the manufacturer’s instructions.

Plasmid DNA samples were shipped to Macrogen (South Korea) for sequencing both strands with universal primers, and also primers O-048 (5’-TCTG-GTCTCTTCTCAAGG-3’), and O-049 (5’-GAGAT-GTCCTATATTAAAGGG-3’), binding to a subsequence of the central domain (Figure 1). They were designed based on the MCR from CV, using the FastPCR software application [20].

Sequence analysis
MCR sequences from both CP specimens were aligned to their corresponding orthologs in CV and OD using Clustal X 2.0.10 [23], analyzing nucleotide composition with DAMBE v5.0.48 [24] and PAUP 4.10 beta [25]. Genetic distance values used to estimate sequence homology between the three species were calculated with MEGA 4.0 [26], using Kimura’s 2-parameter evolution model (K2P) [27].

The presence or absence of the main subsequences (ETAS, CD, and CSB) reported for mammalian [5, 6] and, specifically, rodent MCR [8], was ascertained by visual inspection, since they exhibited an acceptable level of homology. The absence or presence of the ETA2 subsequence was corroborated with a separate alignment that included rodent species Mus musculus and Rattus norvegicus which, unlike CV and OD, do have this subsequence previously identified.

Results and discussion
Sequence and characterization of the MCR from C. pilorides
Both CP specimens had an MCR that was 1336 bp long. As shown in previous studies of this region using mammals, and specifically rodents [5, 6, 8], it was also divided into a highly conserved central domain flanked by ETAS and CSB domains. There was also a repetitive DNA segment within the CSB domain (Figure 2). The presence or absence of the ETA2 subsequence was corroborated with a separate alignment that included rodent species Mus musculus and Rattus norvegicus which, unlike CV and OD, do have this subsequence previously identified.

ETAS domain
The ETAS domain is 350 bp long in CP. Two conserved subsequences have been described within this region; they are named ETAS1 and ETAS2. While ETAS2 is conserved across different mammalian species [5, 6, 28, 29], it is reportedly absent in certain rodents [8]. Using comparisons with MCR sequences from CV, OD, M. musculus, and R. norvegicus, it was possible to corroborate the presence of both subsequences in CP (Figure 1, supplementary material). Likewise, an ongoing phylogenetic study (Silva A., unpublished observations), using ETAS sequences from 20 species of hystricognath rodents, has also confirmed the presence of ETA2 subsequences. Although a previous study reported a repetitive region within this domain in rodents [8], we did not find it in CP.

Central Domain
This domain is 309 bp long in CP. Subsequences A, B and C (Figure 2), involved in the binding of cytoskeletal elements associated to the mitochondria [30], were confirmed.

CSB domain
The CSB domain was 676 bp long in CP, structured into the three canonical sequence blocks of this region (CSB1, CSB2 and CSB3). Additionally, CSB from CP has a 300-hp-long repetitive DNA region between CSB1 and CSB2 (Figure 2), in agreement with previously published data for other mammals and, especially, rodents [5, 6, 8, 29, 31, 32]. In CP the repetitive DNA region is composed of 50 hexamers, not all of which are identical (Table 2).

Comparison to CV and OD
The fundamental goal of this study was to determine the sequence and structure of the mitochondrial control region of a representative species of Cuban rodents from the Capromyinae subfamily to apply molecular genetic tools to future efforts for their conservation. It was therefore necessary to compare the MCR sequence from CP to that from phylogenetically close rodents to evaluate the feasibility of using our results as the basis of future population, inter-species and supra-species studies.

The species chosen for the comparison, CV and OD, are also New World hystricogasth rodents. OD is evolutionarily closer to CP than CV [33-36]; it is therefore expected to cluster with CP and away from CV on the basis of sequence similarity alone. Results shown on Table 3 confirm these expectations regarding both domain length and genetic distance (homology).

