Marmoset Phylogenetics, Conservation Perspectives, and Evolution of the mtDNA Control Region

Claudia H. Tagliaro,*† Maria Paula C. Schneider,‡ Horacio Schneider,‡ Iracilda C. Sampaio,‡ and Michael J. Stanhope*

*Biogeography and Biochemistry, Queen's University Belfast, U.K.; †Departamento de Fisiologia and ‡Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal do Pará, Brazil

Marmosets (genus Callithrix) are a diverse group of platyrrhine primates with 13–15 purported taxa, many of them considered endangered. Morphological analyses constitute most of the basis for recognition of these forms as distinct taxa. The purpose of this study was to provide a molecular view, based on mitochondrial control region sequences, of the evolutionary history of the marmosets, concomitant with a molecular phylogenetic perspective on species diversity within the group. An additional purpose was to provide the first comparative examination of a complete New World monkey control region sequence with those of other mammals. The phylogenetic analyses provide convincing support for a split between the Atlantic forest and Amazonian marmosets, with the inclusion of the pygmy marmoset (Cebuella pygmaea) at the base of the Amazonian clade. The earliest branch of the Atlantic forest group was C. aurita. In the Amazonian group, the analyses do not support the recognition of C. humeralifer and the recently described C. mauesi as distinct taxa. They do, however, support a clear distinction between C. argentina and a strongly supported mixed clade of C. humeralifer and C. mauesi. In the Atlantic forest group, the phylogenetic tree suggests mixing between C. penicillata, C. kuhli, and possibly C. jacchus. Most of the sequence features characteristic of other mammal control regions were also evident in marmosets, with the exception that conserved sequence blocks (CSBs) 2 and 3 were not clearly identifiable. Tandem repeat units often associated with heteroplasmy in a variety of other mammals were not evident in the marmoset sequences.

Introduction

The marmosets are a diverse group of platyrrhine primates within the subfamily Callitrichinae, genus Callithrix. Depending on the authority, there are currently 13–15 distinct taxa listed as members of this genus (reviewed in Rylands, Coimbra-Filho, and Mittermeier 1993). This includes recent discoveries of two new species of marmoset, C. mauesi (Mittermeier, Schwarz, and Ayres 1992) and C. nigriceps (Ferrari and Lopes 1992), both from the state of Amazonas, in central Brazilian Amazonia. Most of the recent and current systematic discussion regarding this group has centered around whether variously recognized taxa are distinct species or subspecies (reviewed in Rylands, Coimbra-Filho, and Mittermeier 1993). There is, however, no convincing molecular genetic evidence indicating that the various forms are even distinct evolutionary entities, whether you regard them as species or subspecies. There is no published account of a molecular phylogenetic analysis of this issue at the DNA sequence level, and much of the protein electrophoretic data have been inconclusive, with only a few of the taxa represented (e.g., Meireles et al. 1992). Morphological analyses constitute most, if not all (in some cases), of the basis for recognition of these forms as distinct taxa.

Most of the recognized marmoset taxa are vulnerable or endangered, many being currently classified within Appendix 1 or 2 of the CITES (Rylands, Coimbra-Filho, and Mittermeier 1993). From a conservation perspective, it is clearly of fundamental importance to have a confident understanding of which forms represent distinct evolutionary entities, based on a reliable phylogenetic framework, before effective conservation management programs can be implemented. One approach for the establishment of such a phylogenetic framework should be through the principles and techniques of molecular evolutionary genetics. The absence of such information has been shown to result in serious conservation management mistakes, the most widely documented example being the case of the dusky seaside sparrows (Avise and Nelson 1989; Avise 1994).

In addition to the concerns related to conservation, there is a wide range of phylogenetic issues in marmosets that remain virtually unexplored at the DNA sequence level. The only widely accepted phylogenetic hypothesis is that the Atlantic forest and Amazonian marmosets represent two distinct clades; however, even this view has no published support at the DNA sequence level. Relative relationships within each of these clades are not well understood, and some species within each group are certainly. The longstanding view has been to place Cebuella, the pygmy marmoset, at the base of the marmoset clade, no more closely related to either the Amazonian or Atlantic group; however, even this idea has little convincing support.

The mitochondrial control region, in the major of taxa so far investigated, is the most rapidly evolving region of the mtDNA molecule (see, e.g., Aquadro and Greenberg 1983; Horai and Hayasaka 1990; Brow Beckenbach, and Smith 1993; Zhu et al. 1994). It is therefore a putatively informative region for addressing evolutionary relationships of closely related species and subspecies. Complete control region sequences are available from a number of hominid primates (see, e.g., Anderson et al. 1981; Foran, Hixson, and Brown 1981), but there are no complete sequences available from a...
New World monkeys. Comparisons of various mammalian control region sequences have already identified several sequence features broadly characteristic of the control region including a conserved central domain, a divergent left (L; adjacent to the tRNA\textsuperscript{Pro}) and right (R; adjacent to the tRNA\textsuperscript{Phe}) domain, the presence of termination-associated sequences (TAS elements) in the L domain, and several conserved sequence blocks (CSBs) in the R domain which have been implicated in the initiation of H strand replication (Saccone, Attimonelli, and Sbisà 1987; Saccone, Pesole, and Sbisà 1991; Gemmell et al. 1996). Since no similar comparisons have been made involving New World monkeys, it is not clear to what extent the hominid features are also characteristic of other groups of primates. For example, there is an insertion sequence present in chimpanzee, gorilla, and human (Foran, Hixson, and Brown 1988; Saccone, Pesole, and Sbisà 1991) that is not present in other mammals; however, it is not known more specifically when this feature might have arisen in the evolution of the mammalian control region or whether it may have any conserved features representing a possible functional role unique to the group possessing it. Intraspécifc and interspecific mitochondrial DNA length variants are now widely documented from a diversity of animal groups (reviewed in Rand 1993), including Japanese monkeys (Hayasaka, Ishida, and Horai 1991), often due to variation in copy number of tandemly repeated sequences in the control region. It is not known whether this is a feature also characteristic of New World monkey mtDNAs.

