Mitochondrial DNA-like sequence in the nuclear genome of *Saguinus* (Callitrichinae, Primates): transfer estimation

Marcelo Vallinoto¹, Leonardo Sena¹, Iracilda Sampaio², Horacio Schneider² and Maria Paula Schneider¹

**Abstract**

Mitochondrial DNA-like sequences have been found in the nuclei of a variety of organisms. These nuclear pseudogenes can be used to estimate relative evolutionary rates of mitochondrial genes, and can be used as outgroups in phylogenetic analyses. In this study, mitochondrial sequences with pseudogene-like characteristics, including deletions and/or insertions and stop codons, were found in tamarins (*Saguinus* spp., Callitrichinae, Primates). Phylogenetic analysis allowed estimation of the timing of the migration of these sequences to the nuclear genome, and also permitted inferences on the phylogeny of the genus. The choice of an inadequate outgroup (*Aotus infulatus*) may have favored confounding homoplasies.

**INTRODUCTION**

Mitochondrial DNA-like sequences have been observed in the nuclear genome of several vertebrates and invertebrates (Gellissen et al., 1983; Jacobs et al., 1983; Fukuda et al., 1985; Thorness and Fox, 1990; Smith et al., 1992; Lopez et al., 1994; Arctander, 1995; Moreira, 1996; Sunnucks and Hales, 1996). When transferred to the nucleus, these sequences lose their function and become pseudogenes as a result of the different genetic code and to regulatory factors present in this organelle. Lopez et al. (1994) designated such sequences as *Numt* (nuclear mitochondrial) to distinguish them from *Cymt* (cytoplasmic mitochondrial) sequences.

When non-specific primers are used, *Numt* sequences may be preferentially amplified because of the better matches between primer and pseudogene (Smith et al., 1992; Collura and Stewart, 1995; van der Kuyl et al., 1995). As a result they may confound phylogenetic analyses. However, such pseudogenes can be used as a powerful tool to estimate the relative evolutionary rates of mitochondrial genes (Arctander, 1995; Lopez et al., 1997). As these sequences evolve more slowly than their mitochondrial counterparts, and are thus generally more similar to the ancestral sequences, they can be used as outgroups in phylogenetic analyses (Zhang and Hewitt, 1996). In addition, the timing of migration to the nuclear genome can be estimated.

The tamarins, *Saguinus* spp., are the basal members of the Callitrichinae, which also includes the genera *Callithrix*, *Leontopithecus* and *Callimico* (Primates - Platyrrhini) (Schneider et al., 1993, 1996; Almeida, 1995; Barroso et al., 1997). Schneider et al. (1996) estimated that this subfamily emerged approximately 10 million years ago (Mya). According to the classic review of Hershkovitz (1977), *Saguinus* is divided into three “sections” based on the facial pelage (hairy face, mottled face and bare face), and six species groups, with a total of 10 species (Table I). More recently, Rylands et al. (1993) recognized 12 species, including *Saguinus tripartitus* (Thorington Jr., 1988) and *Saguinus Geoffroyi* (Mittermeier and Coimbra-Filho, 1982), and also suggested a new grouping of *Saguinus midas* ssp. and *Saguinus bicolor* ssp. (see Natori and Hanihara, 1990; Ferrari, 1993; Meireles et al., 1997).

Jacobs et al. (1995) and Cropp et al. (1999) sequenced the D-loop region and a small part of cytochrome *b* from the mitochondrial DNA to clarify the taxonomy of *Saguinus*. According to these authors, *Saguinus* is composed of two clusters, one including *S. fuscicolis*, *S. tripartitus* and *S. nigricolis* (called small-bodied group), and the other with the rest of the species that compose the genus (large-bodied group). This new grouping has some important differences from the classical taxonomy of this genus (Hershkovitz, 1977), some of which are supported by other authors. For example, *midas* and *bicolor* are grouped while the Colombian tamarins are a sister group.

In the present study, we sequenced 549 bp of the cytochrome *c* oxidase subunit II (COII) mitochondrial gene of *Saguinus*, and included additional sequences from *Callithrix, Callimico* and *Aotus*, in order to i) identify mitochondrial-DNA like sequences, ii) determine the timing of the migration of the COII gene to the nucleus, and iii) evaluate the phylogenetic relationships among the callitrichines.

