Timing the Origin of New World Monkeys

Carlos G. Schrago and Claudia A. M. Russo

Laboratorio Biodiversidade Molecular, Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Brazil

The origin of New World monkeys (Infraorder Platyrrhini) has been an extensively debated issue. In this study, we analyzed mitochondrial genomes from Cebus (Platyrrhini), Homo, Hyllobates, Pan, Pongo (Hominoids), Macaca, Papio (Cercopithecoids), and Tarsius (outgroup) to investigate this matter. Two distinct methodologies were employed on mitochondrial genes to estimate divergence times: the traditional likelihood ratio test performed in ML analyses of individual and concatenated gene sequences and the recent multigene Bayesian approach. Using the Cercopithecoid-Hominoid split as calibration point (25 MYA), our results show consistently that Platyrrhines split from Catarrhines at around 35 MYA. Although the main focus of the study is New World monkey origins, we have also estimated other primate divergence times: Homo-Pan at 5–7 MYA; Pongo-(Homo)Pan at 13–16 MYA; Hyllobates-(Pongo)Homo/Pan at 15–19 MYA; and Macaca-Papio at 10–12 MYA. Our estimate for the origin of New World monkeys is in agreement with the hypothesis of a transatlantic journey from Africa to South America, as suggested by the fossil record.

Introduction

Human interest in primate evolution is obviously motivated by the fact that we belong to this mammalian order. However, despite the anthropocentric importance of the group, major events in primate evolution remain debatable (Kay, Ross, and Williams 1997). One of the main controversies involves the origin of New World Monkeys (NWM, Infraorder Platyrhini). Although a sister group relationship of NWM and Old World Anthropoids (OWA, Infraorder Catarrhini) is well recognized (Ciochon and Chiarelli 1980; Ross, Williams, and Kay 1998), it is not clear how or when this divergence took place.

Those issues are not enlightened by the fossil record either, because most available fossils fall well into modern families. Furthermore, the few old primate fossils that do exist are very fragmentary, and it is not clear which lineages they represent (Kay, Ross, and Williams 1997; Gunnell and Miller 2001). The lack of fossils to depict early moments of primate evolution opens a wide door to molecular data in clarifying the origin of New World monkeys. Several molecular studies have recently approached this matter, but there are major incongruities between their estimates (e.g., 60, 70 MYA: Arnason et al. 2000; 40 MYA: Goodman et al. 1998; 47 MYA: Kumar and Hedges 1998; 33 MYA: Nei and Glazko 2002).

Therefore, we decided to examine this issue in detail using primate mitochondrial genomes. Paralogous copies of mitochondrial genes have been found in the nuclear genome (e.g., Collura and Stewart 1995); we avoided this problem by studying whole mitochondrial genomes (e.g., Collura and Stewart 1995). Furthermore, we also chose to analyze genes individually rather than using solely a concatenated sequence approach or a multigene approach for dating the split. This was done on account of the incongruity between molecular dates and, therefore, a consistent estimate, derived from several different genes, seems to be necessary in order to settle this controversial matter.

Key words: Platyrhini, biogeography, molecular clock.

E-mail: claudia@biologia.ufrj.br.

Methods

A large number of mitochondrial genomes are available for primates. Our data set was comprised of eight species: a Platyrrhini, the white-fronted capuchin monkey (Cebus albifrons NC002763, Arnason et al. 2000), two cercopithecoids: the Barbary macaque and the hamadryas baboon (Macaca sylvanus NC002764: Arnason et al. 2000; and Papio hamadryas NC001992: Arnason, Gullberg, and Janke 1998), four hominoids (Homo sapiens NC001807: Ingman et al. 2000; Hyllobates lar NC002082: Arnason, Gullberg, and Xu 1996; Pan troglodytes NC001643 and Pongo pygmaeus NC001646: Horai et al. 1995) and an outgroup sequence (Tarsius bancanus NC002811: Schmitz, Ohme, and Zischler 2002).

