

12-1-2005

A mobile element based phylogeny of Old World monkeys

Jinchuan Xing
Louisiana State University

Hui Wang
Louisiana State University

Kyudong Han
Louisiana State University

David A. Ray
Louisiana State University

Cheney H. Huang
Louisiana State University

See next page for additional authors

Follow this and additional works at: https://repository.lsu.edu/biosci_pubs

Recommended Citation

Xing, J., Wang, H., Han, K., Ray, D., Huang, C., Chemnick, L., Stewart, C., Disotell, T., Ryder, O., & Batzer, M. (2005). A mobile element based phylogeny of Old World monkeys. *Molecular Phylogenetics and Evolution*, 37 (3), 872-880. <https://doi.org/10.1016/j.ympev.2005.04.015>

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact ir@lsu.edu.

Authors

Jinchuan Xing, Hui Wang, Kyudong Han, David A. Ray, Cheney H. Huang, Leona G. Chemnick, Caro Beth Stewart, Todd R. Disotell, Oliver A. Ryder, and Mark A. Batzer

A mobile element based phylogeny of Old World monkeys

Jinchuan Xing^a, Hui Wang^a, Kyudong Han^a, David A. Ray^a, Cheney H. Huang^a,
Leona G. Chemnick^b, Caro-Beth Stewart^c, Todd R. Disotell^d, 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, Baton Rouge, LA 70803, USA

^b Conservation and Research for Endangered Species (CRES), Zoological Society of San Diego, San Diego, CA 92112, USA

^c Department of Biological Sciences, University at Albany, State University of New York, Albany, NY 12222, USA

^d Department of Anthropology, New York University, New York, NY 10003, USA

Received 21 February 2005; revised 15 April 2005

Available online 3 June 2005

Abstract

SINEs (Short INterspersed Elements) are a class of non-autonomous mobile elements that are <500 bp in length and have no open reading frames. Individual SINE elements are essentially homoplasly free with known ancestral states, making them useful genetic systems for phylogenetic studies. *Alu* elements are the most successful SINE in primate genomes and have been utilized for resolving primate phylogenetic relationships and human population genetics. However, no *Alu* based phylogenetic analysis has yet been performed to resolve relationships among Old World monkeys. Using both a computational approach and polymerase chain reaction display methodology, we identified 285 new *Alu* insertions from sixteen Old World monkey taxa that were informative at various levels of catarrhine phylogeny. We have utilized these elements along with 12 previously reported loci to construct a phylogenetic tree of the selected taxa. Relationships among all major clades are in general agreement with other molecular and morphological data sets but have stronger statistical support.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Cercopithecidae; Primates; Phylogeny; *Alu*; SINE

1. Introduction

Old World monkeys (family Cercopithecidae) represent one of the largest and most diverse primate families. Together with family Hominidae (humans and apes), they form the infraorder Catarrhini. Extant Old World monkeys can be divided into two ecologically and morphologically distinct subfamilies: Cercopithecinae (cheek-pouched monkeys) and Colobinae (leaf-eating monkeys) (Delson, 1992; Disotell, 2003; Groves, 1993, 2001). Cercopithecidae encompasses at least 21 genera,

minimally 11 in subfamily Cercopithecinae and ten in Colobinae (Groves, 2001; Disotell, 2003). Subfamily Cercopithecinae includes *Erythrocebus*, *Chlorocebus*, *Cercopithecus*, *Miopithecus*, *Allenopithecus*, *Cercocebus*, *Lophocebus*, *Macaca*, *Papio*, *Mandrillus*, and *Theropithecus*. With the exception of genus *Macaca* and some small populations of Arabian baboons, the genera have solely African distributions. The subfamily Colobinae is subdivided into two clades: the African genera *Colobus*, *Procolobus* and *Piliocolobus*, and the Asian genera *Nasalis*, *Rhinopithecus*, *Presbytis*, *Pygathrix*, *Semnopithecus*, *Trachypithecus*, and *Simias*. The existence of other genera such as, *Allochrocebus* and *Presbyticus* are under debate (Groves, 2001). In addition to the classification system described above, Goodman et al. (1998) proposed

* Corresponding author. Fax: +1 225 578 7113.
E-mail address: mbatzer@lsu.edu (M.A. Batzer).

an alternative taxonomic classification based on the ages of the clades. In this system, all the extant members of family Cercopithecidae have been placed in the single subfamily Cercopithecinae which subdivided into tribe Cercopithecini and Colobini, corresponding to subfamilies Cercopithecinae and Colobinae mentioned previously (Groves, 2001). For clarity and consistency we have chosen to follow the nomenclature of Groves (2001).

Several molecular phylogenetic studies of Cercopithecidae have been conducted previously (Disotell, 1994; Harris and Disotell, 1998; Messier and Stewart, 1997; Page et al., 1999; Page and Goodman, 2001; Tosi et al., 2004). For subfamily Cercopithecinae, the tribe and subtribe level relationships inferred from the molecular based studies (Page et al., 1999; Page and Goodman, 2001) are congruent with classifications based on morphological data (Delson, 1992; Goodman et al., 1998). However, considerable disagreement still exists for some relationships at the genus level. For example, for the tribe Papionini, the morphology-based classification (Delson, 1992) is discordant with molecular evidence (Harris and Disotell, 1998; Page et al., 1999). Molecular data also suggest relationships among Cercopithecini that have not been proposed in morphology-based phylogenetic hypotheses (Tosi et al., 2004).

