ELSEVIER

# *Alu* insertion loci and platyrrhine primate phylogeny

David A. Ray[a], Jinchuan Xing[a], Dale J. Hedges[a], Michael A. Hall[a], Meredith E. Laborde[a],
Bridget A. Anders[a], Brittany R. White[a], Nadica Stoilova[a], Justin D. Fowlkes[a],
Kate E. Landry[a], Leona G. Chemnick[b], Oliver A. Ryder[b], Mark A. Batzer[a],*

[a] *Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Multi-scale systems,
Louisiana State University, 202 Life Sciences Bldg., Baton Rouge, LA 70803, USA*
[b] *Center for Reproduction of Endangered Species, Zoological Society of San Diego, San Diego, CA 92101, USA*

## Abstract

Short INterspersed Elements (SINEs) make very useful phylogenetic markers because the integration of a particular element at a location in the genome is irreversible and of known polarity. These attributes make analysis of SINEs as phylogenetic characters an essentially homoplasy-free affair. *Alu* elements are primate-specific SINEs that make up a large portion of the human genome and are also widespread in other primates. Using a combination wet-bench and computational approach we recovered 190 *Alu* insertions, 183 of which are specific to the genomes of nine New World primates. We used these loci to investigate branching order and have produced a cladogram that supports a sister relationship between Atelidae (spider, woolly, and howler monkeys) and Cebidae (marmosets, tamarins, and owl monkeys) and then the joining of this two family clade to Pitheciidae (titi and saki monkeys). The data support these relationships with a homoplasy index of 0.00. In this study, we report one of the largest applications of SINE elements to phylogenetic analysis to date, and the results provide a robust molecular phylogeny for platyrrhine primates.
© 2004 Elsevier Inc. All rights reserved.

*Keywords:* *Alu*; Primate; Phylogeny; Platyrrhine

## 1. Introduction

Short INterspersed Elements (SINEs) are powerful tools for systematic biologists (Hillis, 1999; Shedlock and Okada, 2000). For example, Shimamura et al. (1997) used SINEs to support the hypothesis that cetaceans (whales, dolphins, and porpoises) form a clade within Artiodactyla. Takahashi et al. (2001) also used SINEs to elucidate the relationships among the cichlid fishes of Lake Malawi. Studies of primate phylogenetics have also been successfully addressed using SINE markers. Salem et al. (2003b) used primate SINEs to produce a definitive

picture of hominid phylogeny and Schmitz et al. (2001) supported the close relationship between tarsiers and anthropoids. In each of these studies, the presence of a SINE in any particular lineage provided strong evidence to cluster members of that node, with only two cases of potential homoplasy being introduced either by lineage sorting or via interspecies hybridization.

One reason for the success of SINEs as phylogenetic and population genetic markers is that their mode of evolution is unidirectional (Hillis, 1999). This characteristic allows for a confident inference that the ancestral state is the absence of the SINE for each locus under examination. Because there is no known mechanism for the specific removal of SINEs from any genome (Batzer and Deininger, 2002; Shedlock and Okada, 2000), individual SINEs are generally thought to be homoplasy-free

---

characters (Batzer and Deininger, 2002; Hamdi et al., 1999; Hillis, 1999; Miyamoto, 1999; Roy-Engel et al., 2002; Salem et al., 2003a; Shedlock and Okada, 2000).

This does not mean that SINEs are without problems with regard to phylogenetic analysis. It is known that insertion homoplasy may occur across distantly related taxa as a function of evolutionary time and variable retroposition rates among species (Cantrell et al., 2001; Hillis, 1999; Miyamoto, 1999). This can limit the application of SINEs in examinations of more diverse taxa. Random sorting of the ancestral allelic lineages, sequence convergence, and sequence exchanges between alleles or duplicated loci have also been identified as likely factors confounding the interpretation of the interrelationships among species (Hillis, 1999).

*Alu* elements are primate-specific SINEs of ∼300 bp. These elements have been extremely successful at propagating in primate genomes as evidenced by the fact that they make up ∼10% of the human genome by mass (Batzer and Deininger, 2002; Lander et al., 2001). Distinct families of *Alu* elements in the human genome have been described in detail (Batzer and Deininger, 1991, 2002; Batzer et al., 1991, 1995; Carroll et al., 2001; Salem et al., 2003a; Xing et al., 2003). Examination of these young subfamilies has provided us with clues to the mobilization dynamics and evolution of *Alu* elements in the hominid line. However, characterization of *Alu* mobilization in non-human primates has not been as complete. The ascertainment of these types of markers would increase our understanding of mobile element evolution in distinct lineages. In addition, because they are useful as tools in evolutionary and population biology, the recovery of non-human primate-specific *Alu* markers would be advantageous.

Platyrrhine primates, New World monkeys (NWM), are generally considered a monophyletic group consisting of 15–16 genera. These genera are typically described as belonging to six or seven clades (Barroso et al., 1997; Groves, 1989; Schneider et al., 1993): the callitrichines (*Callithrix*, including *Cebuella*, *Leontopithecus*, *Saguinus*, and *Callimico*); the capuchins (*Cebus*) and squirrel monkeys (*Saimiri*); the owl monkeys (*Aotus*); the saki monkeys (*Pithecia*, *Chiropotes*, and *Cacajao*); the titi monkeys (*Callicebus*); and the spider, woolly, and howler monkeys, and muriqui (*Ateles*, *Lagothrix*, *Alouatta*, and *Brachyteles*, respectively). Three families are typically resolved in analyses using various genetic systems (see Goodman et al., 1998 for a review): Pitheciidae (titi and saki monkeys), Cebidae (callitrichids, capuchins, squirrel monkeys, and owl monkeys), and Atelidae (spider, woolly, and howler monkeys) (Canavez et al., 1999; Harada et al., 1995; Horovitz and Meyer, 1995; Porter et al., 1995, 1996, 1997a, 1999; Schneider et al., 1993, 1996; Singer et al., 2003; Steiper and Ruvolo, 2003; von Dornum and Ruvolo, 1999). Unfortunately, as Schneider (2000) illus-

trates, the branching order of these three families has not been confidently resolved.