The assumption of a global clock does not hold for mammals (Gissi et al. 2000) and, hence an internal (anthropoid) calibration was needed. For our data set, the most reliable calibration point was the hominoid-cercopithecoid (H/C) split. This divergence is well corroborated by fossil evidence that includes the cercopithecoids Victoriapithecus (Benefit 1993; Fleagle 1999) and Prohylobates from the early Miocene (21 MYA) and the hominoid Kamoya pithecus (Harland et al. 1990; Boschetto Brown, and McDougall 1992; Leakey, Ungar, and Walker 1995) from the late Oligocene (ca. 25 MYA). Also, a recent molecular clock study attained an estimate close to 25 MYA for this split, with a particularly small standard error (Kumar and Hedges 1998; see also Staufer et al. 2001). Hence, the adoption of a divergence at 25 MYA for hominoids and cercopithecoids appears to be appropriate as an anthropoid calibration.

After selected mitochondrial genome sequences for anthropoids (Platyrrhini and Catarrhini) and Tarsius were retrieved, amino acid sequences were independently selected from GenBank. Multiple alignments for nucleotide and amino acid were conducted online for individual genes by ClustalW (www.ebi.ac.uk/clustalw) and then manually corrected. Most ribosomal and protein coding gene sequences of the mitochondrial genome were used in this study. The few exceptions were ND4L, ATPase8, and tRNA genes, excluded on account of their small size. The nucleotide alignment for the ND6 gene was unreliable because of an unusual substitution pattern (Russo, Take-
zaki, and Nei 1996; Arnason et al. 2000; Yoder and Yang 2000), and it was removed from the analysis. For protein-coding genes, we selected first and second codon positions only. This was done because analyses using third codon positions, although theoretically more appropriate because of weaker selective pressures (Russo, Takezaki, and Nei 1995), presented huge branch lengths and large associated errors. This preliminary result indicates a high degree of saturation that severely affected the statistical significance of the estimates (results not shown) and, thus, we excluded these sites from the analysis.

Because the phylogenetic relationships between species were known, the topology in figure 1 was assumed a priori, and the analyses concerned solely branch length (i.e., divergence time) computation. The analysis of NWM-OWA divergence time was conducted with two different approaches, the likelihood ratio test (LRT: Felsenstein 1988) and the multigene Bayesian approach (Thorne and Kishino 2002).

In the first approach, individual genes and concatenated sequences were submitted to the following analysis: ML phylogenetic analyses were performed by BASEML (DNA sequences) and CODEML (amino acid sequences) in the PAML 3.12 program package (Yang 1997), by inputting the known topology (fig. 1). The model of DNA substitution of Hasegawa, Kishino, and Yano (1985) with an eight-category discrete gamma correction (HKY85 + θ) was used, because it was the best fit to the data by LRT. This model not only corrects for unequal nucleotide frequencies and transition/transversion bias but it also allows eight different rate-categories for the sites. Protein sequences were corrected with the mtREV24 matrix (Adachi and Hasegawa 1996).

All ML trees were submitted to rate constancy tests before the clock was assumed. Log-likelihoods of clock-constrained and unconstrained trees were compared with the LRT. For sequences in which the clock was not rejected, individual divergence times were averaged following the procedure of Nei, Xu, and Glazko (2001) to yield global estimates for DNA and amino acid sequences.

In the second approach, the MULTIDISTRIBUTE program package (kindly provided by J. Thorne) was used to perform the multigene Bayesian analysis to estimate divergence times (Thorne and Kishino 2002). The ESTBRANCHES program was used to compute branch lengths by inputting the topology in figure 1. For DNA sequences, the program requested parameters that were calculated by PAML. For amino acid data, the JTT substitution matrix (Jones, Taylor, and Thornton 1992) was adopted.