Unlike the cercopithecids, only a few DNA sequence based studies have been performed concerning the branching order of the Colobinae (Messier and Stewart, 1997; Zhang and Ryder, 1998), primarily due to the difficulty of obtaining DNA samples from these rare and endangered species. Genus level relationships of colobines are still in flux due to discordant evidence. For example, the branching order of Colobinae inferred from morphological and fossil studies are different from that derived from karyotype studies, especially for the relationships among the Asian colobines (Bigoni et al., 2003, 2004; Jablonski and Peng, 1993). An independent set of molecular markers may help to resolve the phylogeny for both groups of catarrhine primates.

SINE (Short INterspersed Element) insertions are genetic markers that have proven useful at several levels of phylogenetic analysis (Nikaido et al., 2001; Ray et al., 2005; Roos et al., 2004; Salem et al., 2003b; Takasaki et al., 1997; Zietkiewicz et al., 1999). SINEs are a class of non-autonomous mobile elements that are <500 bp in length and have no open reading frames. SINEs are ubiquitous in all mammalian genomes examined as well as in many non-mammalian genomes (Deininger and Batzer, 1993, 2002; Okada, 1991). The utilization of SINEs as phylogenetic markers was proposed over 15 years ago (Murata et al., 1993; Okada, 1991; Ryan and Dugaiczyk, 1989). Since then, SINEs have become widely recognized as powerful tools for phylogenetic

studies and multiple controversial phylogenetic relationships that can not be solved using traditional molecular data have been successfully elucidated (Murata et al., 1993; Nikaido et al., 1999; Okada et al., 2004; Ray et al., 2005; Roos et al., 2004; Salem et al., 2003b; Schmitz et al., 2005; Shedlock et al., 2000, 2004; Takahashi et al., 1998, 2001).

While there may be some disadvantages to using SINEs as phylogenetic markers (see Hillis, 1999 for a review), SINE insertions have several unique characteristics that make them particularly promising in evolutionary analyses. These properties have been reviewed several times (Schmitz et al., 2005; Shedlock et al., 2004; Shedlock and Okada, 2000) and we will only briefly summarize them here. First, the probability of two identical SINEs independently inserting in the exact position with the same orientation is essentially zero. Adjacent independent *Alu* insertions that happen to occur near the same site can readily be distinguished by base sequence analysis. Second, the insertion of a SINE can be assumed to be unidirectional since there is no known mechanism to precisely remove SINEs after their fixation in the genome and the removal of SINEs is very unlikely to happen in multiple genomes.

These two characteristics suggest that SINEs are essentially homoplasy free characters. A third advantage of SINEs, when compared to other molecular phylogenetic systems such as SNPs (single nucleotide polymorphisms) and nuclear and mitochondrial DNA sequences, is that phylogenetic inference using SINEs does not directly rely on acquiring DNA sequence data. This feature makes SINEs an independent complement to traditional DNA sequence based molecular studies (Roy-Engel and El-Sawy, 2002), especially for groups that are known for high levels of sequence homoplasy (Cantrell et al., 2001) and for closely related species for which there is little phylogenetic information in DNA sequences.

As the most successful SINEs in primate genomes, *Alu* elements have enjoyed remarkable proliferation during the primate radiation and have expanded to more than one million copies in the human genome (Batzer and Deininger, 2002; Lander et al., 2001). Recently, *Alu* elements have been used intensively for resolving primate phylogeny at different levels (Hamdi et al., 1999; Ray et al., 2005; Roos et al., 2004; Salem et al., 2003b; Schmitz et al., 2005). However, no SINE based phylogenetic analysis has been performed to resolve the relationships among Old World monkeys. Using both a computational approach and polymerase chain reaction (PCR) display methodology, we identified 285 new *Alu* insertions that integrated into sixteen Old World monkey genomes and used them to construct a robustly supported phylogenetic hypothesis for the family Cercopithecidae.

2. Materials and methods

2.1. Computational data mining

Genomic sequences from *Chlorocebus aethiops* (African green monkey), *Papio anubis* (olive baboon), and *Macaca mulatta* (rhesus monkey) were obtained from the NIH Intramural Sequencing Center, as part of the Comparative Vertebrate Sequencing Initiative. The sequences were broken into 10,000bp fragments and compared to the human genome using the BLAST-Like Alignment Tool (BLAT) program (Kent, 2002) available at <http://genome.ucsc.edu>. Fragments containing insertions/deletions were extracted and annotated to identify putative lineage specific *Alu* insertions using RepeatMasker (<http://repeatmasker.genome.washington.edu>) as previously reported (Hedges et al., 2004). To allow a focus on the Cercopithecoid lineage, only *Alu* elements present in the Old World monkey sequence and missing in the human orthologous regions were excised along with 1000bp of flanking sequence in both directions. Flanking oligonucleotide primers for PCR amplification of each *Alu* element were then designed using Primer3 (Rozen and Skaletsky, 2000). The primers were subsequently screened against the GenBank NR database using Basic Local Alignment Search Tool (BLAST) program available at <http://www.ncbi.nlm.nih.gov/BLAST/> (Altschul et al., 1990) to determine if they resided in unique DNA sequences.

2.2. PCR display methodology

The *Alu* element PCR display methodology described by Ray et al. (2005) was used with minor modifications. These modifications included the use of CHROMAS SPIN-400 columns (BD Biosciences) to select fragments larger than 400bp and different *Alu* selection primers (Table 1).

