Comparative analysis of *Alu* repeats in primate genomes

George E. Liu, Can Alkan, Lu Jiang, et al.

*Genome Res.* 2009 19: 876-885
Access the most recent version at doi:10.1101/gr.083972.108

**Supplemental Material**

http://genome.cshlp.org/content/suppl/2009/05/01/19.5.876.DC1.html

**References**

This article cites 36 articles, 18 of which can be accessed free at:
http://genome.cshlp.org/content/19/5/876.full.html#ref-list-1

Article cited in:
http://genome.cshlp.org/content/19/5/876.full.html#related-urls

**Email alerting service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

---

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions
Comparative analysis of Alu repeats in primate genomes

George E. Liu,1,6 Can Alkan,2,3 Lu Jiang,4 Shaying Zhao,5,6 and Evan E. Eichler2,3

1USDA, ARS, ANRI, Bovine Functional Genomics Laboratory, Beltsville, Maryland 20705, USA; 2Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, Washington 98195, USA; 3Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA; 4Department of Bioengineering, University of Maryland, College Park, Maryland 20742, USA; 5Department of Biochemistry and Department of Molecular Biology, University of Georgia, Athens, Georgia 30602, USA

Using bacteria artificial chromosome (BAC) end sequences (16.9 Mb) and high-quality alignments of genomic sequences (17.4 Mb), we performed a global assessment of the divergence distributions, phylogenies, and consensus sequences for Alu elements in primates including lemur, marmoset, macaque, baboon, and chimpanzee as compared to human. We found that in lemur, Alu elements show a broader and more symmetric sequence divergence distribution, suggesting a steady rate of Alu retrotransposition activity among prosimians. In contrast, Alu elements in anthropoids show a skewed distribution shifted toward more ancient elements with continual declining rates in recent Alu activity along the hominoid lineage of evolution. Using an integrated approach combining mutation profile and insertion/deletion analyses, we identified nine novel lineage-specific Alu subfamilies in lemur (seven), marmoset (one), and baboon/macaque (one) containing multiple diagnostic mutations distinct from their human counterparts—Alu J, S, and Y subfamilies, respectively. Among these primates, we show that that the lemur has the lowest density of Alu repeats (55 repeats/Mb), while marmoset has the greatest abundance (188 repeats/Mb). We estimate that ~70% of lemur and 16% of marmoset Alu elements belong to lineage-specific subfamilies. Our analysis has provided an evolutionary framework for further classification and refinement of the Alu repeat phylogeny. The differences in the distribution and rates of Alu activity have played an important role in subtly reshaping the structure of primate genomes. The functional consequences of these changes among the diverse primate lineages over such short periods of evolutionary time are an important area of future investigation.

[Supplemental material is available online at www.genome.org and at http://bfgl.anri.barc.usda.gov/Alusite.]
waves of fixation from sequential small subsets of master elements (Batzer and Deininger 2002).