The purpose of this study is to provide a mtDNA view of the evolutionary history of the marmosets concomitant with a molecular phylogenetic perspective of subspecies/species diversity in the group. This latter issue employs a consideration of the congruence or lack of congruence between purported morphological taxa and a molecular phylogenetic species concept (i.e., strongly supported monophyletic groups). An additional purpose is to provide a comparative examination of a New World monkey control region sequence with those of other vertebrates and in particular other mammals.

Materials and Methods

DNA sequences of the mtDNA control region were determined by direct sequencing of PCR-amplified fragments. Primers for amplification of this region were L15174 (5'-TGAGGACAAATATCTTCTGAGGGGC-3'), located in the cytochrome \textit{b} gene, and H00651 (Kocher et al. 1989). A second PCR was performed on the resulting fragment using L15926 (Kocher et al. 1989) as an internal primer. This internal PCR eliminated any false priming products that occasionally arose in the original genomic DNA PCR. The DNA sequences were determined using dye terminator cycle sequencing reactions that were subsequently loaded on an Applied Biosystems 373A automatic sequencer, following the manufacturer's protocols. Additional sequencing primers were designed as necessary. All sequences were obtained on both strands. The scientific names of the taxa included in the analyses, as well as the geographic origins of the various individuals from each taxon, appear in table 1; their relative distributions are depicted in figure 1.

Initial sequence alignments were constructed using the Clustal algorithm within the MEGALIGN program of the DNASTAR package and were subsequently perfected by eye using the eyeball sequence editor (Cabot and Beckenbach 1989). A few gaps were evident, and they were included in the analyses (i.e., the PHYLIP DNAPARS algorithm assumes that each nucleotide gap represents a single change). The best alignment was taken as the one that yielded the lowest parsimony score.
Data were analyzed by neighbor-joining (NJ; Saitou and Nei 1987), maximum-parsimony (MP; Wagner parsimony), and maximum-likelihood (DNAML; Felsenstein 1981) methods using PHYLIP (Felsenstein 1993). The robustness of the phylogenetic hypotheses was tested by bootstrapping (Felsenstein 1985a). All bootstrap analyses of DNA sequence data involved 1,000 replications of the data. Neighbor-joining analyses of the DNA sequence data were performed using different distance calculations as input: Jukes and Cantor (1969); Kimura two-parameter (Kimura 1980) with 1.5, 2.0, 2.5, 3.0, and 10.0 transition/transversion ratios; and maximum-likelihood with the same range of transition/transversion ratios (Felsenstein 1993). Maximum-parsimony analyses of the DNA sequence data were performed with the total data unweighted, and with transversions only. Most-parsimonious trees were determined by randomizing the input order 50 times. In several instances, significance tests were performed between user-defined or constrained trees and the maximum-parsimony tree; such tests were conducted using the method proposed by Templeton (1983) and Felsenstein (1985b) available in PHYLIP (Felsenstein 1993), in which the mean and variance of step differences between trees are evaluated. Maximum-likelihood analyses of the DNA sequence data were performed with the global branch-swapping option and expected transition/transversion ratios of 1.5, 2.0, 2.5, 3.0, and 5.0. All trees were rooted at the lion tamarin, *Leontopithecus chrysomelas*.

**Results**

**Sequence Characteristics and Patterns of Substitution**

The marmoset control region ranged from 1,081 to 1,142 bp in length. Complete control region sequences were obtained from *C. penicillata* (sample 89), *C. argentata* (sample 21), *C. mauesi* (sample 10), and *Cebuella* (sample 104); these four complete sequences (presented in fig. 2) were used in all the sequence characteristic comparisons that involved the 3' quarter (most of the R domain) of the control region, because data for all other taxa were obtained solely from the remaining 75% of the D loop (i.e., L and central domains). This was because approximately 300 bp of the region adjac-
FIG. 2.—Alignment of four complete marmoset control region sequences. Car = Callithrix argentutu, Cma = Callithrix maues, Cpe = Callithrix penicillata, Cpy = Cebuellu pygmeu. The 3' ends of the tRNA{Thr} and tRNA{Pro} are indicated by ]; the 5' end of the tRNA{Phe} is indicated by [. The sequence orthologous to the human termination-associated sequence (TAS) is indicated as TAS. Other sequences homologous to the TAS are indicated with a solid line. The direct repeats flanking the "primate insertion sequence" are indicated by dots followed by >> and << for 5' and 3', respectively. The boundaries of the conserved central domain as demarcated by Saccone, Pesole, and Sbisi (1991) are indicated by \[\text{TAS}\].
cent to the 12S gene could not be unambiguously aligned with the outgroup and therefore were eliminated from the phylogenetic sequence comparisons. The resulting phylogenetic sequence alignment was 920 bp for all individuals, which included 14 bp of the 3' end of the tRNA<sup>Thr</sup> gene, the tRNA<sup>Pro</sup> gene, and 830 bp of the subsequent control region.