2.3. PCR and DNA sequencing

All oligonucleotide primer pairs were initially tested for amplification using human DNA templates with a

Table 1
Linkers and primers sequences used for PCR display methodology

Name	Sequence (5'–3')
Linkers	
Top	TAGAAGGAGAGGACGCTGTCTGTGCGAAGG
Bottom	GAGCGAATTCGTCAACATAGCATTCTCTGTCCTCTCCTC
Primers	
LNP ^a	GAATTCGTCAACATAGCATTCT
<i>Alu</i> PY1 ^b	GGAGCTTGACAGTGAGCTGAGATCC
<i>Alu</i> PY2 ^b	GCGACAGAGCGAGACTCC
<i>Alu</i> Yd ^b	AGATCGAGACCACGGTGAA

^a Linker primer.

^b *Alu* internal primers.

temperature gradient PCR (48–60 °C) to determine the most appropriate annealing temperature for further analysis of non-human primate genomes. All loci were screened on a primate panel that was composed of human HeLa genomic DNA and DNA samples from nineteen non-human primate species (Table 2). For some taxa only 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). The amplified samples were then used as templates for locus specific PCR analysis.

PCR amplification of each locus was performed in 25 µl reactions using 10–50 ng of target DNA, 200 nM of each oligonucleotide primer, 200 µM dNTPs in 50 mM KCl, 1.5 mM MgCl₂, 10 mM Tris–HCl (pH 8.4), and 2.5 U *Taq* DNA polymerase. Each sample was subjected to an initial denaturation step of 94 °C for 150 s, followed by 32 cycles of one minute of denaturation at 94 °C, 1 min of annealing at indicated annealing temperature, 1 min of extension at 72 °C, followed by a final extension step at 72 °C for 10 min. Resulting PCR products were run on a 2% agarose gel with 0.25 µg ethidium bromide and visualized using UV fluorescence. Detailed information on each locus including primer sequences, annealing temperature, PCR product sizes, chromosomal locations and amplification results are available on our web site (<http://batzerlab.lsu.edu>) as supplemental data.

To confirm the phylogenetic distribution of *Alu* insertions on the primate panel, loci from selected taxa were also chosen for sequence analysis to verify the presence of the experimentally derived *Alu* element. In addition, when the PCR amplification patterns were different from that suggested by the majority of genetic systems analyzed, representative PCR products were selected for DNA sequence analysis to resolve the data points. Individual PCR products were cloned and sequenced as described previously (Ray et al., 2005). Sequences for loci identified experimentally were aligned with the orthologous human sequence obtained via the BLAT search and sequence alignments of these elements are available from our web site (<http://batzerlab.lsu.edu>) under publications. The DNA sequences generated for this project have been deposited in GenBank under Accession Nos. AY879605 to AY879769.

2.4. Phylogenetic analysis

Alu insertion loci were included in phylogenetic analysis if the amplicons were generated in at least 10 out of 16 cercopithecoid taxa and only two distinct classes of amplicons were generated (*Alu* filled size and pre-integration or *Alu* empty size). Any primer pair that generated multiple paralogous fragments across the panel was excluded from the analysis. Four examples of gel chromatographs of amplification result are shown in Fig. 1.

Table 2
DNA samples of all species examined in this study

Species names	Common names	Origin	ID number
<i>Homo sapiens</i>	Human	ATCC ^a	CCL2
<i>Pan troglodytes</i>	Common Chimpanzee	Coriell ^b	NG06939
<i>Hylobates syndactylus</i>	Siamang	SDFZ ^c	KB 11539
<i>Macaca nemestrina</i>	Pigtailed Macaque	Coriell	NG08452
<i>Macaca silenus</i>	Lion-tailed Macaque	SDFZ	OR 1890
<i>Macaca mulatta</i>	Rhesus Macaque	Coriell	NG07109A
<i>Papio cynocephalus</i>	Yellow Baboon	SFBR ^d	9656
<i>Papio anubis</i>	Olive Baboon	SFBR	8229
<i>Theropithecus gelada</i>	Gelada Baboon	SDFZ	KB 10538
<i>Cercocebus agilis</i>	Agile Mangabey	Disotell ^e	N/A
<i>Chlorocebus aethiops</i>	African Green Monkey	ATCC	CCL70
<i>Erythrocebus patas</i>	Patas Monkey	SDFZ	KB5435
<i>Miopithecus talapoin</i>	Talapoin	SDFZ	OR 755
<i>Colobus guereza</i>	Black and White Colobus	Stewart ^f	N/A
<i>Colobus guereza kikuyuensis</i>	Kikuyu Colobus	SDFZ	OR 160
<i>Pygathrix nemaeus</i>	Douc Langur	SDFZ	OR 259
<i>Nasalis larvatus</i>	Proboscis Monkey	Stewart	N/A
<i>Trachypithecus cristatus</i>	Silvered Leaf Langur	SDFZ	B 4381
<i>Trachypithecus vetulus nestor</i>	Western Purple-faced Langur	SDFZ	OR 219
<i>Aotus trivirgatus</i>	Three-striped Owl Monkey	ATCC	CRL1556

^a From cell lines provided by the American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108.

^b Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103.

^c Frozen Zoo, San Diego Zoo, conservationandscience.org.

^d Southwest Foundation for Biomedical Research.

^e Blood sample from Dr. Todd Disotell.

^f DNA sample from Dr. Caro-Beth Stewart.

We implemented a heuristic search in PAUP* 4.0b10 (Swofford, 2003) using Dollo parsimony and designating owl monkey as an outgroup taxon. 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, “?” For loci containing more than one insertion event (see Section 3), the independent insertion events were treated as two independent markers. Ten thousand bootstrap replicates were performed on the data. A statistical test for evaluating SINE insertions based on a likelihood model (Waddell et al., 2001) was also performed to assess the statistical significance of each branch of the resulting tree.