**MATERIAL AND METHODS**

Blood samples were collected from seven tamarins representing six taxa (Table II). DNA was extracted using...
phenol extraction, followed by sodium acetate precipitation. After extraction, the samples underwent 30 cycles of the polymerase chain reaction (PCR) with the following conditions: denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min and extension at 72 °C for 1.5 min. The sequences of the primers used in PCR were

\[
\begin{align*}
\Sigma_24 - & 5' \text{ CCA TCC AGC CCA ACT AGG CTT A 3'} \\
\Sigma_25 - & 5' \text{ GGC TCA TAC TTC AAA GTC TTT G 3'}
\end{align*}
\]

The PCR product was electrophoresed in 1% agarose gels. The band corresponding to the COII gene was excised and purified using the Qiaex gel extraction protocol (Qiagen). The purified DNA was inserted into the vector PGEM \(^\text{\textregistered}\) (Promega) according to the manufacturers recommendations, and then cloned into \textit{E. coli} using a pT7Blue kit (Novagen Corp.). The clones were infected with Helper Phage M13K07 (Invitrogen) and the single-strand DNA (ssDNA) was purified by PEG-NaCl precipitation, followed by phenol-chloroform extraction and ethanol precipitation. The ssDNA was sequenced by the dideoxy chain-termination method (Sanger \textit{et al.}, 1977) using a Sequenase kit (Amersham Life Science).

The sequences were aligned using the ESEE200b sequence editor (Cabot and Beckenbach, 1989) by constructing a common alignment relative to the sequence from \textit{Aotus infulatus}, which was used as the outgroup.

Divergence was estimated with the MEGA program (Kumar \textit{et al.}, 1993), using the algorithm of Tamura and Nei (1993) which is generally employed in analyses involving mitochondrial sequences. The most parsimonious tree was obtained using the SEQBOOT, DNAPARS and CONSENSE programs of the Phylogenetic Inference Package (PHYLIP) (Felsenstein, 1993).

**RESULTS**

The nucleotide sequences of the cytochrome \(c\) oxidase subunit II (COII) mitochondrial gene in seven OTUs (operation taxonomic units) of \textit{Saguinus}, three \textit{Callithrix} (\textit{C. aurita} (CAU), \textit{C. argentata} (CAR), \textit{C. jacchus} (CJA)), one \textit{Callimico goeldii} (CGO) and one \textit{Aotus infulatus} (AI) are shown in Figure 1. The SMM, SMN, SBB1 and SOE sequences (cluster 2) presented deletions at positions 197, 239 and 384. SOE also showed an insertion at position 519. Such mutations are characteristics of Numt. On the other hand, SMY, SFM and SBB2 (cluster 1) showed no such mutations.

The OTUs from cluster 2 had no nucleotide variation at any codon position, as is generally observed in noncoding sequences (Figure 2). The inferred amino acid sequences of the two clusters were highly homogeneous,
Figure 1 - Alignment of 549 base pairs of the cytochrome c oxidase subunit II (COII) mitochondrial gene in four New World primate genera (Saguinus, Callithrix, Callimico and Aotus). Dots indicate nucleotides identical to Saguinus bicolor bicolor (SBB1), "?" indicates missing data and "-" represents an insertion or deletion. SMM = S. m. midas, SMN = S. m. niger, SBB1, SBB2 = S. b. bicolor, SOE = S. oedipus, SMM = S. m. mystax, SFM = S. f. melanoleucus, CJA = C. jacchus, CAR = C. argentata, CAU = C. aurita, CGO = Callimico goeldi, and AI = Aotus infulatus.

<table>
<thead>
<tr>
<th></th>
<th>Aotus infulatus (AI)</th>
<th>Callimico goeldi (CGO)</th>
<th>C. jacchus (CJA)</th>
<th>C. argentata (CAR)</th>
<th>Callithrix argentata (CAR)</th>
<th>Callithrix aurita (CAU)</th>
<th>Saguinus bicolor bicolor (SBB1)</th>
<th>Saguinus midas midas (SMM)</th>
<th>Saguinus mystax mystax (SMY)</th>
<th>Saguinus oedipus (SOE)</th>
<th>Saguinus midas niger (SMN)</th>
<th>S. m. midas (SMM)</th>
<th>S. m. niger (SMN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COTAT</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>CAC</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>CTT</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>TCC</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>ATT</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>TCC</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>CTC</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>CTA</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>CTC</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>T</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>T</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>T</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>T</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Continued on next page
Table I - Continued