For example, several analyses based on gene sequences suggested Cebidae as sister to an Atelidae–Pitheciidae clade (Canavez et al., 1999; Harada et al., 1995; Schneider et al., 1993). On the other hand, one examination of interstitial retinol-binding protein (IRBP) gene sequence data places Atelidae as basal to the remainder of the group (Schneider et al., 1996). Depending on the analysis implemented (Bayesian, parsimony, or maximum-likelihood), glucose-6-phosphate dehydrogenase (G6PD) sequence data produced different topologies (Steiper and Ruvolo, 2003). Furthermore, analysis of 16S RNA sequences separated *Pithecia* and *Callicebus* into distinct groups (Horovitz and Meyer, 1995).

Only one study has attempted to address the issue of platyrrhine phylogeny using SINE insertion polymorphisms (Singer et al., 2003). In their study, 74 intronic *Alu* insertions were examined and they found six elements that were informative at various levels of platyrrhine phylogeny. These markers were used to confirm New World monkey monophyly (three insertions), a close relationship between *Callithrix* and *Cebuella* (one insertion), monophyly of the callitrichines (one insertion), and a relatively close affiliation for *Aotus*, *Saimiri*, and *Cebus* to the callitrichine monkeys (one insertion). While informative, this relatively small set of SINE insertions is limited in its utility. Here, we report a phylogenetic analysis of 183 newly identified *Alu* elements that integrated into various New World monkey (NWM) genomes after divergence of the group from catarrhines (Old World monkeys and hominids).

## 2. Materials and methods

### 2.1. Computational methodology

We identified *Alu* elements in DNA sequences obtained from Bacterial Artificial Chromosome clones (BACs) (Shizuya et al., 1992) of three New World primates (*Callithrix jacchus*, *Saimiri boliviensis boliviensis*, and *Callicebus moloch*) currently available from the GenBank database. These sequences were deposited by the NIH Intramural Sequencing Center (NISC) as part of the Comparative Vertebrate Sequencing Initiative (AC146661, AC146662, AC146674, AC146675, AC146740, AC146768, AC146883–AC146886, AC146925, AC146926, AC147845, AC147932, AC148057–AC148059, AC148120, AC148136, AC148137, AC148186–AC148188, AC148203–AC148208, and AC148246). All of these BAC clone sequences were identified by the NISC database as originating as part of Target 1, a ∼1.5 MB region homologous to chromomsome 7q31 in *Homo sapiens*. *Alu* elements were identified using RepeatMasker

based analysis of the BAC sequences (A.F.A. Smit & P. Green, unpublished data. Current Version: 20040306-web; http://www.repeatmasker.org). Each *Alu* element, along with ~500 bp of flanking sequence in both directions, was excised and subjected to a BLAT search of the human genome [UCSC Genome Browser: http://genome.ucsc.edu/, (Kent et al., 2002)]. This search indicated whether the *Alu* element being investigated was also present in the human genome. The *Alu* elements were then divided into two groups. 'Old' elements were defined as *Alu* insertions present in both the human and the new world monkey genome under investigation. 'Young' elements were defined as *Alu* insertions present in one or more of the platyrrhine primates but not present in *Homo sapiens*. Oligonucleotide primers for polymerase chain reaction (PCR) amplification of each locus were designed for the young or lineage-specific *Alu* insertions using Primer3 (Rozen and Skaletsky, 1998).

## 2.2. Experimental methodology

We employed a modification of the PCR display based ascertainment procedure reported by Roy et al. (1999) to identify potentially informative *Alu* insertions in the genomes of nine additional platyrrhine primates from across the taxonomic distribution. The taxa included were *Aotus trivirgatus* (owl monkey), *Saguinus labiatus* (red-bellied tamarin), *Lagothrix lagotricha* (woolly monkey), *Ateles geoffroyi* (black-handed spider monkey), *Callithrix pygmaea* (pygmy marmoset), *Pithecia p. pithecia* (northern white-faced saki), *Callicebus d. donacophilus* (Bolivian gray titi), *Saimiri s. sciureus* (squirrel monkey), and *Alouatta sara* (Bolivian red howler). The DNA sources for each taxon are listed in Table 1.

Five hundred nanograms of genomic DNA from each taxon were digested (see Fig. 1) using any of three restriction enzymes *Mse*I, *Aci*I, and *Nde*I (New England Biolabs, Beverly, MA) in 120 μl reactions, followed by heat inactivation of the enzyme at 65 °C for 20 min. After several rounds, we began using the enzyme *Nde*I exclusively because the six-base recognition site provided us with longer flanking unique sequences than did the four-base recognition sites of *Mse*I or *Aci*I. This modification aided our search for overlapping unique DNA sequences in the human genome later in the procedure. After digestion, one of two sets of linkers (Table 2) was annealed to the restriction digested DNA using the protocol of Munroe et al. (1994).

Following the protocol described in Roy et al. (1999), we amplified the ligation products using the primer LNP (5′-GAATTCGTCAACATAGCATTTCT-3′) and one of several *Alu*-specific primers (Table 3) to obtain partial *Alu* sequences and the accompanying flanking unique sequences from each NWM taxon. PCR products were separated on 2% agarose gels and fragments of 500 bp or larger were isolated using the Wizard gel purification kit (Promega). The resulting amplicons were either cloned directly into the TOPO-TA cloning vector (Invitrogen) or subjected to a second round of PCR using a nested *Alu*-specific primer and LNP prior to TA cloning. A minimum of 100 clones were randomly selected from each NWM taxon for sequence analysis using chain termination sequencing (Sanger et al., 1992) on an ABI 3100 genetic analyzer.