The nucleotide sequence composition of the marmoset control region was 33% A, 27% T, 25% C, and 15% G (average of the four complete sequences). Guanine was particularly underrepresented in the L and R domains, but exhibited a significant increase in the middle domain, accompanied by a comparable drop in adenine for this same region. Most of the sequence homology with human (outside the tRNAs) occurred in the 5' 550 bp of the marmoset control region and, in particular, between 16319 and 1 of the published human sequence (Anderson et al. 1981). In this middle 250-bp region, there were several highly conserved sequence blocks, some of them as long as 30 bp, with an overall marmoset/human sequence divergence of approximately 30%. After bp 1 in human (representing the 3' third of the control region), the human/marmoset homology dropped off considerably, with interspecific marmoset homology remaining high for another 225 bp, until the region known as CSB 1 (approximately bp 765 of the marmoset control region), after which even the marmoset sequences were considerably more divergent. The two other conserved sequence blocks, known as CSBs 2 and 3, characteristic of several other mammalian control region sequences, were not clearly evident in these marmosets. There was also no evidence of any tandem repeat motifs similar to those that have been identified in a wide range of other mammalian taxa (reviewed in Rand 1993) and which have been associated with mtDNA heteroplasmacy.

An "insertion" sequence of approximately 136 nt has been identified for some Old World primates (Saccomane, Pesole, and Sbisa 1991) within the L domain, and in humans it is flanked by the following 11 nt direct repeat: 5'-TAGTACATAAA-3'. In marmosets, both these perfect repeats can be identified in roughly orthologous position to that in human (fig. 2), and the repeats had, for the vast majority of the 26 individuals, identical sequence to that for human. The marmoset "insertion" sequence is about 156 nt long, and shows no significant homology to human with the exception of the following identical stretch of 16 nt, occurring about 30 nt downstream of the 5' direct repeat: 5'-CATGCTTACAAGCAAG-3'. The termination-associated sequence (TAS) in the L domain identified in humans as the putative signal for termination of D-loop synthesis was also present in the marmosets and had a high degree of homology to the human TAS (human: 5'-TACATAAAAACCCAAT-3'; marmoset consensus: 5'-TACATAAAAATCTCAAC-3'; differing by only two transitions and one transversion). The R domain exhibited greater interspecies sequence divergence than did the L domain (mean sequence divergence for L and R, 0.207 and 0.244, respectively), particularly so for Cebuella; this R domain also had a greater number of indels, at least in the four individuals for which complete sequence was determined.