After branch length estimation, the MULTIDIVTIME program was used to estimate “actual time nodes” and the 95% credibility interval of the estimates. Genes were allowed to differ in evolutionary rates and the prior distribution for v (the autocorrelation parameter) was set according to the strategy described elsewhere (J. Thorne, MULTIDIVTIME manual). The method also permits entering time constraints, and four were input based on the fossil record. The first was the minimum age for the Homo-Pan divergence that was set at 6 MYA (minimum age of Sahelanthropus tchadensis—Brunet et al. 2002); the second was the minimum age for the Macaca-Papio divergence at 5 MYA (the age of the oldest fossil recognized as Macaca sp.: Fleagle 1999); the third was the minimum age for the cercopithecid-hominoid divergence at 21 MYA and, finally, the minimum age for the Platyrhini-Catarrhini divergence at 27 MYA (the age of the oldest Platyrhini, Braniella boliviana: Hoffstetter 1969).

Markov chain Monte Carlo analyses were performed to approximate posterior distributions of node times. The analyses were run for 100,000 cycles before the first sample of Markov chain was taken. Another 100 cycles were run, and the second sample was taken. This procedure was repeated until 10,000 samples were obtained every 100 cycles. The entire analysis was conducted twice to check for consistency of the posterior approximations.

**Results**

The most appealing result of this paper is the fact that all NWM-OWA estimates (table 1) were very close to classical 35–40 MYA paleontological dating (Lavocat 1980; Takai and Anaya 1996; Fleagle 2000). The time estimates also varied little when comparisons were made of individual nucleotide (35.4 ± 5.2 MYA) and amino acid (36.8 ± 4.6 MYA) averages (table 2). Rate constancy could not be assumed in 16S rRNA, ATPase 6, COII, ND2, and ND4 DNA sequences and in ATPase 6, COI, COII, and ND4 amino acid sequences. We were expecting that the clock would be rejected for COI, because a faster evolutionary rate has been reported in the branches leading to hominoids and platyrrhines for this gene (Adkins and Honeycutt 1994).

The molecular clock could not be assumed for the concatenated sequences of amino acids and nucleotides because hominoid proteins evolved significantly slower than those of other primates. Even not behaving clocklike, if the clock is assumed for the concatenated sequences,
a time estimate of 39.8 MYA is obtained for amino acids and 39.7 MYA for nucleotides. This estimate falls to 36.8 MYA and 35.4 MYA, respectively, if only amino acids and nucleotide sequences for individually clocklike genes are used (table 1). Remarkably, these concatenated estimates are very close to individual nondeviant estimates and corroborates the suggestion that molecular clock estimates are minimally affected by a few deviant sequences (Nei and Kumar 2000).

The multigene Bayesian approach estimated the NWM-OWM divergence at 32.8 MYA for DNA sequences, and the 95% credibility interval (28.1–42.3 MYA) was similar to the confidence interval calculated for the average of individual genes for DNA sequences. As expected, the multigene credibility interval was considerably narrower than individual Bayesian estimations made by the same method (see also Thorne and Kishino 2002).

Nevertheless, for amino acid sequences, the multigene Bayesian approach presented a much older estimate for the Platyrrhini-Catarrhini split, mean of 41.9 MYA, with a large 95% credibility interval of 32.4–59.7 MYA. This Bayesian estimate exceeds the remaining divergence time estimates by at least 5 myr.

One explanation for this difference in means and the larger associated error could be the use of the JTT matrix to correct amino acid sequences in the Bayesian analysis. Unfortunately, the more appropriate mtREV24 matrix (Adachi and Hasegawa 1996), used in the ML analyses, was unavailable in the MULTIDISTRIBUTE program.

In our results, we have several estimates that converge for a split around 35 MYA for Platyrrhini and Catarrhini, namely individual and concatenated nucleotide sequences; individual and concatenated amino acid sequences—both calculated by LRT—and the multigene Bayesian nucleotide analysis. In addition, this result is also consistent with those of a recent paper including mitochondrial and nuclear gene sequences analyzed with distance approaches (Glazko and Nei 2003; 33 MYA). Hence, we feel that the use of an inappropriate amino acid transition matrix might be causing the difference between nucleotide and amino acid sequences in the multigene Bayesian approach.