Clones in which at least 100 bp of flanking sequence was identifiable were subjected to a BLAT search to identify the corresponding human genomic region. We then "backtracked" ~400 bp from the identified overlapping sequence in the human genome to verify that the same *Alu* element was not present at that locus in the

Table 1
List of DNA sources for all species studied.

| Species | Common name | Origin | ID number |
|---|---|---|---|
| *Homo sapiens* | Human | ATCC[a] | CCL2 |
| *Chlorocebus aethiops sabaeus* | Grivet monkey | ATCC | CCL70 |
| *Aotus trivirgatus* | Three-striped owl monkey | ATCC | CRL1556 |
| *Galago senegalensis* | Senegal galago | Cell line[b] | |
| *Pan paniscus* | Bonobo | Coriell[c] | NG05253A |
| *Saguinus labiatus* | Red-bellied tamarin | Coriell | NG05308A |
| *Ateles geoffroyi* | Black-handed spider monkey | Coriell | NG053052 |
| *Lagothrix lagotricha* | Woolly monkey | Coriell | NG05356 |
| *Lemur catta* | Ring-tailed lemur | Coriell | NG07099A |
| *Macaca mulatta* | Rhesus macaque | Coriell | NG07109A |
| *Callithrix pygmaea* | Pygmy marmoset | SDFZ[d] | OR690 |
| *Alouatta sara* | Bolivian red howler monkey | SDFZ | OR749 |
| *Saimiri s. sciureus* | Squirrel monkey | SDFZ | KB4544 |
| *Pithecia p. pithecia* | Northern white-faced saki | SDFZ | OR842 |
| *Callicebus d. donacophilus* | Bolivian gray titi | SDFZ | OR1522 |

[a] From cell lines provided by the American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108.
[b] Adenovirus 12 SV-40-transformed fibroblast cell line maintained in the laboratory of Dr. Mark Batzer, Louisiana State University.
[c] Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103.
[d] Frozen Zoo, San Diego Zoo Research, PO Box 120551, San Diego, CA 92112.
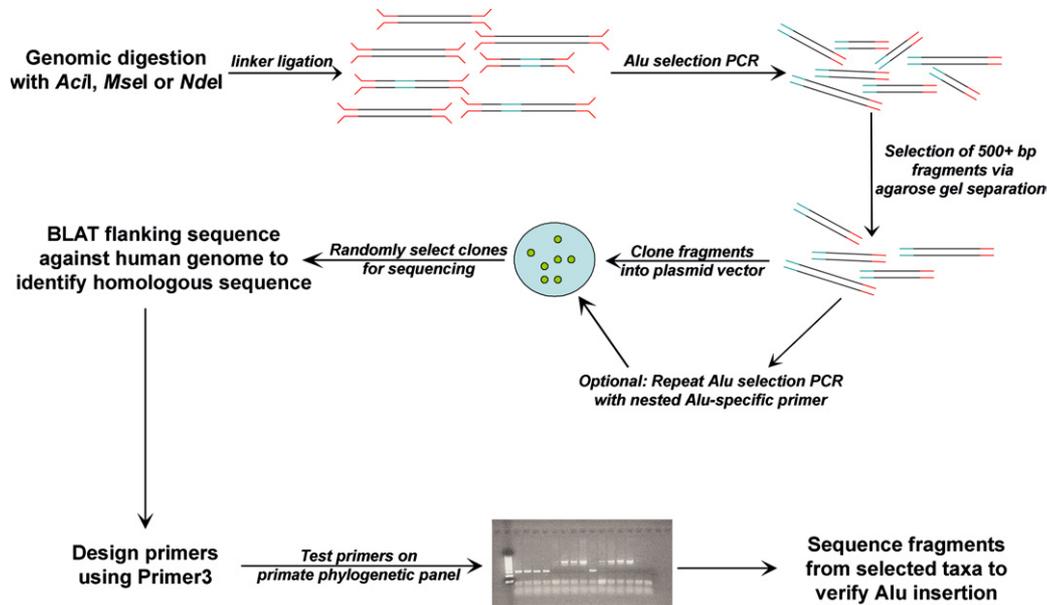
Fig. 1. Illustration of the protocol used to identify NWM-specific *Alu* insertions. Red lines indicate linker fragments ligated to digested genomic DNA (blue/black). Blue lines indicate *Alu* elements in genomic DNA fragments.

Table 2
Linker sequences[a] for *Alu* detection protocol

| Name | Sequence |
|---|---|
| Aci-Top[b] | CGGAAGGAGAGGACGCTGTCTGTCGAAGG |
| Mse-Top[c] | TAGAAGGAGAGGACGCTGTCTGTCGAAGG |
| Aci-Bottom/ Mse-Bottom | GAGCGAATTCGTCAACATAGCATTTCTGTC CTCTCCTTC |

[a] "Top" sequences are identical for each linker pair except for underlined bases which correspond to the two-base overhang at the restriction enzyme cut site.

[b] For use with genomic DNA cut using *Aci*I.

[c] For use with genomic DNA cut using either *Mse*I or *Nde*I, as the overhangs from these restriction enzymes are identical.

Table 3
*Alu* selection primers

| Name | Sequence (5'–3') |
|---|---|
| A*Alu*1A | GGGAGGCCGAGGTGGGT |
| A*Alu*1B | CTTTGGGTGGCTGAGGCA |
| *Alu*1D | CAAGAGATCGAGACCATCCTGGT |
| *Alu*2D | TGCCTGTAATCCCAGCTACTCA |
| *Alu*2E | CTGGGAGTGGTGGCGCGT |
| *Alu*i2B | ACTCCAGCCTGGYRCCTG |
| *Alu*i2C | GGYGCCTGGYRAYRGAGTG |
| *Alu*R1 | GCCTCCTGGGTTCAGGCA |
| Fam1 | GGAGAATTGCTTGAAACCGGAAA |
| Fam2 | CACTCCAGCCTGGCGCCT |
| Fam3 | GCTGAGGCAGGAGAATTGCC |
| S*Alu*1 | AGAGATCGAGACCATCCTGGT |
| S*Alu*2A | CTGGGCATGGTGGCAAGT |

human genome. If the same *Alu* insertion was identified in the human genome, the insertion was considered an old shared character, having inserted prior to the platyr-rhine–catarrhine divergence, and the *Alu* element was discarded. In many cases the lineage-specific *Alu* inser-tions had integrated in or near other repetitive sequences making primer design within unique DNA sequences difficult. When this was the case, oligonucleotide primers were designed only when the repetitive sequences exhib-ited $\geq 25\%$ divergence from the consensus sequence of the repetitive motif. Using a combination of the sequences from the taxon used to isolate the locus and the orthologous human DNA sequences, we were gener-ally able to design oligonucleotide primers for PCR amplification of the loci within $\sim 300\,\text{bp}$ of the insertion using Primer3 (Rozen and Skaletsky, 1998).