| SBB1 | AGA | ATC | CTT | TAC | ATA | ACA | GAC | G-G | TTT | ATT | AAA | CCC | TAC | TTA | ACC | CTT | AAA | GCA |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| SMN  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SMY  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SOE  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SBB2 | C.C | T.A | T   | T   | T.A | A   | C   | T   | T   | A   | A   | G   | A   | A   | G   | A   | A   | A   | A   | A   |
| SFM  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |
| SMY  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |
| CCO  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |
| CAR  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |
| CCA  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |
| CJA  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |
| CJA  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |
| CJA  | T.C | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   | C   |

Continued on next page
although there were several differences. The OTUs from cluster 2 had two stop codons (nucleotide positions 217-219 and 502-504). The former codon and a deletion were located in membrane-spanning domains. Other important changes included a tyrosine (cluster 1) - cysteine (cluster 2) swap (301-303) in the sequence of aromatic residues in a region believed to favor electron transfer and which is highly conserved in eutherians (Adkins and Honeycutt, 1994; Adkins et al., 1996).

The genetic divergence (Tamura and Nei, 1993) among *Saguinus* species ranged from 0.37% (SBB1 and SMN) to 22.29% (SOE and SBB2). However, when the other species were included, the maximum divergence increased to 23.56% (SMM and CAR). The divergence among cluster 1 OTUs ranged from 1.31% (SMY and SFM) to 6.14% (SBB2 and SMY), and in cluster 2, from 0.37% (SBB1 and SMN) to 2.33% (SMM and SOE). Divergence rates varied little when individuals of different clusters
were compared (from 20.99% for SFM and SBB1 to 22.29% for SOE and SBB2) (Figure 4).

There was a high degree of divergence (21.77%) between the two samples of *S. bicolor bicolor* used in the present study (SBB1 and SBB2), which was higher than divergences among distinct genera, such as *Saguinus* and *Callimico* (SBB2 and CGO) (19.73%).

**DISCUSSION**

The characteristics of the sequences of cluster 2 (SOE, SBB1, SMN, and SMM), such as the marked divergence (even among samples of the same species - *S. b. bicolor* samples), equivalent rates of change in all codon positions, deletions and/or insertions, stop codons and changes in highly conserved amino acids, lead us to conclude that we had obtained mitochondrial DNA-like nuclear sequences (pseudogenes). This is why SBB1 and SBB2 appear in different clades and consequently, the high divergence (Figure 4) should reflect the time of the migration of this sequence to the nucleus. However, other theories could account for this high degree of divergence, such as heteroplasmy (two different copies of the mitochondrial genome, one containing the intact COII gene and the other with mutations).

We excluded heteroplasmy as a possible explanation for these cases, because it can be maintained in an organism only if the differences between the two molecules are substitutions (which may even be non-synonymous, as long as protein function is unaffected), but not where deletions and/or insertions have occurred. Olivo *et al.* (1983) suggested that selective pressures resulted in the survival of only one of the heteroplasmic mitochondrial DNA molecules (the one without deletions or insertions) during the early development of an organism. However, this hypothesis is plausible only for functional mitochondrial sequences. In the case of noncoding sequences, such as the D-loop region, heteroplasmy can be maintained because there will be no selective pressure on molecules with deletions and/or insertions, except in regions where heavy chain replication is initiated (see Hayasaka *et al.*, 1991).

The number of substitutions in cluster 2 OTUs was much smaller than in cluster 1 OTUs, which would be expected when mitochondrial and nuclear sequences are compared (Figure 2). In addition, in the absence of the outgroup, cluster 2 OTUs exhibited no variation in nucleotides at any codon position.