To date, genome-wide characterization of Alu repeats in nonhuman primates has been limited to chimpanzee and macaque (The Chimpanzee Sequencing and Analysis Consortium 2005; Gibb et al. 2007). Most chimpanzee-specific elements belong to a subfamily (AluYc1) that is very similar to the source gene in the human–chimpanzee last common ancestor. In macaque, Alu elements have evolved into four currently active lineages: AluYRa1-4, AluYRb1-4, AluYRc1-2, and AluYRd1-4 (Han et al. 2007). Currently, there are three macaque consensus sequences: AluMacYa3, AluMacYb2, and AluMacYb4 in Repbase (Version 13.5). For other primate genomes, most studies have been based on PCR cross-amplification among diverse primate taxa and, therefore, are potentially biased to either conserved regions or limited to closely related species. Ray and Batzer (2005) recovered 48 NWM-specific Alu elements using a combination of PCR and computational approaches and reported three NWM-specific subfamilies: AluTa7, AluTa10, and AluTa1. In another publication, Herke et al. (2007) reported a few loci (such as DQ822065) from the lemur derived from PCR display. Initial comparative analysis based on small samples of primate genomic sequences demonstrated that the fixation rates of retroelements (especially SINE/Alu) vary radically in different primate lineages (Liu et al. 2003; Hedges et al. 2004). In this study, we analyze Alu elements in randomly sampled BAC end sequences (BES) and finished genomic sequence alignments (ALN) from five nonhuman primate species (Supplemental Table S1). We identified all Alu repeat elements whose insert length was ≥80% of the corresponding consensus sequence length (Table 2). Compared to all other primates analyzed in this study, the marmoset genome shows the greatest density of Alu repeats (188 repeats/Mb), while the lemur genome shows the least (55 repeats/Mb) (Table 2). In human RES, the density of Alu repeats is 104 repeats/Mb, which is lower than the genome-wide density of human Alu repeats at 315 repeats/Mb, mainly because of the short length of BES. We performed an all-by-all pairwise sequence divergence analysis of all available Alu elements within each species (210–718 Alu repeat elements) and computed the genetic distance among all alignments using the Kimura two-parameter model. We plotted the distribution of pairwise divergences within each species (Fig. 1B, with Kimura distance ≥0.10) also as a function of genetic distance. Notable differences among the K-plots were observed when lemur was compared to other primates. All anthropoids including human, great apes (chimpanzee), OWM (baboon and macaque), and NWM (marmoset) show a similar asymmetric divergence profile with a mode at 0.23 substitutions/ site and a relative small fraction of high-identity Alu repeat elements. In contrast, the lemur shows a broader, more symmetric distribution with a much greater abundance of highly identical (potentially evolutionarily “young”) Alu repeats when compared to other primates. A detailed inspection of the most identical Alu repeats (Fig. 1B, with Kimura distance <0.10) also provides evidence of a slight increase in the fraction of most elements in nonhuman primates, especially those lineage-specific Alu elements and/or those in more distantly related species like marmoset and lemur, may differ significantly from human consensus sequences; therefore, they may be difficult to recognize by RepeatMasker. To eliminate this bias and exclude the possibility of incomplete annotation, we separately analyzed all indels (insertions or deletions >100 bp) based on human–marmoset and human–lemur genomic sequence alignments using previously described methods (Liu et al. 2003). In total, we identified 1475 human and 1507 marmoset Alu elements from human–marmoset sequence alignments; 1569 human and 340 lemur Alu elements were identified from human–lemur alignments. No additional Alu repeats were identified based on our independent analysis of indels (>100 bp).

### Table 1. Alu elements in primate genomes sequences

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Accession count</th>
<th>Base pair</th>
<th>Total Lineage specific</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Human</td>
<td>NHP</td>
</tr>
<tr>
<td>Human–chimpanzee</td>
<td>51</td>
<td>4,938,130</td>
<td>4,883,663</td>
</tr>
<tr>
<td>Human–baboon</td>
<td>42</td>
<td>4,739,969</td>
<td>4,685,021</td>
</tr>
<tr>
<td>Human–marmoset</td>
<td>45</td>
<td>4,222,126</td>
<td>4,182,575</td>
</tr>
<tr>
<td>Human–lemur</td>
<td>29</td>
<td>3,615,410</td>
<td>2,885,250</td>
</tr>
</tbody>
</table>

Human–macaque comparison was not performed.

*Counts of Alu elements ≥80% of the corresponding consensus sequence length.

See Results. An additional analysis was performed on lemur Alu elements using the Alucode developed by Pevzner and colleagues (Price et al. 2004).
identical Alu repeats (<0.01) in human as compared to chimpanzee, consistent with previous observations (Liu et al. 2003; Hedges et al. 2004; Watanabe et al. 2004; The Chimpanzee Sequencing and Analysis Consortium 2005). Similar K-plots were obtained for Alu elements derived from finished primate genomic sequences (data not shown).

Characterization of lineage-specific Alu repeat elements from BAC end sequences

We used two distinct approaches to study lineage-specific Alu subfamilies. First, we categorized Alu subfamilies using the program Alucode (Price et al. 2004). Based on our analysis of 2128 Alu repeats from six primate species, we identified 18 distinct subfamilies: subfamily composition ranges from 15 to 691 with most subfamilies containing 50–100 elements (P-value for subfamily partition ranges from $2 \times 10^{-30}$ to $2 \times 10^{-37}$) (see Price et al. 2004 for the P-value definition and calculation). We next constructed a minimum spanning (MS) tree for these 18 Alu subfamilies to summarize their evolutionary relationship (Fig. 2). We identified 11 subfamilies shared among different species (Nodes 1–11) and seven putative lineage-specific subfamilies (Nodes 12–18, named BES_MS_BM1, BES_MS_R1-2, and BES_MS_L1-4).