### Table 2

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cja33</td>
<td>0.001</td>
<td>0.065</td>
<td>0.061</td>
<td>0.064</td>
<td>0.055</td>
<td>0.065</td>
<td>0.073</td>
<td>0.075</td>
<td>0.075</td>
<td>0.074</td>
<td>0.062</td>
</tr>
<tr>
<td>Cja43</td>
<td>0.064</td>
<td>0.060</td>
<td>0.062</td>
<td>0.054</td>
<td>0.064</td>
<td>0.074</td>
<td>0.076</td>
<td>0.076</td>
<td>0.075</td>
<td>0.061</td>
<td>0.061</td>
</tr>
<tr>
<td>Cku94</td>
<td>4.86</td>
<td>5.67</td>
<td>0.065</td>
<td>0.001</td>
<td>0.056</td>
<td>0.000</td>
<td>0.065</td>
<td>0.065</td>
<td>0.066</td>
<td>0.065</td>
<td>0.065</td>
</tr>
<tr>
<td>Cku95</td>
<td>4.25</td>
<td>4.86</td>
<td>3.78</td>
<td>0.064</td>
<td>0.025</td>
<td>0.065</td>
<td>0.066</td>
<td>0.070</td>
<td>0.069</td>
<td>0.069</td>
<td>0.060</td>
</tr>
<tr>
<td>Cku96</td>
<td>4.57</td>
<td>5.67</td>
<td>*</td>
<td>3.78</td>
<td>0.055</td>
<td>0.001</td>
<td>0.065</td>
<td>0.068</td>
<td>0.067</td>
<td>0.066</td>
<td>0.058</td>
</tr>
<tr>
<td>Cku112</td>
<td>4.57</td>
<td>5.33</td>
<td>4.83</td>
<td>5.33</td>
<td>4.83</td>
<td>0.066</td>
<td>0.061</td>
<td>0.064</td>
<td>0.067</td>
<td>0.063</td>
<td>0.047</td>
</tr>
<tr>
<td>Cku123</td>
<td>4.86</td>
<td>5.67</td>
<td>*</td>
<td>3.78</td>
<td>*</td>
<td>4.83</td>
<td>0.065</td>
<td>0.066</td>
<td>0.067</td>
<td>0.066</td>
<td>0.058</td>
</tr>
<tr>
<td>Cgu81</td>
<td>5.00</td>
<td>4.44</td>
<td>2.91</td>
<td>2.21</td>
<td>2.91</td>
<td>2.08</td>
<td>2.91</td>
<td>2.00</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Cgu83</td>
<td>5.00</td>
<td>4.44</td>
<td>2.91</td>
<td>2.21</td>
<td>2.91</td>
<td>2.08</td>
<td>2.91</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Cgu85</td>
<td>5.00</td>
<td>4.44</td>
<td>2.91</td>
<td>2.21</td>
<td>2.91</td>
<td>2.08</td>
<td>2.91</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*061</td>
</tr>
<tr>
<td>Cge87</td>
<td>5.00</td>
<td>4.44</td>
<td>2.91</td>
<td>2.21</td>
<td>2.91</td>
<td>2.08</td>
<td>2.91</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Cpe89</td>
<td>4.37</td>
<td>3.80</td>
<td>2.78</td>
<td>3.00</td>
<td>2.78</td>
<td>2.56</td>
<td>2.78</td>
<td>2.80</td>
<td>2.70</td>
<td>2.80</td>
<td>2.56</td>
</tr>
<tr>
<td>Cpe129</td>
<td>4.17</td>
<td>5.00</td>
<td>6.40</td>
<td>4.75</td>
<td>6.40</td>
<td>7.40</td>
<td>6.40</td>
<td>3.42</td>
<td>3.42</td>
<td>3.42</td>
<td>4.50</td>
</tr>
<tr>
<td>Cau120</td>
<td>4.92</td>
<td>4.57</td>
<td>3.19</td>
<td>4.38</td>
<td>3.19</td>
<td>4.17</td>
<td>3.19</td>
<td>2.63</td>
<td>2.63</td>
<td>2.63</td>
<td>3.60</td>
</tr>
<tr>
<td>Cau121</td>
<td>4.43</td>
<td>4.13</td>
<td>2.88</td>
<td>3.93</td>
<td>2.88</td>
<td>3.69</td>
<td>2.88</td>
<td>2.40</td>
<td>2.40</td>
<td>2.40</td>
<td>2.40</td>
</tr>
<tr>
<td>Car21</td>
<td>2.10</td>
<td>2.03</td>
<td>2.04</td>
<td>1.97</td>
<td>2.04</td>
<td>1.80</td>
<td>2.04</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
</tr>
<tr>
<td>Car23</td>
<td>2.10</td>
<td>2.03</td>
<td>2.04</td>
<td>1.97</td>
<td>2.04</td>
<td>1.80</td>
<td>2.04</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
</tr>
<tr>
<td>Car98</td>
<td>2.14</td>
<td>2.07</td>
<td>2.07</td>
<td>1.97</td>
<td>2.07</td>
<td>1.93</td>
<td>2.07</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
</tr>
<tr>
<td>Cmaa9</td>
<td>1.73</td>
<td>1.69</td>
<td>1.61</td>
<td>1.61</td>
<td>1.61</td>
<td>1.51</td>
<td>1.61</td>
<td>1.31</td>
<td>1.31</td>
<td>1.31</td>
<td>1.28</td>
</tr>
<tr>
<td>Cmaa10</td>
<td>1.73</td>
<td>1.69</td>
<td>1.61</td>
<td>1.61</td>
<td>1.61</td>
<td>1.51</td>
<td>1.61</td>
<td>1.31</td>
<td>1.31</td>
<td>1.31</td>
<td>1.28</td>
</tr>
<tr>
<td>Cmaa11</td>
<td>1.59</td>
<td>1.54</td>
<td>1.67</td>
<td>1.58</td>
<td>1.67</td>
<td>1.43</td>
<td>1.67</td>
<td>1.19</td>
<td>1.19</td>
<td>1.19</td>
<td>1.39</td>
</tr>
<tr>
<td>Chu29</td>
<td>1.66</td>
<td>1.61</td>
<td>1.73</td>
<td>1.59</td>
<td>1.73</td>
<td>1.44</td>
<td>1.73</td>
<td>1.35</td>
<td>1.35</td>
<td>1.35</td>
<td>1.49</td>
</tr>
<tr>
<td>Chu31</td>
<td>1.64</td>
<td>1.59</td>
<td>1.72</td>
<td>1.51</td>
<td>1.72</td>
<td>1.47</td>
<td>1.72</td>
<td>1.37</td>
<td>1.37</td>
<td>1.37</td>
<td>1.49</td>
</tr>
<tr>
<td>Cpy104</td>
<td>2.03</td>
<td>2.11</td>
<td>2.29</td>
<td>1.97</td>
<td>2.29</td>
<td>1.90</td>
<td>2.29</td>
<td>1.67</td>
<td>1.67</td>
<td>1.67</td>
<td>1.73</td>
</tr>
<tr>
<td>Cpy105</td>
<td>1.52</td>
<td>1.56</td>
<td>1.54</td>
<td>1.50</td>
<td>1.54</td>
<td>1.49</td>
<td>1.54</td>
<td>1.37</td>
<td>1.37</td>
<td>1.37</td>
<td>1.43</td>
</tr>
<tr>
<td>Lch108</td>
<td>1.26</td>
<td>1.29</td>
<td>1.23</td>
<td>1.24</td>
<td>1.23</td>
<td>1.25</td>
<td>1.23</td>
<td>1.13</td>
<td>1.11</td>
<td>1.11</td>
<td>1.13</td>
</tr>
</tbody>
</table>