3. Results and discussion

In total, 285 new loci were identified for phylogenetic analysis. Seventy-five loci were collected from computational data mining and 196 loci were collected from PCR display methodology. Fourteen of the loci contained two independent *Alu* insertion events and were treated as two independent markers. An additional 12 loci were selected from a previous study (Salem et al., 2003b) to distinguish the hominid lineage from the remainder of Catarrhini. Thus, a total 297 markers were used for the phylogenetic analysis. Of the 297 markers, 218 loci were parsimony informative. The analysis resulted in a single

most parsimonious tree (Fig. 2; 302 steps, CI=0.983; HI=0.017; RI=0.993). Because at least five unambiguous *Alu* insertion events were recovered for each clade, the likelihood test for every branch was significant at the 0.005 level (Waddell et al., 2001).

Within family Cercopithecidae, the two previously established subfamilies were distinguishable. Thirty-three loci supported the monophyly of Cercopithecinae and thirteen loci supported monophyly of Colobinae. Within subfamily Cercopithecinae, our tree is congruent with previous studies, but provides a much higher level of statistical support. Two tribes are recognized, Papionini and Cercopithecini (Groves 2001), and these tribes were supported by 15 and 7 insertions, respectively.

Within Papionini, nine *Alu* insertions suggest that *Papio* and *Theropithecus* form a monophyletic clade. These taxa are joined by *Cercocebus* to form subtribe Papionina, which is supported by five unambiguous loci. Members of *Macaca* were clearly defined as a sister clade to subtribe Papionina, by 15 unambiguous loci. These results are congruent with several previous phylogenetic studies (Disotell et al., 1992, 1994; Page et al., 1999; Page and Goodman, 2001). Within genus *Macaca* our three representatives formed a monophyletic group supported by eight loci. *M. nemestrina* and *M. silenus* shared a closer relationship (six insertions) than either did with *M. mulatta*. These results are consistent with a comprehensive study of macaque phylogeny by Hayasaka et al. (1996).

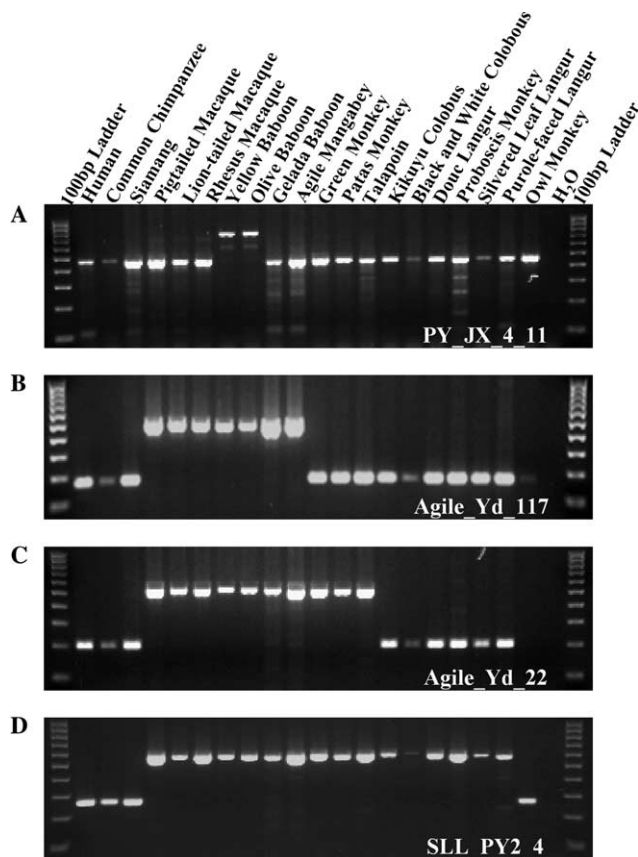


Fig. 1. Phylogenetic analysis of several *Alu* insertions in Old World monkey genomes. Four examples of gel chromatographs used to determine the phylogenetic origin of individual *Alu* insertions in Old World primates are shown. Upper DNA fragments indicate “filled” sites where an *Alu* has inserted. The DNA template used in each reaction and the locus designation are shown. (A) An *Alu* insertion specific for genus *Papio*. (B) An *Alu* insertion restricted to the tribe Papionini and absent from other Old World monkeys. (C) An *Alu* element present in the subfamily Cercopithecinae. (D) An *Alu* insertion found within the genomes of all Old World monkeys examined.

Unfortunately, we had only partial access to DNA from members of tribe Cercopithecini. However, several conclusions can be made from the available data. For example, the monophyly of the tribe is supported by seven *Alu* insertions. Within the tribe, six additional insertions joined *Chlorocebus* and *Erythrocebus*. Within this tribe, it has been suggested that a distinct arboreal-terrestrial split exists that can be resolved using molecular data (Tosi et al., 2004). We hope to address this issue with additional *Alu* insertion data in the near future.