*Saguinus* pseudogenes (cluster 2) and COII sequences (cluster 1) are phylogenetically distinct (Figure 3). The divergence between *Saguinus* of different clusters is greater than that between *Saguinus* (cluster 1) and *Callithrix*. According to our most parsimonious tree, the pseudogene would have appeared prior to the emergence of the Callitrichinae, that is, before 10 Mya (according to Schneider *et al.*, 1996), and the appearance of *Saguinus*. Consequently, all species in clade 2 possess the same se-
quency that migrates to the nucleus. However, this does not mean that other Saguinus taxa, not present in this clade, do not have the same pseudogene. The PCR products were cloned; so, by chance, it is possible that one of the fragments amplified (the COII gene or its pseudogene counterpart) could have been cloned and subsequently sequenced. That is why some samples of Saguinus were sequenced for the COII gene (clade 1) and others for the COII pseudogene (clade 2).

However, the possibility that the migration occurred after the emergence of the callitrichines cannot be excluded, in which case divergent haplotypes may have existed, one of which migrated to the nucleus. This would explain the degree of divergence observed. If true, it may be possible to find haplotypes that are very similar to the pseudogenes in present-day Saguinus populations, if they have not been extinguished. The former hypothesis is, however, the most parsimonious.

The poor phylogenetic resolution of the present study may reflect the choice of Aotus as the outgroup, given that Schneider et al. (1996) estimated the emergence of the Cebidae at approximately 17.5 Mya, which might favor homoplasy between the Callitrichinae and Aotinae. The divergence rate between Saguinus (cluster 1) and Aotus (or even Callithrix and Aotus) is slightly higher than that between Saguinus (cluster 1) and Callithrix. This probably reflects saturation of the bases as a consequence of the relatively ancient divergence between the Callitrichinae and Aotinae-Cebinae. The use of the Aotus haplotype may thus favor the appearance of reverse mutations, turning potentially informative sites into homoplastic ones.

Other studies have shown that COII is a useful gene for resolving phylogenetic questions within platyrrhine genera (Anselmo, 1996; Sena, 1998; Figueiredo, 1999), but not among very divergent platyrrhine genera (von Dornum, 1997). Despite these problems, it is possible to draw two conclusions from our data. First, the close relationship between S. mystax and S. fuscicollis indicated by Hershkovitz (1977) was confirmed (cluster 1), contrary to Jacobs et al. (1995) and Cropp et al. (1999). However, more data should be added to conclude if these two forms are really close taxa or not. Second, S. midas and S. bicolor were placed in the same group (cluster 2), as suggested by Natori and Hanihara (1990), Ferrari (1993), Rylands et al. (1993), Jacobs et al. (1995), Meireles et al. (1997) and Cropp et al. (1999), although a more detailed analysis will be required before the relationship between these two forms, i.e., whether S. bicolor is a subspecies of S. midas, can be defined.

In spite of its limitations, the present comparison of mitochondrial DNA in two different evolutionary scenarios (nucleus and mitochondria) offers a useful tool for tracing the evolutionary history of COII in callitrichines.

ACKNOWLEDGMENTS

We thank Arlindo Junior for technical support, Dr. José Augusto Pereira Carneiro Muniz (Centro Nacional de Primatas) and Dr. Alcides Pissinatti (Centro de Primatologia do Rio de Janeiro) for supplying samples and Dr. Stephen Francis Ferrari for reviewing the manuscript.

RESUMO

Seqüências tipo mitocondriais têm comumente sido encontradas no genoma nuclear de diversos organismos. Quando acidentalmente incluídas em estudos de seqüências mitocondriais, diversas conclusões errôneas podem ser obtidas. No entanto, estes pseudogenes nucleares tipo mitocondriais podem ser usados para a estimativa da taxa relativa de evolução de genes mitocondriais e também como grupo externo em análises filogenéticas. No presente trabalho, seqüências mitocondriais com características do tipo de pseudogene, tais como deleções e/ou inserções e códons de parada, foram encontradas em tamarins (Saguinus spp., Callitrichinae, Primates). A análise filogenética permitiu a estimativa do tempo da migração da seqüência mitocondrial para o genoma nuclear e algumas inferências filogenéticas. A escolha de um grupo externo não adequado (Aotus infulatus) não permitiu uma reconstrução filogenética confiável da subfamília Callitrichinae. A divergência bastante antiga de Cebidae (Callitrichinae, Aotinae e Cebinae) pode ter favorecido o aparecimento de homoplasias, obscurecendo a análise.

Figure 4 - Genetic divergence calculated using the algorithm of Tamura and Nei (1993). For abbreviations, see legend to Figure 1.
REFERENCES


(Received September 21, 1998)