When comparing domain length (Table 3), however, there is an important disparity in the case of ETAS in OD. This is not a contradictory finding in itself, however, as the length of this domain is known to vary in rodents [8], although this is clearly not a conclusive structural and functional explanation. Apart from this exception, the remaining domains have similar lengths across all three species compared.

An examination of the calculated genetic distance values (Table 3) indicates that the homologies of domains ETAS and CD (Table 3) are similar to those described for other mammalian families [37, 38]. In the specific case of domain CD in the CP/OD pair, the computed genetic distances are even close to the average for genera within the same rodent family [39], although both species belong to different families (CP to Capromyidae and OD to Octodontidae). This confirms the close phylogenetic relationship of these families, which, not coincidentally, are grouped together in superfamily Octoentoidea.

CV, on the other hand, belongs to family Caviidae belonging to the Caviidae superfamily. Consequently, its genetic divergence (inverse of homology) is lower when compared to the other two species, because they are not so closely related from an evolutionary viewpoint [37].

Bioinformatic analysis

29. Matson CW, Baker RJ. DNA sequence variation in the mitochondrial control region of red-backed voles (Clethrionomy.}
Figure 2. Complete sequence of the mitochondrial control region of *Capromys pilorides*, showing the ETAS, Central and CSB domains and subsequences ETAS1, ETAS2, A, B, C, CSB1, CSB2, CSB3 and repetitive DNA.
The above results are confirmed on examining the alignments for domains ETAS (Figure 1, supplementary material), CD (Figure 2, supplementary material) and CSB (Figure 3, supplementary material; excluding the repetitive DNA region from each species) as well as Table 3.

The three alignments demonstrate a greater similarity between domain sequences from CP and OD, evidencing that the incidence of insertions and deletions between these two species is much lower to that of these two and CV.

The largest genetic distance, largest numbers of insertions and deletions, and highest proportion of insertions and deletions larger than 1 bp (parentheses in Table 3, insertions/deletions) are observed in the specific case of domain CSB, confirming the greater variability of this domain compared to ETAS and CD. This is even more evident in CV in respect to the other two species, underscoring once again the degree of evolutionary divergence between these species.

A repetitive DNA region was also observed in domain CSB for the three species, located between subsequences CSB1 and CSB2. This region, however, had inter-species differences for the number of repeats and their composition (Table 2). For instance, CV had several copies of a single repeat, whereas CP and OD were heterogeneous in repeat sequence and numbers.

The presence of repetitive DNA in mammalian CSB domains has been well documented. This repetitive region is highly variable, and may even be entirely absent [5, 6, 8]. In any case, its role within the context of the mitochondrial control region is still unknown.

The alignments for domain CSB (Figure 3, supplementary material) demonstrate the presence of sequence blocks CSB1, CSB2, and CSB3, with a high degree of sequence conservation except for small variations in CSB1 (Figure 3). These three blocks are not a conserved, general feature in rodents or in mammals, in general, since out of the 7 hystricognath with published full-length sequences of the mitochondrial control region, only those examined here have all three blocks present.

In summary, despite the availability of previous sequences from CP and other capromids published in the literature in studies of intra- and supra-species phylogenetic relationships within the Capromyidae family [18, 19, 40], this is the first published full-length D-loop sequence for a member of this taxon, and does not only enhance the knowledge on the genetic resources of our country, but it is a starting point for exploring this region in mitochondrial DNA of other Cuban capromid species.

Results indicate that the sequence and structure of the MCR in CP correspond to those published for other rodents, in complete agreement with already established phylogenetic relationships within the Hystricognathi suborder.

The strong homology between the two full-length MCR CP specimens sequences (98%), and the coherence of the values obtained from comparisons of ETAS and CD domains between species, regarding their length, genetic distance and number of insertions and deletions with those obtained for these domains in other rodents [41-43] in population genetics studies, lead to the conclusion that these sequences may be useful for population studies of Cuban capromids focused on their conservation.