With regard to the second subfamily, Colobinae, our results clearly separate African species (*Colobus guereza kikuyuensis* and *Co. guereza*) from Asian species (*Pygathrix nemaeus*, *Nasalis larvatus*, *Trachypithecus cristatus*, and *Tr. vetulus nestor*), reinforcing the existence of an African clade (*Colobus* monkeys) and an Asian clade (langurs) (Page and Goodman, 2001). Branching order within the Asian lineage is one of the more hotly debated

areas in Old World monkey phylogeny. One view, based on fossil and morphology studies, indicates that *Nasalis* was the first to diverge from the rest of the group (Peng et al., 1993). On the other hand, karyotypic studies indicate that the *Pygathrix* was the first to diverge and a relatively close relationship between *Nasalis* and *Trachypithecus* exists (Bigoni et al., 2003, 2004). In contrast, one molecular study based on mitochondrial cytochrome *b* data suggests that *Nasalis* and *Rhinopithecus* form a sister group to *Pygathrix*. They also suggested that a third genus, *Trachypithecus*, is paraphyletic and that one subgroup of the genus forms a basal clade to the remainder of the langurs (Zhang and Ryder, 1998). Using a consensus approach, Disotell (2003) was unable to resolve the relationships between many of these genera. The *Alu* insertion data presented here are useful in resolving some of these problematic relationships.

In our study, a sister relationship between *Nasalis* and *Pygathrix* is strongly supported by six shared *Alu* insertions. This result is similar to the one found by Zhang and Ryder (1998) but with higher statistical support. The *Nasalis* *Pygathrix* clade is subsequently joined by *Trachypithecus* (18 insertions), clearly delineating at least part of the branching order of the Asian colobines. Unfortunately, we were unable to obtain samples from the remaining genera. However, the power of this approach in resolving problematic phylogenies is clear and we would suggest that these relationships should be examined in the near future.

One drawback in phylogenetic analysis using SINE insertion loci involve two scenarios that lead to confounding loci in the dataset—adjacent independent insertions and lineage sorting. In the present data set, there were nineteen loci at which the PCR amplification pattern indicated alternative branching orders when compared to the final tree. Sequence analysis revealed that fourteen of the nineteen questionable loci were caused by adjacent independent *Alu* insertions. One example of two adjacent independent insertions at locus “TA_PY2_17” is shown (Fig. 3A). This result is not surprising since similar events have been reported in multiple previous studies (Ray et al., 2005; Salem et al., 2003b; Xing et al., 2003) and the likelihood of these events increases as the divergence time between taxa increases (Hillis, 1999). However, since none of these insertions occurred in precisely the same position, DNA sequence analysis resolves these ambiguous characters.

Four additional loci exhibited more complex evolutionary histories. At locus “V11RE” (Fig. 3B), *Alu* insertions are present in all species in subfamily Cercopithecinae except the three macaque species (rhesus monkey, pig-tailed macaque, and lion-tailed macaque). At locus “PYJX12,” the *Alu* insertions are present in the all of the Old World monkey species examined except the three macaque species. Sequence analysis indicates all of the smaller fragments were gen-

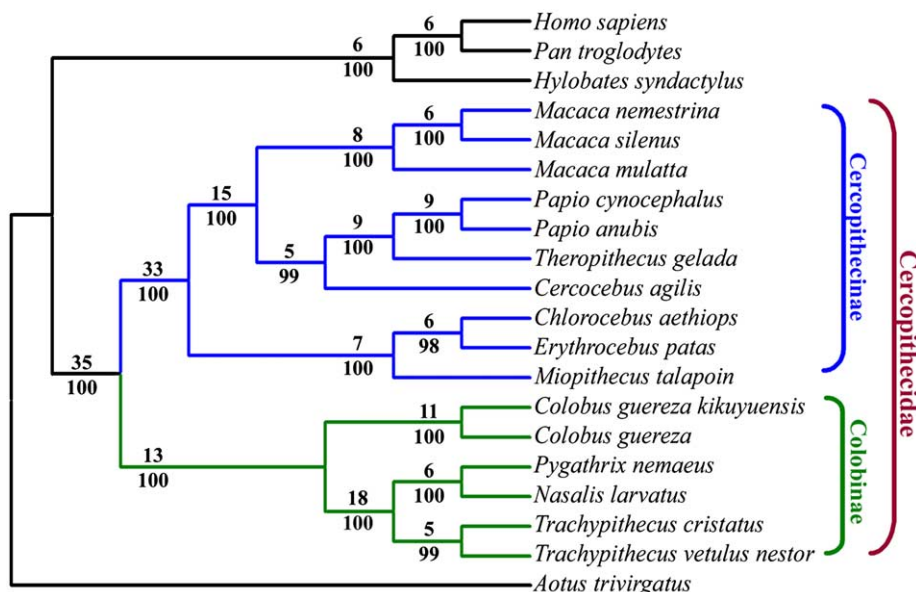


Fig. 2. A cladogram of Old World monkey phylogenetic relationships. The most parsimonious tree generated from analysis of 297 *Alu* insertion polymorphisms. The numbers below the branches indicate the percent of bootstrap replicates (10,000 iterations) producing trees that including that node. Numbers above the branches indicate the numbers of unambiguous insertions supporting each node. For the likelihood test, all nodes are significant at the 0.005 level.

erated from amplification of authentic pre-integration sites. We believe the discordant amplification patterns at these two loci are likely the products of incomplete lineage sorting of ancestral polymorphisms for two reasons: (1) the *Alu* insertions are all flanked by identical target site duplications (TSDs) in multiple taxa, (2) the *Alu* insertions in all the species are in the same orientation and have very similar mutation patterns.

At locus “Agilis_Yd_114,” *Alu* insertions are present in all examined species of subfamily Cercopithecinae except the African green monkey (*C. aethiops*). The most parsimonious explanation is that the *Alu* element was removed from the green monkey genome. Although there is no known mechanism for specific removal of *Alu* element from genome, it is thought to occur on rare occasions via recombination and/or other mechanisms (Deininger and Batzer, 1999; Edwards and Gibbs, 1992). However, we cannot rule out the possibility of multiple incomplete lineage sorting events, that might generate a similar pattern.