### 2.3. PCR and sequencing

All primer pairs were initially tested on human DNA using a temperature gradient PCR (45–60 °C) to deter-mine the appropriate annealing temperature for further testing on a primate phylogenetic panel. This panel included four catarrhine taxa (*Homo sapiens*, *Pan panis-cus*, *Chlorocebus aethiops sabaeus*, and *Macaca mulatta*), one prosimian (*Lemur catta*), and the nine previously described platyrrhine monkeys. Twenty-five microliters PCR amplifications were performed under the following conditions: 10–50 ng template DNA, 7 pM of each oligo-nucleotide primer, 200 mM dNTPs, in 50 mM KCl, 10 mM Tris–HCl (pH 8.4), 2.0 mM MgCl$_2$, and *Taq* DNA polymerase (1.25 U). An initial denaturation at 94 °C for 2 min was followed by 32 cycles of 94 °C for 15 s, the annealing temperature (Supplemental Table, available at http://batzerlab.lsu.edu) for 15 s, and 72 °C for 1 min and 10 s. A final incubation at 72 °C for 3 min prepared the fragments for potential cloning. For the five taxa obtained from the San Diego Zoo (Table 1), very limited amounts of genomic DNA were available. These samples

were subjected to whole genome pre-amplification using the Genomiphi genome amplification kit (Amersham, Sunnyvale, CA) prior to locus-specific PCR analysis. The genome-amplified products were then used as templates for primate panel amplifications.

Sequences from filled and empty sites were ascertained for each locus. In addition, when amplified product sizes varied widely ($\geqslant 50$ bp) among taxa, representative products were selected for DNA sequence analysis to identify the source of the disparity. Individual PCR products were cloned using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA) and inserts were sequenced using chain termination sequencing on an ABI 3100 Genetic Analyzer. Sequences for loci identified experimentally were aligned with the orthologous human sequence obtained via the BLAT search. Sequences and alignments are available at our website, http://batzerlab.lsu.edu. Sequences generated for this project have been deposited in Genbank under Accession Nos. AY620468–AY620746.

## 2.4. Phylogenetic analysis

Alu insertion loci were included in phylogenetic analysis if we were able to amplify a clearly distinguishable



Fig. 2. Results of primate phylogenetic panel PCR amplification for four loci. Locus names and template DNA are indicated.

product (either empty or filled) in at least five of the nine platyrrhine taxa and one of the catarrhine primates (Fig. 2). PAUP* 4.0b10 (Swofford, 2000) was used to perform a parsimony analysis on 190 distinct characters at 177 amplified loci (11 loci were home to multiple insertions distinguishing different sets of taxa). Presence of the insert was coded as "1" and absence of the insert as '0'. If no amplification was observed for a given locus in any taxon, the character state was coded as unknown, '?.' As would be expected, amplification patterns in humans and bonobo were identical with the exception of a few loci where the absence of amplification was uninformative in one or the other species. In addition, successful amplification in Lemur was rare. Because of these observations and because PAUP* allows a maximum of 12 taxa when performing an exhaustive search of all possible trees, we removed P. paniscus and Lemur from the analysis. Dollo parsimony analysis is most appropriately applied when handling presence/absence characters for which the ancestral state can safely be assumed. Thus, we implemented an exhaustive search in PAUP* using Dollo parsimony and designating Homo, Macaca, and Chlorocebus as outgroup taxa. We used the methods developed by Waddell et al. (2001) to determine the statistical strength of each node.

## 3. Results

### 3.1. PCR and sequencing

By the criteria described above, 92 computationally derived and 85 experimentally derived Alu insertion loci were found to be potentially useful for phylogenetic analysis of New World primates. When amplification patterns suggested either a tree that was different from any of the most common topologies suggested by sequence data or when large differences in the size of orthologous loci prompted sequencing of the locus, the sequences invariably revealed the presence of a parallel insertion or some other insertion/deletion event. For example, there were 11 loci for which the amplification pattern alone indicated relationships different from the final tree. Using one of these loci as an example (locus 181), large fragments indicating filled sites were noted in Macaca and Chlorocebus as well as in one or more New World taxa. Sequence analysis at these types of loci showed that the pattern was due to a secondary insertion event in the amplified region but not in exactly the same position, as has been reported previously (Salem et al., 2003a; Schmitz et al., 2001; Vincent et al., 2003). At all other sequenced loci, there was no evidence that the Alu element initially observed had been cleanly removed or that a second element had inserted in exactly the same location in other taxa.
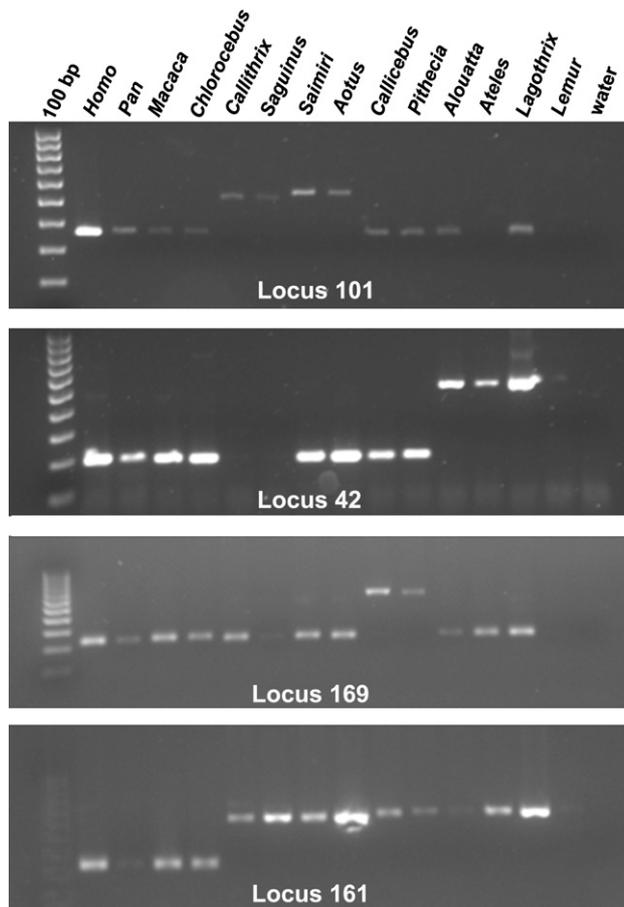
Fig. 3. The most parsimonious tree generated from analysis of 189 *Alu* insertions for nine platyrrhine primates and three outgroup taxa. Numbers with arrows indicate the number of unambiguous insertions supporting that node.