As a second method, we constructed Alu neighbor-joining (NJ) trees independently for genomic sequences from lemur (Supplemental Fig. S3) and marmoset (Supplemental Fig. S4) as well as from all six primate species including human (Supplemental Fig. S5). We used the tree topology to cluster related Alu elements into groups. The groups were named as follows: lemur (BES_NJ_L1-12), marmoset (BES_NJ_R1-11), and baboon/macaque (BES_NJ_BM1). The analysis clearly identified monophyletic clades that appear lineage specific with modest bootstrap support (Supplemental Fig. S5). These six putative lineage-specific subfamilies are lemur's BES_NJ_L10–12 (green, labeled as "Lemur Alu J"), marmoset's BES_NJ_R10–11 (purple, labeled as "Marmoset Alu S"), and baboon/macaque's BES_NJ_BM1 elements (red, labeled as "Baboon/macaque Alu Y"). Based on the majority rule, Alu consensus sequences were derived from each group. We constructed a NJ tree using all derived Alu consensus sequences with known primate Alu consensus sequences (Supplemental Fig. S6).

Characterization of lineage-specific Alu repeat elements from orthologous sequence alignments

As a second source of data, we constructed optimal global sequence alignments between finished nonhuman primate genomic BAC clones and the human genome reference sequence using previously described methods (Liu et al. 2003; She et al. 2006). We generated a total of 51 human–chimpanzee, 42 human–baboon, 45 human–marmoset, and 29 human–lemur genomic alignments (Table 1; http://bfgl.anri.barc.usda.gov/Alusite). Based on these alignments, we classified all Alu elements into two categories (lineage specific or shared) based on the presence or absence of an ~300-bp insertion deletion event within the alignment. We limited our analysis to full-length Alu repeats that are not chimeric (single subfamily designation) and show flanking target site

Table 2. Alu elements in primate BAC end sequences

<table>
<thead>
<tr>
<th>Species</th>
<th>Lemur</th>
<th>Marmoset</th>
<th>Baboon</th>
<th>Macaque</th>
<th>Chimpanzee (+ Riken)*</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>BES sequence</td>
<td>6533</td>
<td>5173</td>
<td>7303</td>
<td>5504</td>
<td>5969 (154,071)</td>
<td>743,245</td>
</tr>
<tr>
<td>Total length (bp)</td>
<td>3,798,199</td>
<td>3,825,700</td>
<td>3,670,302</td>
<td>2,873,380</td>
<td>2,784,861 (118,252,885)</td>
<td>354,136,231</td>
</tr>
<tr>
<td>Alu count</td>
<td>513</td>
<td>1437</td>
<td>1520</td>
<td>986</td>
<td>848 (28,835)</td>
<td>111,411</td>
</tr>
<tr>
<td>Alu count/Mb</td>
<td>464</td>
<td>1404</td>
<td>1481</td>
<td>956</td>
<td>816 (27,969)</td>
<td>108,283</td>
</tr>
<tr>
<td>Alu 80% count</td>
<td>122</td>
<td>367</td>
<td>404</td>
<td>333</td>
<td>293 (237)</td>
<td>306</td>
</tr>
<tr>
<td>Alu 80% count/Mb</td>
<td>210</td>
<td>718</td>
<td>524</td>
<td>348</td>
<td>229 (9524)</td>
<td>36,888</td>
</tr>
</tbody>
</table>

*Counts in parentheses included the chimpanzee BES data set from the Riken Institute.

Counts of Alu elements $\geq 80\%$ of the corresponding consensus sequence length.

Figure 1. (A) Sequence divergences of Alu elements. (B) An enlarged view for Kimura Distances <0.10.
human) has experienced a 4.6-fold increase in Alu activity when compared to prosimians (Table 1). Finally, we generated a minimal spanning tree using Alu elements derived from human–lemur, human–marmoset genomic sequences. Similar to the BES analysis (Fig. 2), we identified three marmoset- and four lemur-specific Alu subfamilies with statistical significance (named ALN_MS_R1-3, ALN_MS_L1-4 in Supplemental Fig. S7A,B), respectively.