At locus “Agilis_Yd_10,” *Alu* insertions are present in agile mangabey (*Cercocebus agilis*), patas monkey (*Erythrocebus patas*), and talapoin (*Miopithecus talapoin*). This pattern is difficult to explain because whatever the evolutionary history is, multiple rare events are required to generate this result. One of the possible scenarios is that, there were two independent *Alu* insertions. One inserted in *Cercocebus* lineage and the second one inserted before the divergence of *Chlorocebus*, *Erythrocebus*, and *Miopithecus*. Later, lineage sorting resulted in the second insertion being fixed in *Erythrocebus* and *Miopithecus* but lost in *Cercopithecus*.

One final case (locus “DoucL_Yd_34RE”) represents a new but not unexpected type of confounding locus. In African green monkey and patas monkey, it appears that an *Alu–Alu* recombination occurred between the newly inserted *Alu* repeat and an old *Alu* element adjacent to it in the flanking region. This recombination removed half of the younger *Alu* element and half of the older *Alu* element, generating an amplification product that has similar size to the pre-integration site. Since the recombination products are identical, it probably occurred before the divergence of these two species. Sequence alignments of all confounding loci can be found on our website (<http://batzlerlab.lsu.edu>) under publications.

Despite a low level of potentially confusing results at some nodes, the overall picture remains very robust. All nodes are supported by at least five loci that are completely unambiguous. Lineage sorting and other unusual events occurred at a rate of approximately 1.4% that is directly comparable to that obtained in previous SINE based studies of phylogenetic relationships (Ray et al., 2005; Salem et al., 2003a,b).

We would be remiss if we did not make some mention of the statistical power of SINE phylogenies when compared to the support for standard sequence data analysis. For example, while bootstrap values are typically reported for analyses of sequence data, there is controversy regarding their interpretation (reviewed in Soltis and Soltis, 2003).

By contrast, SINE insertion based analyses provide a more straight-forward version of statistical support (Waddell et al., 2001). This strength of support, along with the relative ease with which insertion patterns are

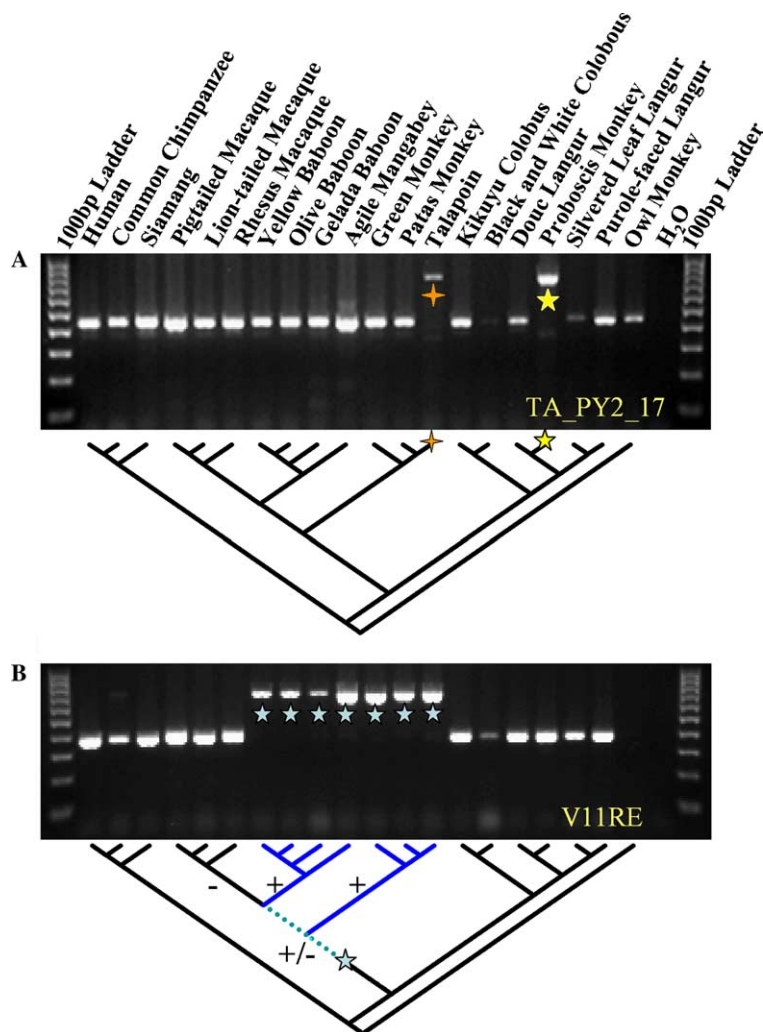


Fig. 3. Potential confounding *Alu* insertions. Two examples of potential confounding loci are shown. The upper part of each figure shows the agarose gel chromatograph with the final tree superimposed on the bottom part. The cross and star denote the approximate times of Old World primate evolution when the insertion of the *Alu* element could have occurred. (A) Parallel independent insertions. At locus “TA_PY2_17,” two *Alu* elements independently inserted in the Talapoin and Proboscis monkey genomes. The integration sites are separated by 42 bp. (B) Ancient incomplete lineage sorting. At locus “VI1RE,” *Alu* insertion happened before the divergence of subfamily Cercopithecinae and was polymorphic (indicated as “±”) in the ancestral population. Random fixation of the alternative alleles (“+” or “−”) resulted in the loss of this *Alu* insertion in Macaques and the fixation of it in all other clades. Green dot line represents the hypothetical polymorphic period of this *Alu* element and blue lines represent the clades that have this *Alu* element fixed present.

analyzed (dolloparsimony vs. maximum likelihood, phenetic analyses, and maximum parsimony with its various weighting schemes, etc.) reinforce the utility of phylogenetic analysis using SINEs.