Two loci appeared to be polymorphic within individuals. Locus 29 exhibited two bands in the only species to have an insertion, *A. geoffroyi*. The primer pair for locus 23 amplified two bands in both *Lagothrix* and *Ateles*. In all cases, sequence analysis confirmed that the smaller fragments are authentic pre-integration sites. Whether these instances represent actual heterozygous genotypes in the individuals or amplification of paralogous sequences cannot be determined by these analyses.

### 3.1.1. Phylogenetic analysis

One hundred twenty-four characters were parsimony-informative. Phylogenetic analysis produced a single most parsimonious tree of 188 steps (Fig. 3). The consistency index (CI) for the final tree was 1, indicating that no markers added information that contradicted the topology of the most parsimonious tree. Each of the two next most parsimonious trees required 1 extra step. For these, the consistency indices were reduced to 0.995 with corresponding increases in the homoplasy indices to 0.005. All but two nodes were significantly supported ($p < 0.05$). The two marginally supported nodes include the branch defining the family Cebidae (two insertions) and the branch indicating a sister relationship between *Aotus* and *Saimiri* (one insertion).

## 4. Discussion

Unlike several important studies placing Pitheciidae as a sister group to Atelidae (Ford, 1986; Rosenberger, 1981, 1984), our analysis of SINE loci strongly suggests that Atelidae and Cebidae join together first and are then joined by Pitheciidae. Seven loci unambiguously

support a topology in which *Pithecia* and *Callicebus* are the sister group of the remaining Platyrrhines. These two taxa have been grouped as part of the Pitheciidae (Goodman et al., 1998). Four of the seven loci distinguish Atelidae and Cebidae from Pitheciidae. Three more loci supported a close relationship between *Pithecia* and *Callicebus* to the exclusion of the remaining NWM taxa. A similar arrangement was suggested by Kay (1990) after analysis of morphological data. However, the presence of the three insertions shared by *Callicebus* and *Pithecia* suggests a shared lineage for these taxa not reflected in Kay's tree. That Pitheciidae is sister to an Atelidae–Cebidae clade was also supported by analysis of mitochondrial sequence data (Horovitz and Meyer, 1995) and from X-linked G6PD sequence (Steiper and Ruvolo, 2003; von Dornum and Ruvolo, 1999), although support for the branching order was not strong in any analysis and was contradicted in a Bayesian analysis of the latter data set.

Family Atelidae was defined by Goodman et al. (1998) as comprising *Alouatta*, *Ateles*, *Lagothrix*, and *Brachyteles*. Our data are concordant with this definition, given that six loci support the monophyly of *Ateles*, *Lagothrix*, and *Alouatta*. Within this family, there are three loci that unambiguously support a closer relationship between *Lagothrix* and *Ateles* than between either taxon and *Alouatta*. Four additional loci have the potential to lend support to this topology but ambiguity is introduced by non-amplification in *Alouatta*. Ambiguous loci such as these are not indicated on the final tree (Fig. 3). Several studies (Harada et al., 1995; Horovitz and Meyer, 1995; Porter et al., 1997a; Schneider et al., 1996) have suggested that *Brachyteles* is sister to *Lagothrix*. Unfortunately, our inability to obtain samples from this genus renders us unable to comment on this issue.

Two *Alu* insertions support a cebid clade containing *Aotus*, *Saimiri*, *Callithrix*, and *Saguinus*, thus distinguishing them from the atelid monkeys. Results of a BLAST search using these loci indicate that none of them are the same *Alu* (HBGF) insertion identified previously by Singer et al. (2003) to group these taxa. Thus, by including HBGF, the total number of SINE markers unambiguously supporting this family is brought to three. The addition of the third locus locus reduces the *p* value for this node to from 0.111 to 0.037. Within this family, a close relationship between *Callithrix* and *Saguinus*, both members of subfamily Callitrichinae (Goodman et al., 1998), is very well supported. This is not surprising given that almost all attempts to resolve NWM phylogenetics using molecular data have suggested a close relationship (Canavez et al., 1999; Horovitz and Meyer, 1995; Schneider et al., 1993; von Dornum and Ruvolo, 1999). BLAST searches confirm that none of the markers we have used to define this node were previously identified.

A second clade within Cebidae consisting of *Aotus* and *Saimiri* is unambiguously supported by one locus. A recent review of molecular and fossil evidence (Goodman et al., 1998) proposed that *Aotus* and *Saimiri* belong to separate subfamilies. Our data provide some support for this classification. Of the 63 *Alu* insertion events identified from either *Saguinus* or *Callithrix*, 11 (17.5%) were shared by only these two taxa. In contrast, of the 43 *Alu* elements isolated from either *Aotus* or *Saimiri*, only one (2.3%) was shared exclusively by these taxa. This pattern suggests the following scenario for evolution within the family.

The ancestor of Callitrichinae may have diverged from the *Aotus Saimiri* (common ancestor relatively soon after the entire clade's separation from Atelidae. A rapid separation of owl (subfamily Aotinae) and squirrel (subfamily Cebinae, tribe Saimiriini) monkeys followed, during which relatively few *Alu* elements had the opportunity to become fixed prior to their divergence from a common ancestor. *Callithrix* and *Saguinus*, on the other hand, shared an ancestral lineage for a much longer period and accumulated numerous additional *Alu* insertions in common. This scenario is further supported by the relative number of insertions on the terminal branches of the tree in these four taxa. Assuming a uniform rate of *Alu* insertion in all taxa and given the scenario described, more genus-specific insertions would be expected in *Aotus* and *Saimiri* than in the two callitrichine monkeys.