Subfamily consensus sequences and phylogeny

Table 3 summarizes all 26 putative lineage-specific Alu subfamilies identified using four combinations of data (ALN vs. BES) and methods (NJ vs. MS) in the three nonhuman primate species while seven were lineage-specific: (black) baboon–macaque; (green) marmoset; (red) lemur. The number of Alu elements (in parentheses) and the P-value within each group are indicated.

duplications. We assume that the majority of 300-bp insertions arise as a result of new retrotransposition events as opposed to precise deletion of the repeat. The term “lineage specific” is relative only to the two species being compared. We constructed NJ trees based on multiple sequence alignments of these lineage-specific Alu repeat elements (Fig. 3A,B) and Alu subfamily consensus sequences (Repbase).

The phylogenetic analysis of lineage-specific Alu repeats derived from human–baboon and human–chimpanzee orthologous sequence alignments reveals three different categories of repeat (Fig. 3A): (1) an interleaved set of divergent human- and baboon-specific copies that are equivalent in number between the two species; (2) a monophyletic set of chimpanzee- and human-specific repeats with high sequence similarity to recently active AluY (Y lineage), Ya5/8 (ALN_NJ_H1), and Yb8/9 (ALN_NJ_H2) subfamilies; and (3) a more abundant set of baboon-specific AluY elements (ALN_NJ_B1 and ALN_NJ_B2) including both ancestral and young elements. There have been 60% more baboon-specific Alu retrotransposition events as a result of the expansion of the third category (Table 1).

A similar topology was obtained from Alu phylogenetic trees constructed from human and marmoset orthologous sequence alignments (Fig. 3B): We identified (1) an interleaved group of divergent human and marmoset repeats that are related to AluS consensus sequences; (2) a monophyletic marmoset-specific AluS/Sc lineage (ALN_NJ_R1); and (3) a human-specific AluY set (human AluY, ALN_NJ_H3). The last two lineages showed significant bootstrap support. By count, once again, marmoset-specific elements were 70% more abundant than human-specific elements (Table 1).

Although human–lemur genomic sequence alignments are complicated by greater sequence divergence between the two genomes, we identified only four pairs of Alu repeats as orthologous from a total of 1569 human and 340 lemur annotated Alu repeats. These data suggest that the anthropoid lineage (represented by

Figure 2. The minimum spanning tree of 18 Alu subfamilies. The tree is based on an Alucode analysis of 2128 Alu repeats extracted from primate BES data. (Blue) Eleven families were shared among human and at least one nonhuman primate species while seven were lineage specific: (black) baboon–macaque; (green) marmoset; (red) lemur. The number of Alu elements (in parentheses) and the P-value within each group are indicated.
S2). In Figure 5, we compare lemur consensus sequences with human AluSc and Sp, marmoset Alu subfamilies have 14–18 distinct nucleotide changes and an insertion of 3–6 nt between positions 264 and 269 (Supplemental Fig. S12). As discussed above, six marmoset subfamilies (Fig. 4A, gray bracket 3) are essentially the same as AluTa15 sharing almost all its diagnostic nucleotides (Ray and Batzer 2005). One marmoset Alu subfamily (ALN_MS_R1) is related to AluTa10 with a few more mutations and can be assigned as AluTa14 (Supplemental Table S2). BES_NJ_BM1/BES_MS_BM1 consensus is close to human Ye2/5 subfamilies. It is identical to AluMacYa3 with the exception of a transition from “G” to “A” at the position 205 (Supplemental Fig. S13). Thus, it can be assigned as Alu-MacYa4.

We also performed an age/divergence distribution analysis of all currently available lemur sequences using these seven lineage-specific Alu consensus sequences (Landey et al. 2001). The divergence levels reported by RepeatMasker were corrected by the CpG content of each repeat. We plotted the divergence distribution either by summing all seven subfamilies or separately for each subfamily (Fig. 6, bin size = 0.01). In the stacking plot (Fig. 6A), two bursts in Alu amplification can be detected (around 0.05 and 0.08 substitutions/site) and estimated to occur 20 and 32 Mya assuming a substitution rate of 2.5 × 10⁻⁹ substitutions/site per year (Price et al. 2004). Notable differences among the distributions are observed when each subfamily is considered: AluL and AluLa subfamilies are the major divergence profiles that are likely responsible for the two bursts; other minor profiles include AluL5, AluL6, and AluL9, which derived from AluL, while AluLa7a and AluLa7b, which are the youngest subfamilies, derived from AluLa. These results generally agree well with the MS trees in terms of age and fractions (Fig. 4B) and verified the relationship among these seven subfamilies. However, the multiple modes of these distribution profiles suggest that these seven subfamilies may still represent a mixed population and could be further divided into distinct subfamilies when more sequences are available.