Like all genetic systems, there are some drawbacks to the use of SINE insertion data. For example, models of insertion rates allow estimates of divergence times and tools analogous to genetic distances and molecular clocks have not yet been developed for SINE insertions. However, combinatorial approaches using sequence analysis of shared insertions may prove valuable in overcoming these drawbacks. Once a clade has been established using a group of shared insertions, the sequences of the elements themselves and flanking sequences can be safely assumed to have derived from a common

ancestor and standard molecular clock estimations can then be applied. As an extension, data generated for more traditional DNA-based sequence analyses can be imported regardless of whether the analyses of sequence data produced the same topology. Such analyses may help to resolve cases where differential substitution rates or long-branch attraction have caused problems in inferring the correct phylogeny.

In conclusion, mobile element based methods represent powerful tools for resolving phylogenetic relationships. As a result of the essentially homoplasy free nature and known ancestral state, many long standing phylogenetic controversies can be resolved using this approach. Our study represents the first large scale SINE based phylogenetic analysis of Old World monkeys. The

results indicate that SINE systems can be utilized to address phylogenetic questions from the family to species scale. More importantly our results represent an important step toward the construction of a mobile element based phylogeny of the entire Primate order.

Acknowledgments

This research was supported by National Science Foundation BCS-0218338 (M.A.B.), EPS-0346411 (M.A.B.), the State of Louisiana Board of Regents Support Fund (M.A.B.), and grants from the National Institutes of Health (C.B.S.). C. Huang was supported by a Howard Hughes Medical Institute grant through the Undergraduate Biological Sciences Education program to Louisiana State University. O. Ryder was supported by a National Science Foundation Grant BCS-0094993 to the Zoological Society of San Diego for the Integrated Primate Biomaterials and Information Resource (www.ipbir.org).

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410.
- Batzer, M.A., Deininger, P.L., 2002. *Alu* repeats and human genomic diversity. *Nat. Rev. Genet.* 3 (5), 370–379.
- Bigoni, F., Houck, M., Ryder, O., Wienberg, J., Stanyon, R., 2004. Chromosome painting shows that *Pygathrix nemaeus* has the most basal karyotype among Asian Colobinae. *Int. J. Primatol.* 25 (3), 679–688.
- Bigoni, F., Stanyon, R., Wimmer, R., Schempp, W., 2003. Chromosome painting shows that the proboscis monkey (*Nasalis larvatus*) has a derived karyotype and is phylogenetically nested within Asian Colobines. *Am. J. Primatol.* 60 (3), 85–93.
- Cantrell, M.A., Filanowski, 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 (2), 769–777.
- Deininger, P.L., Batzer, M.A., 1993. Evolution of Retroposons. In: Hecht, M.K. (Ed.), *Evolutionary Biology*, vol. 27. Plenum Press, New York, pp. 157–196.
- Deininger, P.L., Batzer, M.A., 1999. *Alu* repeats and human disease. *Mol. Genet. Metab.* 67 (3), 183–193.
- Deininger, P.L., Batzer, M.A., 2002. Mammalian retroelements. *Genome Res.* 12 (10), 1455–1465.
- Delson, E., 1992. Evolution of old world monkeys. In: Johns, J.S., Martin, R.D., Pilbeam, D., Bunney, S. (Eds.), *The Cambridge Encyclopedia of Human Evolution*. Cambridge University Press, Cambridge, pp. 217–222.
- Disotell, T.R., 1994. Generic level relationships of the Papionini (Cercopithecoidea). *Am. J. Phys. Anthropol.* 94 (1), 47–57.
- Disotell, T.R., 2003. Primates: Phylogenetics. *Encyclopedia of the Human Genome*. Nature Publishing Group, London.
- Disotell, T.R., Honeycutt, R.L., Ruvolo, M., 1992. Mitochondrial DNA phylogeny of the Old-World monkey tribe Papionini. *Mol. Biol. Evol.* 9 (1), 1–13.
- Edwards, M.C., Gibbs, R.A., 1992. A human dimorphism resulting from loss of an *Alu*. *Genomics* 14 (3), 590–597.
- 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 (3), 585–598.
- Groves, C., 2001. *Primate Taxonomy*. Pages Smithsonian Press, Washington, DC.
- Groves, C.P., 1993. Order Primates. *Mammal species of the world: a taxonomic and geographic*. In: Wilson, D.E., Reeder, D.M. (Eds.), Smithsonian Institution Press, Washington, DC, pp. 243–278.
- Hamdi, H., Nishio, H., Zielinski, R., Dugaiczky, A., 1999. Origin and phylogenetic distribution of *Alu* DNA repeats: irreversible events in the evolution of primates. *J. Mol. Biol.* 289 (4), 861–871.
- Harris, E.E., Disotell, T.R., 1998. Nuclear gene trees and the phylogenetic relationships of the mangabey (Primates: Papionini). *Mol. Biol. Evol.* 15 (7), 892–900.
- Hayasaka, K., Fujii, K., Horai, S., 1996. Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. *Mol. Biol. Evol.* 13 (7), 1044–1053.
- Hedges, D.J., Callinan, P.A., Cordaux, R., Xing, J.C., Barnes, E., Batzer, M.A., 2004. Differential *Alu* mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* 14 (6), 1068–1075.
- Hillis, D.M., 1999. SINEs of the perfect character. *Proc. Natl. Acad. Sci. USA* 96 (18), 9979–9981.
- Jablonski, N.G., Peng, Y.Z., 1993. The phylogenetic relationships and classification of the doucs and snub-nosed langurs of China and Vietnam. *Folia Primatol. (Basel)* 60 (1R-2), 36–55.
- Kent, W.J., 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12 (4), 656–664.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al., 2001. Initial sequencing and analysis of the human genome. *Nature* 409 (6822), 860–921.
- Messier, W., Stewart, C.B., 1997. Episodic adaptive evolution of primate lysozymes. *Nature* 385 (6612), 151–154.
- Murata, S., Takasaki, N., Saitoh, M., Okada, N., 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl. Acad. Sci. USA* 90 (15), 6995–6999.
- Nikaido, M., Matsuno, F., Hamilton, H., Brownell Jr., R.L., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A.M., Fordyce, R.E., Hasegawa, M., et al., 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA* 98 (13), 7384–7389.
- Nikaido, M., Rooney, A.P., Okada, N., 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* 96 (18), 10261–10266.
- Okada, N., 1991. SINEs. *Curr. Opin. Genet. Dev.* 1 (4), 498–504.
- Okada, N., Shedlock, A.M., Nikaido, M., 2004. Retroposon mapping in molecular systematics. *Methods Mol. Biol.* 260, 189–226.
- Page, S.L., Chiu, C., Goodman, M., 1999. Molecular phylogeny of Old World monkeys (Cercopithecoidea) as inferred from gamma-globin DNA sequences. *Mol. Phylogenet. Evol.* 13 (2), 348–359.
- Page, S.L., Goodman, M., 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabey and for a human–chimpanzee clade. *Mol. Phylogenet. Evol.* 18 (1), 14–25.
- Peng, Y.Z., Pan, R.L., Jablonski, N.G., 1993. Classification and evolution of Asian colobines. *Folia Primatol. (Basel)* 60 (1-2), 106–117.
- Ray, D.A., Xing, J.C., Hedges, D.J., Hall, M.A., Laborde, M.E., Anders, B.A., White, B.R., Stoilova, N., Fowlkes, J.D., Landry, K.E., et al., 2005. *Alu* insertion loci and platyrrhine primate phylogeny. *Mol. Phylogenet. Evol.* 35, 117–126.
- Roos, C., Schmitz, J., Zischler, H., 2004. Primate jumping genes elucidate strepsirrhine phylogeny. *Proc. Natl. Acad. Sci. USA* 101 (29), 10650–10654.
- Roy-Engel, A.M.E., El-Sawy, M., 2002. Evolutionary history of B1 retroposons in the genus *Mus*. *J. Mol. Biol.* 51 (5), 256–264.
- Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S.