If one considers various divergence time estimates for taxa within Cebidae (Barroso et al., 1997; Goodman et al., 1998; Porter et al., 1997a,b) this picture is further validated. Estimates calculated with combined IRBP and ε-globin sequences place the divergence of the three subfamilies at 17.5 mya. The same data suggest the divergence between Aotinae and Cebinae occurred relatively soon thereafter (16.7 mya). Finally, the tamarins and marmosets are thought to have shared a common ancestor as recently as 11.1 mya. A more rigorous search for *Alu* insertions and information on the relative rates of retrotransposition in these taxa will be necessary to precisely elucidate these relationships.

### 4.1. Parallel Insertions

Multiple *Alu* insertions within a single amplified locus in taxa can lead to problems interpreting results in phylogenetic analyses that use SINE insertions. As the divergence time between taxa increases, the likelihood of such events should also increase (Hillis, 1999). In this study, we examined loci from taxa that were separated as many as 35–40 mya (Schrago and Russo, 2003). In spite of that fact, we observed only 11 examples of loci with parallel forward insertions. Ten of these were the products of single secondary insertions, but one locus contained four distinct *Alu* insertion events within 300 bp of one

another. Thus, there were thirteen examples of at least one secondary insertion that might confound phylogenetic analysis. However, upon detailed DNA sequence analysis each of these examples was resolved. Six of the multiple insertions were instances where a secondary element had retrotransposed into one or more of the catarrhine taxa when compared to the platyrrhines. Given the time since divergence of these major groups these types of multiple insertions are not unexpected.

Seven other secondary insertions were evident within Platyrrhini. Of these, three could easily be distinguished as instances where a second *Alu* element had inserted within or near the first. Electrophoresis analysis showed a larger than expected PCR product in the taxa containing the secondary elements. Locus 37 provides an example of an instance in which two insertions occurred very close together. In *Ateles*, a truncated *Alu* Sp [nomenclature as in Batzer et al. (1996)] element exists with 5′-AAATGAGAATAAGCTA-3′ direct repeats on either side. At nearly the same location in *Lagothrix*, a full-length *Alu* Sc is present with the direct repeat, 5′-AAGCTAW-3′, almost identical to the final 6 bp of the *Ateles* direct repeat. The truncation of the element in *Ateles* provided a clue that the elements in these taxa were different and DNA sequence analysis supported this hypothesis. At another locus, 46, the amplification pattern suggested that an *Alu* insertion was shared by *Pithecia*, *Ateles*, *Lagothrix*, and *Alouatta*. Nevertheless, DNA sequence analysis of the locus revealed that a secondary *Alu* insertion was present in *Pithecia* downstream of the originally identified insertion, which was present only in the atelids.

Locus 58 was the most complex and most easily misinterpreted locus. Analysis by electrophoresis (see Fig. 4) suggested that an *Alu* element was shared by macaque, green monkey, marmoset, tamarin and squirrel monkey. The catarrhine amplicons could easily be explained as a secondary insertion; however, the pattern in *Saimiri*, *Callithrix*, and *Saguinus* was another matter because it represented a clear and not unlikely alternative to the phylogeny suggested by other insertions. The anomalous amplification pattern was determined to be due to the independent insertion of three different *Alu* elements in the same genomic location in addition to the element originally identified in tamarin. First, an *Alu* Y element integrated in two outgroup taxa, *Chlorocebus* and *Macaca*, ~80 bp downstream of the originally identified insertion, which was present in *Saguinus* and *Callithrix*. A truncated *Alu* Sq/x element was identified in *Aotus*, further downstream of the originally identified insertion. A fourth insertion, of a full-length *Alu* Sq element, had occurred in *Saimiri* at almost the same position as the truncated element in *Aotus*. In fact, these two elements have inserted in such close proximity that, like the *Alu* elements at locus 37, they share a portion of their target site duplications. Thus, locus 58 contained four indepen-
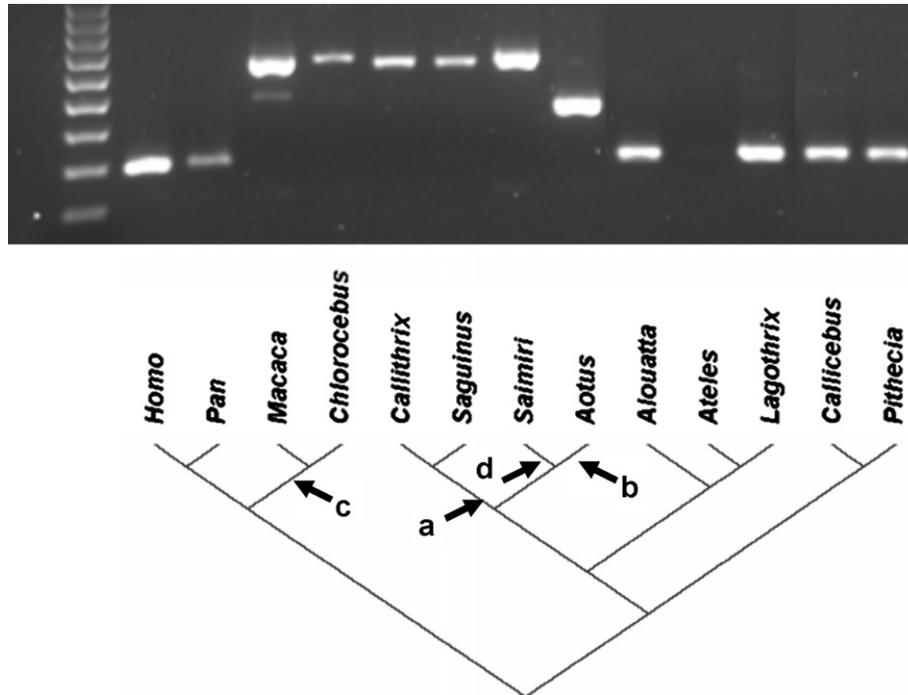
Fig. 4. The *Alu* insertion pattern exhibited at locus 58. Four distinct *Alu* insertions occurred along several lineages. Element "a" inserted in the Callitrichine common ancestor. A second *Alu*, "c," inserted in the common ancestor of macaques and green monkeys. A truncated *Alu* element, "b," inserted in the owl monkey lineage (*Aotus*) after the divergence from squirrel monkeys. Finally, a fourth, full-length insertion, "d," occurred in the lineage leading to *Saimiri sciureus*.

dent *Alu* insertions within the same 300 bp genomic span.