For the sake of comparison, we have also shown the divergence times of the other anthropoids used in this study (table 2). Among these, the Homo-Pan split is of special interest because it has been intensively studied (Sarich and Wilson 1967; Hasegawa, Kishino, and Yano 1985; Yoder and Yang 2000; Nei and Glazko 2002). Concatenated sequences rendered time estimates for this split at around 5 MYA, whereas the Bayesian approach dates the origin of hominids at 7.6 MYA (with the 95% credibility interval ranging from 6.1 to 10.2 MYA). This is slightly older than Glazko and Nei’s estimate of 6 MYA. Naturally, the allowance of time constraints in the MULTIDISTRIBUTE program, including the lower limit of 6 MYA for the Homo-Pan divergence, may account for these differences.

Our estimates for the origin of the lineages leading to Hylobates (15–19 MYA) and Pongo (13–16 MYA) also present small errors by both methods and are in agreement with previous molecular studies (Hasegawa, Kishino, and Yano 1985; Stauffer et al. 2001). These estimates are also corroborated by a recent fossil finding, Lufengpithecus chiangmuensis, recovered from Middle Miocene deposits of Thailand (Chaimanee et al. 2003). Finally, the Papio-Macaca (10–12 MYA) divergence was relatively old compared to the fossil diversity of cercopithecines, showing that the major lineages of the Cercopithecoida were already diversified by the late Miocene (Fleagle 1999).

Discussion

The origin of Platyrrhini has been one of most intriguing questions in primatology (Ciochon and Chiarelli 1980; Martin 1990; Fleagle 1999). Although internally consistent at around 35 MYA, our divergence time estimates are at odds with those obtained by Arnason and coworkers (2000), who used virtually the same mitochondrial data set, yet they estimated the Platyrrhini-Catarrhini divergence at
around 70 MYA. A possible explanation for this discrepancy is that they used a calibration point outside primates, implying a global clock assumption for mammals that is not expected to hold (Gissi et al. 2000). This twofold difference in the divergence time for NWM and OWA illustrates the importance of a reliable calibration point in molecular clock studies.

It has been advocated that mitochondrial genes should not be used for divergence time studies for reasons such as rate heterogeneity between lineages (e.g., Saccone et al. 2000). However, we feel that rate heterogeneity can be overcome if adequate calibration points and molecular clock tests are applied, even without the necessity of sophisticated statistical tools.

Most nuclear estimates show much older divergence times for this split (Takahata and Satta 1997; Kumar and Hedges 1998), but a single recent study provided a nuclear estimate that falls very close to our estimate (ca. 33 myr: Nei and Glazko 2002; see also Glazko and Nei 2003). Interestingly, in their study, Nei and Glazko explicitly tried to select orthologous sequences to estimate time, probably avoiding most problems associated with paralogy. The fact that our several mitochondrial estimates were consistent and very close to this nuclear estimate suggests that we might have reached a divergence time very close to the actual date of Platyrhini and Catarrhini split.

Several vicariant and dispersal hypotheses have been suggested to explain how and when the Platyrhini ancestor arrived on the South American continent. Naturally, the establishment of a reliable time estimate for the origin of NWM is crucial to the understanding of the early evolution of this group. In this sense, many hypotheses may be promptly discarded in light of a consistent temporal estimate for the divergence.

For instance, the vicariant scenario for the origin of Platyrhini from primitive prosimian stocks in Gondwana-land seems improbable because the separation of Africa and South America happened between 120 and 100 MYA and our divergence time dates a much younger split. Therefore, the assumption that Platyrhini ancestors dispersed to the South American island continent at about the Eocene-Oligocene boundary seems likely, but a source continent must be elected.

