Alu insertion loci and platyrrhine primate phylogeny

David A. Raya, Jinchuan Xing, Dale J. Hedges, Michael A. Hall, Meredith E. Laborde, Bridget A. Anders, Brittany R. White, Nadica Stoilova, Justin D. Fowlkes, Kate E. Landry, Leona G. Chemnicking, Oliver A. Ryder, Mark A. Batzer

a Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Microsystems, 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

Received 12 August 2004; revised 11 October 2004

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 (marasmots, 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.

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 can occur across distantly related taxa as a function of evolutionary time and variable recombination 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 confirming 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 (Batzner and Deininger, 2002; Lander et al., 2001). Distinct families of Alu elements in the human genome have been described in detail (Batzner and Deininger, 1991, 2002; Batzer et al., 1991, 1995; Carroll et al., 2001; Salem et al., 2003; 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 Cebuellae, 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 (Callitrichus); and the spider, woolly, and howler monkeys, and muriqii (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): Cebidae (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 Dormann and Ruvolo, 1999). Unfortunately, as Schneider (2000) illustrates, 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–Pithecidae 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 are integrated into various New World monkey (NWM) genomes after divergence of the group from catarhines (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 157 sequences were identified using the NISC database as originating as part of Target 1, a ~1.5 MB region homologous to chromosome 7q31 in Homo sapiens. Alu elements were identified using RepeatMasker based analysis of the BAC sequences (A.F.A. Smit & P. Green, 1995; Schneider et al., 1993): the callitrichines (one insertion), monophyly of the callitrichines (one insertion), and a relatively close affinity 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 are integrated into various New World monkey (NWM) genomes after divergence of the group from catarhines (Old World monkeys and hominids).
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 \textit{Alu} element being investigated was also present in the human genome. The \textit{Alu} elements were then divided into two groups, ‘Old’ elements were defined as \textit{Alu} insertions present in both the human and the new world monkey genome under investigation. ‘Young’ elements were defined as \textit{Alu} insertions present in one or more of the platyrrhine primates but not present in \textit{Homo sapiens}. Oligonucleotide primers for polymerase chain reaction (PCR) amplification of each locus were designed for the young or lineage-specific \textit{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 \textit{Alu} insertions in the genomes of nine additional platyrrhine primates from across the taxonomic distribution. The taxa included were \textit{Aotus trivirgatus} (owl monkey), \textit{Saguinus labiatus} (red-bellied tamarin), \textit{Lagothrix lagotricha} (woolly monkey), \textit{Ateles geoffroyi} (black-handed spider monkey), \textit{Callithrix pygmaea} (pygmy marmoset), \textit{Pithecia p. pithecia} (northern white-faced saki), \textit{Callicebus d. donacophilus} (Bolivian gray titi), \textit{Pithecia p. pithecia} (northern white-faced saki), \textit{Alouatta sara} (Bolivian red howler monkey), \textit{Saimiri s. sciureus} (squirrel monkey), \textit{Aotus trivirgatus} (owl monkey), \textit{Saguinus labiatus} (red-bellied tamarin), \textit{Lagothrix lagotricha} (woolly monkey), \textit{Ateles geoffroyi} (black-handed spider monkey), \textit{Callithrix pygmaea} (pygmy marmoset), \textit{Pithecia p. pithecia} (northern white-faced saki), \textit{Callicebus d. donacophilus} (Bolivian gray titi), and \textit{Saimiri s. sciureus} (squirrel monkey). 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 \textit{MseI}, \textit{AcI}, and \textit{NdeI} (New England Biolabs, Beverly, MA) in 120\,\mu l reactions, followed by heat inactivation of the enzyme at 65 °C for 20 min. After several rounds, we began using the enzyme \textit{NdeI} exclusively because the six-base recognition site provided us with longer flanking unique sequences than did the four-base recognition sites of \textit{MseI} or \textit{AcI}. 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 (‘-GAATTCGTCACATAGCATTCT-3’) and one of several \textit{Alu}-specific primers (Table 3) to obtain partial \textit{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 \textit{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 \textit{Alu} element was not present at that locus in the human genome. If the same \textit{Alu} insertion was identified in the human genome, the insertion was considered an old shared character, having inserted prior to the platyrrhine–catarrhine divergence, and the \textit{Alu} element was
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

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aci-Top</td>
<td>CGGAAGGAGAGAGCCTGCTGTCAGGAG</td>
</tr>
<tr>
<td>Mse-Top</td>
<td>TAGAAGGAGGAGCAGCTGCTGTCAGG</td>
</tr>
<tr>
<td>Aci-Bottom</td>
<td>GAGCGAATTCGTAACATAGCATTTCTGTC</td>
</tr>
<tr>
<td>Mse-Bottom</td>
<td>CTTCCTTC</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5′–3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAlu1A</td>
<td>GGGAGATCGAGACGCTGTTGCGT</td>
</tr>
<tr>
<td>AAlu1B</td>
<td>CTGGTGGTGGGAGGCTGAGGCA</td>
</tr>
<tr>
<td>AAlu1D</td>
<td>CAAAGATCGAGCATCTCCGTGT</td>
</tr>
<tr>
<td>AAlu2D</td>
<td>TGCCGTGTAATATCGCTCTA</td>
</tr>
<tr>
<td>AAlu2E</td>
<td>CTGGGGATGCTGAGGCGGT</td>
</tr>
<tr>
<td>AAlu2B</td>
<td>ACTCCAGCTCGYRCCGT</td>
</tr>
<tr>
<td>AAlu2C</td>
<td>GGYGCCTGYRAGYRGAGTG</td>
</tr>
<tr>
<td>AAluR1</td>
<td>GCCCTCTGAGGTCACGCA</td>
</tr>
<tr>
<td>Fam1</td>
<td>GGAGAAATTGCTTGAACCACGAAAA</td>
</tr>
<tr>
<td>Fam2</td>
<td>CACTCCAGCTGCGGCT</td>
</tr>
<tr>
<td>Fam3</td>
<td>GCTGGAAGGAGAGAATTGCC</td>
</tr>
<tr>
<td>SAlu1</td>
<td>AGAGATCGAGACGCTGTTGGT</td>
</tr>
<tr>
<td>SAlu2A</td>
<td>CTGGGCGATGTTGGCGAAGT</td>
</tr>
</tbody>
</table>