- (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Source code Available at: <<http://fokker.wi.mit.edu/primer3/>> Humana Press, Totowa, NJ, pp. 365–386.
- Ryan, S.C., Dugaiczky, A., 1989. Newly arisen DNA repeats in primate phylogeny. *Proc. Natl. Acad. Sci. USA* 86 (23), 9360–9364.
- 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 (8), 1349–1361.
- Salem, A.H., Ray, D.A., Xing, J.C., 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 (22), 12787–12791.
- Schmitz, J., Roos, C., Zischler, H., 2005. Primate phylogeny: molecular evidence from retroposons. *Cytogenet. Genome Res.* 108 (1–3), 26–37.
- Shedlock, A., Takahashi, K., Okada, N., 2004. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol. Evol.* 19 (10), 545–553.
- Shedlock, A.M., Milinkovitch, M.C., Okada, N., 2000. SINE evolution, missing data, and the origin of whales. *Syst. Biol.* 49 (4), 808–817.
- Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays* 22 (2), 148–160.
- Soltis, D.E., Soltis, P.S., 2003. The role of phylogenetics in comparative genetics. *Plant Physiol.* 132 (4), 1790–1800.
- Swofford, D.L., 2003. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4.0b10 Sunderland, Massachusetts, Sinauer Associates.
- Takahashi, K., Terai, Y., Nishida, M., Okada, N., 1998. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. *Mol. Biol. Evol.* 15 (4), 391–407.
- Takahashi, K., Terai, Y., Nishida, M., Okada, N., 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. *Mol. Biol. Evol.* 18 (11), 2057–2066.
- Takasaki, N., Yamaki, T., Hamada, M., Park, L., Okada, N., 1997. The salmon *SmaI* family of short interspersed repetitive elements (SINEs): interspecific and intraspecific variation of the insertion of SINEs in the genomes of chum and pink salmon. *Genetics* 146 (1), 369–380.
- Tosi, A.J., Melnick, D.J., Disotell, T.R., 2004. Sex chromosome phylogenetics indicate a single transition to terrestriality in the guenons (tribe Cercopitheciini). *J. Hum. Evol.* 46 (2), 223–237.
- 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.
- Zhang, Y., Ryder, O., 1998. Mitochondrial cytochrome *b* gene sequences of Old World monkeys: With special reference on evolution of Asian colobines. *Primates* 39 (1), 39–49.
- Zietkiewicz, E., Richer, C., Labuda, D., 1999. Phylogenetic affinities of tarsier in the context of primate *Alu* repeats. *Mol. Phylogenet. Evol.* 11 (1), 77–83.