A special case of multiple insertions at a locus is the "precise" parallel forward insertion. This event is defined as the insertion of a different *Alu* element at an identical target site in different taxa. Only two instances similar to a precise parallel insertion were observed, at locus 37 and locus 58. We are able to distinguish these as separate events, however, due to the truncations of one *Alu* element in each case as well as through differences in target site duplications. This rate of precise parallel insertion, 1.04%, is very similar to the levels obtained in other studies (Salem et al., 2003a; Vincent et al., 2003).

## 5. Conclusions

This study represents the third large-scale application (>100 markers) of SINE elements to primate phylogenetic analysis (Salem et al., 2003b; Schmitz et al., 2001). As with the study by Schmitz et al. only a small subset of markers were informative along some of the more recent branches of the tree. In our case, 28 insertions represent data supporting branches within a monophyletic Platyrrhini. Recently, hominid phylogeny was conclusively resolved using 133 *Alu* insertions (Salem et al., 2003b). In that case only one instance of potential lineage sorting of a SINE was found that may have led to misinterpretation of the branching order. In this study, we detected no

examples of lineage sorting and relatively few potentially confounding loci. Thus, the utility of SINEs as phylogenetic markers continues to be supported. We conclude by suggesting that identifying additional *Alu* elements that resolve other relationships in the primate lineage will be an important step forward in completely resolving the primate phylogenetic tree. Over one million *Alu* insertions are known to exist in the human genome (Batzer and Deininger, 2002) and there is reason to believe that similar numbers will be found in other primate species. A recent analysis in our laboratory compared *Alu* elements on human chromosome 21 and *Pan troglodytes* chromosome 22 (Hedges et al., 2004). Results of this study suggested that over 6000 *Alu* elements are specific to *Homo sapiens*. If similar numbers of species-specific *Alu* insertions have occurred in other primate species, one could imagine the presence of multiple *Alu* insertions to elucidate each and every branch of the primate order.

## References

Barroso, C.M.L., Schneider, H., Schneider, M.P.C., Sampaio, I., Harada, M.L., Czelusniak, J., Goodman, M., 1997. Update on the phylogenetic systematics of new world monkeys: further DNA evidence for placing the pygmy marmoset (*Cebuella*) within the genus *Callithrix*. Int. J. Primatol. 18, 651–674.

Batzer, M.A., Deininger, P.L., 1991. A human-specific subfamily of *Alu* sequences. Genomics 9, 481–487.

Batzer, M.A., Deininger, P.L., 2002. *Alu* repeats and human genomic diversity. Nat. Rev. Genet. 3, 370–379.

Batzer, M.A., Deininger, P.L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C.M., Schmid, C.W., Zietkiewicz, E., Zuckerkandl, E., 1996. Standardized nomenclature for *Alu* repeats. J. Mol. Evol. 42, 3–6.

Batzer, M.A., Gudi, V.A., Mena, J.C., Foltz, D.W., Herrera, R.J., Deininger, P.L., 1991. Amplification dynamics of human-specific (HS) *Alu* family members. Nucleic Acids Res. 19, 3619–3623.

Batzer, M.A., Rubin, C.M., Hellmann-Blumberg, U., Alegria-Hartman, M., Leeflang, E.P., Stern, J.D., Bazan, H.A., Shaikh, T.H., Deininger, P.L., Schmid, C.W., 1995. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats. J. Mol. Biol. 247, 418–427.

Canavez, F.C., Moreira, M.A., Ladasky, J.J., Pissinatti, A., Parham, P., Seuanez, H.N., 1999. Molecular phylogeny of new world primates (Platyrrhini) based on beta2-microglobulin DNA sequences. Mol. Phylogenet. Evol. 12, 74–82.

Cantrell, M.A., Filanoski, B.J., Ingermann, A.R., Olsson, K., DiLuglio, N., Lister, Z., Wichman, H.A., 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. Genetics 158, 769–777.

Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen, L., Sammarco, M., Watkins, W.S., Henke, J., Makalowski, W., Jorde, L.B., Deininger, P.L., Batzer, M.A., 2001. Large-scale analysis of the *Alu* Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J. Mol. Biol. 311, 17–40.

Ford, S.M., 1986. Systematics of the New World Monkeys. In: Swindler, D.R., Erwin, J. (Eds.), Comparative Primate Biology. A.R. Liss, New York, pp. 73–135.

Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G., Groves, C.P., 1998. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. 9, 585–598.

Groves, C.P., 1989. A Theory of Human and Primate Evolution. Oxford University Press, New York.

Hamdi, H., Nishio, H., Zielinski, R., Dugaiczyk, A., 1999. Origin and phylogenetic distribution of *Alu* DNA repeats: irreversible events in the evolution of primates. J. Mol. Biol. 289, 861–871.

Harada, M.L., Schneider, H., Schneider, M.P., Sampaio, I., Czelusniak, J., Goodman, M., 1995. DNA evidence on the phylogenetic systematics of New World monkeys: support for the sister-grouping of *Cebus* and *Saimiri* from two unlinked nuclear genes. Mol. Phylogenet. Evol. 4, 331–349.

Hedges, D.J., Callinan, P.A., Cordaux, R., Xing, J., Barnes, E., Batzer, M.A., 2004. Differential *Alu* mobilization and polymorphism among the human and chimpanzee lineages. Genome Res. 14, 1068–1075.

Hillis, D.M., 1999. SINEs of the perfect character. Proc. Natl. Acad. Sci. USA 96, 9979–9981.

Horovitz, I., Meyer, A., 1995. Systematics of New World monkeys (Platyrrhini, Primates) based on 16S mitochondrial DNA sequences: a comparative analysis of different weighting methods in cladistic analysis. Mol. Phylogenet. Evol. 4, 448–456.

Kay, R.F., 1990. The phyletic relationships of extant and fossil Pitheciinae (Platyrrhini, Anthropoidea). J. Human Evol. 19, 175–208.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Miyamoto, M.M., 1999. Molecular systematics: perfect SINEs of evolutionary history?. Curr. Biol. 9, R816–R819.