**Note:** Jukes and Cantor distances are above the diagonal and transition/transversion ratios are below the diagonal. * refers to transition/transversion with transversions equal to 0.
In the resulting phylogenetic sequence alignment (all 26 individuals represented over 920 bp), a total of 359 positions were variable, 248 of which were phylogenetically informative. The majority (i.e., 66%) of the variable positions were located between base pairs 84 and 105 of the control region (fig. 2). A hypervariable region between base pairs 237 and 406 of this alignment contained only 2.5% of the variable positions in 15% of the total sequence, and, conversely, a highly conserved region between base pairs 126 and 137 corresponded to most of the conserved central domain. A total of 49 indels of 1-5 nt in length (the vast majority being 1-3 nt in length) were apparent, with a single longer deletion of 15 bp in position 762.

Sequence divergence amongst the various Callithrix individuals ranged from 0.000 to 0.148 (table 2). Divergence figures between Cebuella and the Amazonian Callithrix species were less than those in many of the congeneric Callithrix comparisons. Sequence divergence between the two Cebuella individuals (0.117) was greater than that between many of the purported species of marmosets. Marmoset/lion tamarin divergence ranged from 0.218 to 0.248. Several individuals within purported taxa showed no sequence divergence. Transition/transversion ratios ranged from 1.11 to 28.0 (table 2). The highest transition/transversion ratios involved comparisons within the Amazonian Callithrix species. Several conspecific comparisons exhibited only transitions. Transitions accumulated much more rapidly than transversions, up to a sequence divergence of approximately 9%-10%, after which transversions continued to accumulate to about 15%, with little or no increase in transitions. At about 20% sequence divergence, there was a further increase in the number of transitions, reflecting several changes in conserved blocks of control region sequence in comparisons involving the outgroup. Using only those positions for which it could be unambiguously determined, there was a slight bias in the transition pathways, with A to G representing about 35% of the changes, G to A 16%, C to T 21%, and T to C 28%.

Phylogenetics

All three bootstrap analyses (parsimony, transversion parsimony, and neighbor joining) exhibited a high degree of congruence in their resulting topologies and clade strength. The principal features of the consensus topology (fig. 3) included a split between the Amazonian and Atlantic forest marmosets, with Cebuella joining convincingly at the base of the Amazonian clade. Callithrix aurita was the first to branch off the Atlantic group. A large number of clades were supported by 100% (or very nearly) in all three analyses. Several of the purported morphological species formed strongly supported monophyletic groups in all three analyses, including C. aurita, C. geoffroyi, Cebuella, C. jacchus, and C. argentata. The lowest clade strength occurred in the area of the tree depicting the relative relationships between C. kuhli, C. jacchus, and C. penicillata. Morphological species that did not form monophyletic groups included C. mauesi, C. penicillata, and C. kuhli. Although C. mauesi was not monophyletic it did form a strongly supported clade with C. humeralifer (bootstrap support of 100% for all three analyses). The five
C. kuhli individuals formed two strongly supported clades interrupted by a grouping of C. jacchus and C. penicillata. Different distance-based methods and transition/transversion ratios in the NJ analyses resulted in only very trivial differences in bootstrap values.

The maximum-likelihood analyses (fig. 4) agreed very closely with the bootstrap topologies; different transition/transversion ratios had only a minimal effect on branch length with no effect on topology. Several monophyletic species groupings including C. aurita, C. geoffroyi, C. jacchus, and C. argentata exhibited very short branch lengths separating individuals within the species, but had long branch lengths separating that clade from the adjacent taxa (fig. 4).

A large number of unique synapomorphies defining various groupings were evident in a site-by-site examination of the sequence alignment. For example, 19 such positions were identified in support of a Cebuella/Amazonian marmoset clade, in which all eight Amazonian marmosets and the two pygmy marmosets shared the same nucleotide, that was different from all 18 other sequences in the alignment. A further seven positions were identified in which only 1 of the 10 Cebuella/Amazonian marmoset taxa differed, with all others sharing a unique synapomorphy. Of these 26 synapomorphies, 12 were transversions, 6 were transitions, and 8 were insertions or deletions. A tree that had a monophyletic Cebuella on their own branch of the tree coming off directly after the outgroup, followed by a split defining the respective monophilies of the Amazonian and Atlantic marmosets, added 16 steps to the MP score and was judged significantly different from that MP tree.

Several other clades were similarly well supported by unique synapomorphies. The Atlantic forest clade was supported by 14 such positions (4 transitions, 2 transversions, 8 indels), C. geoffroyi by 10 (6 transitions and 4 transversions), C. argentinata by 9 (8 transitions and 1 transversion), C. mauesi/C. humeralifer by 5 (2 transitions, 3 transversions), and C. aurita by 20 (11 transitions, 6 transversions, 3 indels).

Constraining the tree to support a monophyletic C. mauesi added six substitutions and was judged significantly different from the MP tree. A similar approach defining a monophyletic C. penicillata and C. kuhli added 12 and 1 steps respectively; the monophyletic C. penicillata tree was significantly different from the MP tree, but the C. kuhli arrangement was not.