### 2.3. PCR and sequencing

All primer pairs were initially tested on human DNA using a temperature gradient PCR (45–60 °C) to determine the appropriate annealing temperature for further testing on a non-human primate phylogenetic panel. This panel included four catarrhine taxa (*Homo sapiens*, *Pan paniscus*, *Chlorocebus aethiops sabaue*, and *Macaca mulatta*), one prosimian (*Lemur catta*), and the nine previously described platyrhine monkeys.

Twenty-five microliters PCR amplifications were performed under the following conditions: 10–50 ng template DNA, 7 pmol of each oligonucleotide primer, 200 nm dNTPs, in 50 mM KCl, 10 mM Tris–HCl (pH 8.4), 2.0 mM MgCl₂, 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 pre-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 (≥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 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 unanimous or formative 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 458 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 485), 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.

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
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 three 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 Dormann 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 they are not 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 third locus supports a close relationship between Callithrix and Saguinus, both members of subfamily Callitrichinae (Goodman et al., 2000), 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 between Alouatta (Canavez et al., 1999; Horovitz and Meyer, 1995; Schneider et al., 1993; von Dormann 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 oppor-
tunity 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 inser-
tions 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 uni-
form rate of Alu insertion in all taxa and given the sce-
nario described, more genus-specific insertions would be
expected in Aotus and Saimiri than in the two callitri-
chine 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 β-globin sequences place the divergence of the three
subfamilies at 17.5 mya. The same data suggest the diver-
gence between Aotinae and Cebinae occurred relatively
soon thereafter (16.7 mya). Finally, the tamarins and
marmosets are thought to have shared a common ances-
tor 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 pre-
cisely elucidate these relationships.

4.1. Parallel Insertions

Multiple Alu insertions within a single amplified locus
in taxa can lead to problems interpreting results in phy-
logenetic analyses that use SINE insertions. As the diver-
gence 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 sin-
gle 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 phyloge-
netic analysis. However, upon detailed DNA sequence
analysis each of these examples was resolved. Six of the
multiple insertions were instances where a secondary ele-
ment had retrotransposed into one or more of the catar-
rhine taxa when compared to the platyrrhines. Given the

time since divergence of these major groups these types
of multiple insertions are not unexpected. 670

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 contain-
ing the secondary elements. Locus 37 provides an exam-
ple of an instance in which two insertions occurred very
close together. In Ateles, a truncated Alu Sp [nomencla-
ture 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 686
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
689

pattern suggested that an Alu insertion was shared by 690
Pithecia, Ateles, Lagothrix, and Alouatta. Nevertheless, 691
DNA sequence analysis of the locus revealed that a sec-
ondary Alu insertion was present in Pithecia down-
stream of the originally identified insertion, which was 694
present only in the atelids.

Locus 58 was the most complex and most easily mis-
interpreted locus. Analysis by electrophoresis (see Fig. 4) 697
suggested that an Alu element was shared by macaque, 698
green monkey, marmoset, tamarin and squirrel monkey. 699

The catarrhine amplicons could easily be explained as a 700
secondary insertion; however, the pattern in Saimiri, 701
Callithrix, and Saguinus was another matter because it 702

represented a clear and not unlikely alternative to the 703
phylogeny suggested by other insertions. The anomal-
ous amplification pattern was determined to be due to the 704
independent insertion of three different Alu elements in 705
the same genomic location in addition to the element 706
originally identified in tamarin. First, an Alu Y element 707
integrated in two outgroup taxa, Chlorocebus and 708
Macaca, ~80 bp downstream of the originally identified
insertion, which was present in Saguinus and Callithrix. 709
A truncated Alu Sq/x element was identified in Aotus, 710

further downstream of the originally identified insertion. 711

A fourth insertion, of a full-length Alu Sq element, had 712
occurred in Saimiri at almost the same position as the 713
truncated element in Aotus. In fact, these two elements 714
have inserted in such close proximity that, like the Alu 715
elements at locus 37, they share a portion of their target 716
site duplications. Thus, locus 58 contained four indepen-
dent Alu insertions within the same 300 bp genomic 717
span.

A special case of multiple insertions at a locus is the
"precise" parallel forward insertion. This event is 721
defined as the insertion of a different Alu element at 722
an identical target site in different taxa. Only two instances 726
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 Platyrhini. 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 (Batzinger 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.

Acknowledgments

R. Cordaux and S. Herke contributed comments to earlier drafts of this manuscript. This research was funded by the National Science Foundation BCS-0218338 (M.A.B.). Bridget Anders was supported by a National Science Foundation REU supplement to award BCS-0218338. N. Stoilova and M. Laborde were supported by a Howard Hughes Medical Institute grant through the Undergraduate Biological Sciences Education program to Louisiana State University.
References