Munroe, D.J., Haas, M., Bric, E., Whitton, T., Aburatani, H., Hunter, K., Ward, D., Housman, D.E., 1994. IRE-bubble PCR: a rapid method for efficient and representative amplification of human genomic DNA sequences from complex sources. Genomics 19, 506–514.

Porter, C.A., Czelusniak, J., Schneider, H., Schneider, M.P., Sampaio, I., Goodman, M., 1997a. Sequences of the primate ε-globin gene: implications for systematics of the marmosets and other New World primates. Gene 205, 59–71.

Porter, C.A., Czelusniak, J., Schneider, H., Schneider, M.P., Sampaio, I., Goodman, M., 1999. Sequences from the 5′ flanking region of the ε-globin gene support the relationship of *Callicebus* with the pitheciins. Am. J. Primatol. 48, 69–75.

Porter, C.A., Goodman, M., Stanhope, M.J., 1996. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. Mol. Phylogenet. Evol. 5, 89–101.

Porter, C.A., Page, S.L., Czelusniak, J., Schneider, H., Schneider, M.P., Sampaio, I., Goodman, M., 1997b. Phylogeny and evolution of selected primates as determined by sequences of the ε-globin locus and 5′ flanking regions. Int. J. Primatol. 18, 261–295.

Porter, C.A., Sampaio, I., Schneider, H., Schneider, M.P., Czelusniak, J., Goodman, M., 1995. Evidence on primate phylogeny from ε-globin gene sequences and flanking regions. J. Mol. Evol. 40, 30–55.

Rosenberger, A.L., 1981. Systematics: the higher taxa. In: Coimbra-Filho, A.F., Mittermeier, R. (Eds.), Ecology and Behaviour of Neotropical Primates. Academia Brasileira de Ciencias, Rio de Janeiro, pp. 9–27.

Rosenberger, A.L., 1984. Fossil New World monkeys dispute the molecular clock. J. Human Evol. 13, 737–742.

Roy, A.M., Carroll, M.L., Kass, D.H., Nguyen, S.V., Salem, A.H., Batzer, M.A., Deininger, P.L., 1999. Recently integrated human *Alu* repeats: finding needles in the haystack. Genetica 107, 149–161.

Roy-Engel, A.M., Carroll, M.L., El-Sawy, M., Salem, A.H., Garber, R.K., Nguyen, S.V., Deininger, P.L., Batzer, M.A., 2002. Non-tradi-

tional *Alu* evolution and primate genomic diversity. J. Mol. Biol. 316, 1033–1040.

Rozen, S., Skaletsky, H.J., 1998. Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Salem, A.H., Kilroy, G.E., Watkins, W.S., Jorde, L.B., Batzer, M.A., 2003a. Recently integrated *Alu* elements and human genomic diversity. Mol. Biol. Evol. 20, 1349–1361.

Salem, A.H., Ray, D.A., Xing, J., Callinan, P.A., Myers, J.S., Hedges, D.J., Garber, R.K., Witherspoon, D.J., Jorde, L.B., Batzer, M.A., 2003b. *Alu* elements and hominid phylogenetics. Proc. Natl. Acad. Sci. USA 100, 12787–12791.

Sanger, F., Nicklen, S., Coulson, A.R., 1992. DNA sequencing with chain-terminating inhibitors. 1977. Biotechnology 24, 104–108.

Schmitz, J., Ohme, M., Zischler, H., 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157, 777–784.

Schneider, H., 2000. The current status of the New World monkey phylogeny. An. Acad. Bras. Cienc. 72, 165–172.

Schneider, H., Sampaio, I., Harada, M.L., Barroso, C.M., Schneider, M.P., Czelusniak, J., Goodman, M., 1996. Molecular phylogeny of the New World monkeys (Platyrrhini, primates) based on two unlinked nuclear genes: IRBP intron 1 and ε-globin sequences. Am. J. Phys. Anthropol. 100, 153–179.

Schneider, H., Schneider, M.P., Sampaio, I., Harada, M.L., Stanhope, M., Czelusniak, J., Goodman, M., 1993. Molecular phylogeny of the New World monkeys (Platyrrhini, Primates). Mol. Phylogenet. Evol. 2, 225–242.

Schrago, C.G., Russo, C.A., 2003. Timing the origin of new world monkeys. Mol. Biol. Evol. 20, 1620–1625.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. Bioessays 22, 148–160.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., Okada, N., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388, 666–670.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M., 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89, 8794–8797.

Singer, S.S., Schmitz, J., Schwiegk, C., Zischler, H., 2003. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol. Phylogenet. Evol. 26, 490–501.

Steiper, M.E., Ruvolo, M., 2003. New World monkey phylogeny based on X-linked G6PD DNA sequences. Mol. Phylogenet. Evol. 27, 121–130.

Swofford, D.L., 2000. PAUP: Phylogenetic analysis using parsimony. Sinauer Associates, Sunderland, MA.

Takahashi, K., Nishida, M., Yuma, M., Okada, N., 2001. Retroposition of the AFC family of SINEs (short interspersed repetitive elements) before and during the adaptive radiation of cichlid fishes in Lake Malawi and related inferences about phylogeny. J. Mol. Evol. 53, 496–507.

Vincent, B.J., Myers, J.S., Ho, H.J., Kilroy, G.E., Walker, J.A., Watkins, W.S., Jorde, L.B., Batzer, M.A., 2003. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. Mol. Biol. Evol. 20, 1338–1348.

von Dornum, M., Ruvolo, M., 1999. Phylogenetic relationships of the New World monkeys (Primates, platyrrhini) based on nuclear G6PD DNA sequences. Mol. Phylogenet. Evol. 11, 459–476.

Waddell, P.J., Kishino, H., Ota, R., 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform. Ser. Workshop Genome Inform. 12, 141–154.

Xing, J.C., Salem, A.H., Hedges, D.J., Kilroy, G.E., Watkins, W.S., Schienman, J.E., Stewart, C.B., Jurka, J., Jorde, L.B., Batzer, M.A., 2003. Comprehensive analysis of two *Alu* Yd subfamilies. J. Mol. Evol. 57, S76–S89.