Discussion

D-Loop Sequence Characteristics

The basic sequence features of the marmoset control region were roughly similar to those described for other mammals, with the central highly conserved domain flanked by the considerably more variable L and R domains (Saccone, Pesole, and Sbisa 1991). However,
not all of the CSBs characteristic of the R domain in several other mammals were identifiable in the complete marmoset sequences. CSB 1 was present in all marmosets, but the position and/or presence of CSBs 2 and 3 was not clearly recognizable. CSB 1 appears to be the more universally conserved sequence, present in a wide range of mammals (Saccone, Pesole, and Sbisa 1991); CSBs 2 and 3 are not clearly identifiable in cow or sheep (Saccone, Pesole, and Sbisa 1991; Wood and Phua 1996), with a hybrid CSB 2/3 suggested for these two artiodactyls (Wood and Phua 1996). CSB 3 appears to be absent from dolphin (Saccone, Pesole, and Sbisa 1991) and from platypus (Gemmill et al. 1996). All three CSBs have been suggested to act as processing signals for the enzymes involved in the generation of RNA primers for heavy-strand replication, and Bennet and Clayton (1990) have identified in human and mouse a site-specific endonuclease which recognizes CSBs 2 and 3. Sequence data of ours indicate that CSBs 2 and 3 are also not clearly recognizable in Alouatta belzebul (unpublished data), another New World monkey from a different subfamily. A relatively conserved sequence in the marmosets just upstream of CSB 1 (21 bp upstream) showed some similarity to the human CSB 2 (human: 5'-CCAAACCCCCCCTCCCCC-3'; marmoset: 5'-CCAR-----CCCGCCCCC-3'). This sequence is also conserved in A. belzebul (unpublished data). If this is the same functional domain, it places the three CSBs in a different relative order than that for other mammals: this is not, however, without precedent in other vertebrates (see, e.g., Brown, Beckenbach, and Smith 1993).

In C. argentata and C. mauesi, another region 136 bp downstream of CSB 1 showed some homology to CSB 2; however, this was not the case for Cebuella and C. penicillata, with at least part of this putative element deleted in both of these species (fig. 1). Alouatta belzebul for this same region also has a putative CSB 2 element, but again with part of it missing (i.e., A. belzebul sequence: CCCCCCTACC). Still another region, 187 bp downstream of CSB 1, showed quite close homology to CSB 2 in C. argentata and C. mauesi (Car: CAAAAACTCCCCACCC; Cma: CAAAGCCCCCCCCTCCCCC), but C. penicillata, Cebuella, and A. belzebul for this same region exhibited no such homology (fig. 1). Two regions downstream of CSB 1 (19 bp and 70 bp, respectively) showed weak homology to CSB 3, but neither was particularly conserved across marmoset species (fig. 1). We would suggest that whatever the role of CSBs 2 and 3, there are some salient features that differ in New World monkeys based on the extensive modification or, possibly, complete absence of these sequence features in the Callitrichinae and in howler monkey.

The TAS identified in humans as the putative element signaling the end of D-loop synthesis was very well conserved in marmosets and forms part of the 5' direct repeat which flanks the "primate insertion" sequence. The TACAT pentanucleotide which forms part of this TAS element as well as those of other mammals (Gemmill et al. 1996) is repeated throughout the L domain of the control region in marmosets, is conspicuously absent from the middle conserved block, and is only rarely present in the R domain. This is similar to the situation in other mammals (Saccone, Pesole, and Sbisa 1991). These sequences have been shown to be associated with stable secondary structures in a range of mammals (Wilkinson and Chapman 1991; Saccone, Pesole, and Sbisa 1991; Wood and Phua 1996; Stewart and Baker 1994) and have been implicated as recognition sites for the termination of H-strand synthesis. It is not clear whether it is solely the primary sequence of the TAS element which acts as the signal to stop synthesis or whether its association with secondary structures may also play a role (Stewart and Baker 1994). The complete TAS element, as well as the TACAT pentanucleotide repeats within these marmosets, was generally (but not always) part of the stable secondary structures that are possible in this L domain (determined using the Zuker-Stiegler algorithm in DNASIS).

The insertion sequence in the L domain typical of some Old World primates was also evident in the marmosets, defined by the same conserved direct repeats present in human. The approximately 150-bp insertion sequence did not show any significant homology to human, except for a 16-bp region just inside the 5' direct repeat. The presence of the direct repeats and of the conserved block of sequence inside the proposed pri- mate insertion sequence confirms that this insertion took place prior to the evolutionary splitting event that gave rise to New and Old World monkeys. The conserved block of sequence inside the direct repeat suggests some possible functional role, perhaps unique to primates. This insertion sequence is otherwise a hypervariable region that, along with the tRNAognitive end of the D-loop, should serve as a target area for attempts at genetic discrimination of closely related New World primates.

Similar to results reported for other animal mtDNA control regions (e.g., rainbow fishes [Zhu et al. 1994] and Drosophila [Desalle et al. 1987]), marmoset transitions accumulate much more rapidly than transversions (up to about 4% sequence divergence virtually all changes are transitions) and appear to saturate at about 9%-10% total sequence divergence after which transversions accumulate. This overall pattern has been attributed to high A+T content, and the composition of these marmoset control regions would support this view. The A+T content of the marmoset control region, at 60%, is higher than that in hominid primates, but slightly lower than figures for various other mammals, which, in turn, are lower than those reported for most other vertebrates (Saccone, Attimonelli, and Sbisa 1987; Saccone, Pesole, and Sbisa 1991; Zhu et al. 1994).