### Table 2. Composition of the repetitive DNA region in the three studied species

<table>
<thead>
<tr>
<th>Species</th>
<th>Repeat Unit</th>
<th>Bases/Unit</th>
<th>Copies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavia porcellus</td>
<td>GTACCACAGACGTG</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Octodon degus</td>
<td>TACACACAGTGA</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>TACACACAGTA</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>TACCAGACGTAG</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>GTACACACAGTGA</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TACACACA</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(28)*</td>
</tr>
</tbody>
</table>

**Table 3. Comparison of Capromys pilorides (CP), Octodon degus (OD), and Cavia porcellus (CV) in domain length, sequence homology and deletions**

<table>
<thead>
<tr>
<th>Species</th>
<th>Domain length (bp)</th>
<th>Homology [%]</th>
<th>Insertions/deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETASb</td>
<td>CDc</td>
<td>CSBd</td>
<td>ETASb</td>
</tr>
<tr>
<td>CP</td>
<td>350</td>
<td>309</td>
<td>677</td>
</tr>
<tr>
<td>OD</td>
<td>266</td>
<td>310</td>
<td>679</td>
</tr>
<tr>
<td>CV</td>
<td>351</td>
<td>315</td>
<td>684</td>
</tr>
</tbody>
</table>

**Table 4. Comparison of Capromys pilorides (CP), Octodon degus (OD), and Cavia porcellus (CV) in domain length, sequence homology and deletions**

<table>
<thead>
<tr>
<th>Species</th>
<th>Domain length (bp)</th>
<th>Homology [%]</th>
<th>Insertions/deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETASb</td>
<td>CDc</td>
<td>CSBd</td>
<td>ETASb</td>
</tr>
<tr>
<td>CP</td>
<td>350</td>
<td>309</td>
<td>677</td>
</tr>
<tr>
<td>OD</td>
<td>266</td>
<td>310</td>
<td>679</td>
</tr>
<tr>
<td>CV</td>
<td>351</td>
<td>315</td>
<td>684</td>
</tr>
</tbody>
</table>

*Total number of copies.

*K2P homology. Numbers in parenthesis refer to the number of insertions and deletions longer than 1 bp.

*ETAS: Extended termination-associated sequences.

*CD: Central domain.

*CSB: Conserved sequence block.

*Excluding the repetitive DNA region.

**Figure 3. Alignment of sequence blocks CSB1, CSB2 and CSB3 from domain CSB in Capromys pilorides (CP), Octodon degus (OD) and Cavia porcellus (CV). CSB2 and CSB3 are highly conserved, whereas CSB1 differentiates between the three species. Each subsequence was colored so as to highlight differences compared to the consensus (Cons.).**


33. Huchon D, Douzery EJ. From the Old World to the New World: a molecular chronicle of the phylogeny and biogeogr-
Acknowledgements
The authors wish to thank the direction of the Molecular Genetics Laboratory of Hermanos Ameijeiras Hospital, its specialists and the hospital management for the use of their facilities and their constant support during the experimental stage of this study. This was a project funded by the World Wildlife Fund (WWF) of Canada.


Figure 1. Alignment of the ETAS domain sequences of Capromys pilorides (CP), Octodon degus (OD), Cavia porcellus (CV), Mus musculus (MM) and Rattus norvegicus (RN), showing the ETAS1 and ETAS2 sequence blocks.

Figure 2. Alignment of Central Domain sequences of Capromys pilorides (CP), Octodon degus (OD), and Cavia porcellus (CV), also showing subsequences A, B and C.
Figure 3. Alignment of CSB domain sequences of Capromys pilorides (CP), Octodon degus (OD) and Cavia porcellus (CV), excluding the repetitive DNA regions. Notice the presence of the CSB1, CSB2 and CSB3 sequence blocks, and the frequent deletions and insertions larger than single-base in CV as compared to the other species.