One dispersal hypothesis suggests that primitive Asian anthropoids could have invaded South America by an Antarctic route (Houle 1999). Even though a number of primitive fossil anthropoids have been found in Asia, the lack of primate fossils in Antarctica effectively discards this hypothesis (Kay, Ross, and Williams 1997). A second speculation, the North American origin of NWM, is also unlikely, because no anthropoids were ever found in that continent (Fleagle 1999). Finally, the morphological resemblance between platyrrhines and the African anthropoid fossils, particularly those from Fayum deposits in Egypt, justify the idea that the ancestors of Platyrhini primates probably came from Africa (Ross, Williams, and Kay 1998; Houle 1999).

Another hypothesis binds South American and African primates, but in the reverse order. According to this last hypothesis, anthropoids could have originated in South America and subsequently migrated to Africa (Szalay 1975). However, the lack of prosimians in the South American fossil record sheds doubt on the validity of this hypothesis. Furthermore, African anthropoid fossils date much older than South American fossil primates, which are already recognized as Platyrhini (Fleagle 1999).

On conclusion, if an African origin for the South American platyrrhines is admitted, the issue of how they made the journey remains to be clarified. The problem is that a transatlantic journey from Africa to South America is not an easy feat for primates. It is recognized that, in spite the overall unaltered disposition of continental land masses, several drastic climate changes marked the Eocene-Oligocene boundary (Ivany, Patterson, and Lohmann 2000). These changes also include variation in global temperatures that may have affected sea level. In this scenario, South Atlantic Ocean ridges such as the Sierra Leone Rise and the Walvis Ridge could have become exposed as islands, creating pathways that, in conjunction with favorable water and wind currents, enabled faunal migration to the isolated South America (Houle 1999).

Indeed, other mammals have also supposedly invaded the South American continent from Africa, such as New World caviomorph rodents that suddenly appeared in the South American fossil record at approximately the same time the platyrrhines did (Wyss et al. 1993). Interestingly, these mammals, as NWM do, also have a sister taxon relationship with African groups, the phiomorph rodents (Mouchaty et al. 2001). Then, the existence of a faunal connection between Africa and South America in the Eocene/Oligocene transition is further corroborated.

---

Table 2

| Origin of New World Monkeys 1623 |

| Table 2 Divergence Times of Selected Anthropoid Lineages |

<table>
<thead>
<tr>
<th>Average (clocklike)</th>
<th>Multigene Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Amino Acid</td>
</tr>
<tr>
<td>Homo-Pan</td>
<td>5.4 ± 2.2</td>
</tr>
<tr>
<td>Pongo</td>
<td>13.0 ± 2.2</td>
</tr>
<tr>
<td>Hylobates</td>
<td>15.2 ± 2.3</td>
</tr>
<tr>
<td>Macaca-Papio</td>
<td>12.6 ± 2.2</td>
</tr>
<tr>
<td>Cercopithecoid-Hominoid</td>
<td>*</td>
</tr>
<tr>
<td>Catarrhini-Platyrrhini</td>
<td>35.4 ± 5.2</td>
</tr>
</tbody>
</table>

Note.—In the Multigene columns the posterior means and the 95% credibility interval of divergence times are shown. The average was estimated with the procedure described in Nei, Xu, and Glazko (2001) for clocklike genes. An asterisk indicates that the category is not applicable, because this time was assumed as the calibration point for the method.
Acknowledgments

The authors are indebted to Antonio Solé-Cava, the associate editor, and two anonymous reviewers, for helpful comments on an earlier version of the manuscript. We would also like to thank Jeffrey Thorne for patiently aiding us with the use of the MULTIDISTRIBUT program and Pablo Nehab-Hess for helpful hints on the UNIX system. This paper is part of the Ph.D. thesis of C.G.S. at the Genetics Program at the Federal University of Rio de Janeiro. C.G.S. is supported by CAPES (Brazilian Education Ministry). The authors thank CNPq (Brazilian research council) and FAPERJ (Rio de Janeiro State’s research foundation) for research grants to C.A.M.R.

Literature Cited


Mark Springer, Associate Editor

Accepted May 14, 2003