Phylogenetics and Conservation Perspectives

The various methods of analysis were highly congruent in their overall topologies. The first feature, already widely recognized, is that there was an ancient evolutionary splitting event that gave rise to the Atlantic and Amazonian marmosets. An important distinctive feature of this event, suggested by these control region data, is that the pygmy marmoset lineage arose after the Amazonian/Atlantic split and not as the marmoset sister
group, which is the widely held traditional view based on immunological data (Cronin and Sarich 1978), dental ontogeny (Byrd 1981), cytotaxonomic criteria (Boer 1974), and features of the postcranium (Ford 1980). All of these different types of data have suggested a close evolutionary relationship of pygmy marmosets with marmosets. Rosenberger and Coimbra-Filho (1984), using craniodental features, suggested that pygmy marmosets are in fact so closely related to marmosets that they should be included in the same genus. Snowdon (1993) examined vocal characteristics of callirhichids and found pygmy marmosets and C. argentata to have remarkably similar long calls and contact trills. Their IRBP gene, in agreement with our D-loop sequence, also supports Cebuella inside the Callithrix clade, although with these particular nuclear data it is not clear whether pygmy marmosets have a closer affinity with the Atlantic or the Amazonian clade (M. Goodman, personal communication). Our suggestion of a phylogenetic arrangement that has Cebuella at the base of the Amazonian clade concurs with the geographic distribution of this species, which is in the upper Amazon basin, west of the Rio Purus in Brazil (Rylands, Coimbra-Filho, and Mittermeier 1993). We would suggest that additional characteristics of the biology of pygmy marmosets be reevaluated in light of this phylogenetic proposal.

Because of the time-dependent accumulation of genetic differences in the absence of gene flow, species and subspecies will exhibit phylogenetic partitioning, which, in turn, dictates that most species and subspecies will be monophyletic, a view referred to as the "phylogenetic species concept" (Cracraft 1983). A partial solution to the problem of knowing where to draw the "monophyletic line" in any application of this concept is to have sequence data from a number of individuals from a number of different populations (ideally proximally and distally located) of each of the purported taxon, and to interpret the phylogenetic trees in light of the resulting branch lengths and bootstrap support.

In the case of the present marmoset data there are several purported taxa which do form strongly supported monophyletic clusters, which also have long branch lengths separating them from the adjacent group. These include C. aurita, C. geoffroyi, Cebuella, and C. argentata. Callithrix mauesi and C. humeralifer form a strongly supported clade, with a relatively long branch leading to that node. One of the mauesi (Cma11) is more closely related to the two C. humeralifer based both on sequence divergence and high bootstrap support, suggesting the existence of naturally occurring hybrids of humeralifer and mauesi. The distribution of C. mauesi is parapatric with C. humeralifer and Callithrix chryssoleuca (Mittermeier, Schwarz, and Ayres 1992) and occurs well within the geographic distribution for C. humeralifer described by Hershkovitz (1975). Mittermeier, Schwarz, and Ayres (1992) suggest that, among the Brazilian marmosets, C. mauesi is most closely related to C. humeralifer but that it is readily distinguishable by differences in the shape of the ear tufts and the overall dark color. Our D-loop sequence data do not corroborate this view of C. mauesi as a distinct taxon. Instead, our data suggest the possibility that C. humeralifer and C. mauesi may not be distinct evolutionary entities. We emphasize that this suggestion is a tentative one because of the still-limited sampling in terms of both loci and individuals. Also important to realize is the possibility that C. humeralifer and C. mauesi are valid species, but the gene tree does not support that view because of a polymorphism in the ancestral population.

It remains a possibility, of course, that there is a hybrid zone on the contiguous borders of the mauesi distribution and that individuals in the more internal regions of the proposed distribution may more clearly represent a distinct lineage. Our samples, however, come from the border between C. chryssoleuca and C. mauesi, not from the border with C. humeralifer. The proposed distribution for C. mauesi is west of the Rio Maués, east of the Rio Urarí and north and east of the Rio Abacaxis (Mittermeier, Schwarz, and Ayres 1992). Our samples were collected along the east bank of the Rio Abacaxis, with C. chryssoleuca occurring on the west side of the river (see fig. 1). The closest humeralifer, according to Mittermeier, Schwarz, and Ayres’ estimates of the distributions, is about 150 km east, across an area that they indicate is occupied by C. mauesi. The humeralifer we have sample for were collected around Santarem, which is about 400 km east of the Rio Abacaxis. Santarem is the point where the distributions of C. argentata and C. humeralifer become contiguous. One of our C. argentata samples (98) comes from Santarem, and it clearly groups with the other two C. argentata (21, 23) which were collected about 500 km east (see fig. 1). These sampling points concomitant with the tree topology suggest a clear genetic distinction between C. argentata and C. humeralifer, but not so for C. mauesi/C. humeralifer. We offer the hypothesis that the marmosets west of Santarem, immediately south of the Amazon River, to the River Abacaxis (and possibly further west to the inclusion of the proposed distribution for C. chryssoleuca) may in fact represent a relatively continuous genotype. Similarly, we suggest that the marmosets east of Santarem may all belong to a C. argentata genotype, in agreement with the already-proposed distribution for that species (e.g., Mittermeier, Schwarz, and Ayres 1992).

In regard to the Atlantic forest group, there are, as in the Amazonian clade, examples of purported taxa which are not monophyletic. One such example of this is C. kuhli, which forms a mixed group with C. penicillata and C. jacchus. Callithrix kuhli comes from the Atlantic forests of southern Bahia. Hershkovitz (1975) and Vivo (1991) regard the marmosets from this region as hybrids of C. penicillata and C. geoffroyi. Rylands, Coimbra-Filho, and Mittermeier (1993) disagree and regard them as a distinct species. Our trees certainly do not come down in strong support of a monophyletic C. kuhli, although their paraphyly is also only weakly supported (i.e., a monophyletic C. kuhli adds only one sub-
Molecular Phylogenetics of Marmosets 683

phyly occurs despite the fact that all five animals come from the same relatively small area (table 1 and fig. 1); the greatest distance between our sample collection sites for this species was about 100 km. We interpret these results as not providing any convincing indication that C. kuhl i should be regarded as a distinct taxon since its monophyly cannot be clearly demonstrated. If it were a hybrid form, then it would be closely related to one of the parental types, which it is; if it were clearly a distinct species, then the members would form a strongly supported monophyletic group, which they do not. Coimbra-Filho (1991) argued that the destruction of habitat over large areas of the Atlantic forest since the European discovery of Brazil in 1500, along with frequent and repeated introductions of C. jacchus and C. penicillata, has resulted in a confused picture of hybrids between these species and possibly also between C. penicillata and C. kuhl i. Our trees are certainly consistent with this view regarding these three purported species.

In contrast to the confused picture regarding the C. penicillata, C. kuhl i, C. jacchus group, C. Geoffroyi form a very strongly supported clade, suggesting that they might more confidently be regarded as a distinct evolutionary entity; however, all animals did come from the same private breeding farm, which suggests the possibility that they were simply members of the same population. Similarly convincing support was apparent for C. aurita; however, only two individuals were represented (albeit from different places), and we did not have samples of the species for which there is some argument about hybridization (C. flaviceps). The data do confidently indicate that C. aurita represents an early branch of the Atlantic forest group.

Conclusions

The early parts of the next century are going to be a critical period for the conservation of biodiversity throughout the world, particularly in richly biodiverse countries such as Brazil. Approximately one quarter of the world's living primate species are found in Brazil, of which about 45% are endemic (Mittermeier, Schwarz, and Ayres 1992). An important feature of the conservation effort is going to be the preservation of species genetic diversity, and molecular genetic data should be of considerable assistance in this regard. As Avise (1989, 1994) has pointed out, one area in which molecular evolutionary genetics can be of some assistance is in the identification of "taxonomic mistakes" with their following two possible ramifications: (1) the recognition of groups which actually exhibit very little evolutionary differentiation and (2) the lack of recognition of forms which are phylogenetically distinct. In both such cases, conservation efforts can be misdirected with regard to the protection of biological diversity. In the present example, claims of C. mauesi and C. humeralifer as distinct taxa are not corroborated by our molecular genetic data. We emphasize, however, that we regard these data as a first tentative step toward an attempt at clarification of the phylogenetic diversity within these Brazilian primates and that a larger rep-

resentation of individuals from various geographic locations is necessary before any more firm conclusions should be drawn. Similarly, our data do not support a clear taxonomic distinction between C. kuhl i, C. penicillata, and possibly C. jacchus, which, again, we regard as a tentative proposal but one that needs to be further explored because of its possibly important ramifications in regard to the future conservation management of these endangered primates.

Sequence Availability

The nucleotide sequence data reported in this paper appear in GenBank under the following accession numbers: Cja33, U86526; Cja43, U88840; Cku94, U88841; Cku95, U88842; Cku96, U88843; Cku122, U88991; Cku123, U88992; Cge81, U88993; Cge83, U88994; Cge85, U88995; Cge87, U88996; Cpe129, U88997; Cai23, U88998; Car98, U88999; Cau120, U89000; Cau121, U89001; Cma09, U89002; Cma11, U89003; Cpe89, U89004; Car21, U89005; Cma10, U89006; Chu29, U89007; Chu31, U89008; Cpy105, U89009; Cpy104, U89010; Lch108, U89011.

Acknowledgments

This research was made possible by grants from the Nuffield Foundation, the Royal Society, and Northern Ireland Development and Research to M.J.S. C.H.T. was supported by the Brazilian CNPq. We would like to thank Dr. Jose Augusto Pereira Carneiro Muniz (Centro Nacional de Primates, Belem, Pará, Brazil), Dr. Adelmar Coimbra-Filho, Dr. Alcides Pissinatti (Centro de Primatologia do Rio de Janeiro, Brazil), Arlindo Pinto de Souza Jr., Artur Silva, Milton Thiago de Melo, and Criadouro Barbuse Leal for the marmoset samples used in this paper.

LITERATURE CITED


BROWN, J. R., A. T. BECKENBACH, and M. J. SMITH. 1993. Intraspecific DNA sequence variation of the mitochondrial