**Discussion**

In this project, we performed a global characterization of Alu elements in diverse primate genomes using an integrated approach combining phylogenetic (N) and MS trees and insertion/deletion analysis of orthologous genomic alignments. Our analyses were...
Table 3. Counts of lineage-specific Alu subfamilies

<table>
<thead>
<tr>
<th>Method</th>
<th>Genomic sequence alignments (ALN)</th>
<th>BES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NJ</td>
<td>MS</td>
</tr>
<tr>
<td>Human</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Baboon</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>Marmoset</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Lemur</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

The human–chimpanzee shared subfamilies are not included.

* Baboon shared them with macaque.

These seven lemur subfamilies are derived from both BES and genomics sequences using Alucode.

expansions of retroviral inserted within the genomes of African great apes but not in humans and orangutans (Yohn et al. 2005).

A few exceptions in our phylogenetic analyses shed further insight on the evolutionary forces that shaped Alu elements. We observed, for example, a small subset of lineage-specific events that share diagnostic mutational differences with more ancient Alu repeat elements. Such elements may represent perfect deletions of more ancient elements, perhaps as a result of non-allelic homologous recombination, gene conversion events between Alu events, or the low levels of recent activity of the older subfamilies. We characterized nine new lineage-specific Alu consensus sequences in more diverse primate genomes: seven subfamilies in lemur: AluL, AluL5, AluL6, AluL9, AluLa, AluLa7a, and AluLa7b; one in marmoset: AluTa14; and one in baboon/macaque: AluMacYa4. The phylogenetic clustering of these Alu subfamilies according to species support that they were lineage-specific master genes for Alu amplification in these nonhuman primates. The nine new lineage-specific Alu subfamilies expand our understanding of Alu evolution and their impact on primate genome architecture.

Earlier studies using PCR and bioinformatics strategies also confirmed our discoveries. Our results showed that recent lemur-specific Alu consensus sequences (AluLa) contain a distinct poly(A) linker between the left and right Alu monomers. It agreed with previous data using Alu PCR amplification from lemur, sifaka, and galago (Zietkiewicz et al. 1998). Deininger and colleagues also had similar observations for active galago Alu elements (Daniels and Deininger 1983, 1991). However, it is difficult to associate those limited individual lemur loci (such as DQ822065 amplified by Herke et al. [2007]) with our lemur-specific consensus sequences at this stage. More prosimian sequence data are needed to make a meaningful comparison possible. A comparison with Ray and Batzer (2005) demonstrated that multiple subfamilies identified in NWM are essentially identical to AluTa1S sharing most of its diagnostically marked motifs. Our results derived from a larger subset of Alu elements (446 sequences from both BES and genomic sequences) further confirmed that the AluTa1S subtype expanded later in NWM evolution and may have arisen from AluTa7 or AluTa10 (177 sequences).

In summary, our analysis has provided an evolutionary framework for further classification and refinement of the Alu repeat phylogeny. The differences in the distribution and rates of Alu activity have played an important role in subtly reshaping the structure of primate genomes (Bailey et al. 2003). The functional consequences of these changes among the diverse primate lineages over such short periods of evolutionary time are an important area of future investigation.

Methods

Genomic sequence alignment and analyses

BAC libraries were constructed in Peter de Jong’s laboratory at Children’s Hospital Oakland Research Institute, Oakland, CA (http://www.chori.org/bacpac/) for the common chimpanzee (Pan troglodytes CH251), the (olive) baboon (Papio anubis RP41), the rhesus macaque (Macaca mulatta CH250), and the common marmoset (Callithrix jacchus CH259), while the lemur BAC library (Lemur catta LB2) was constructed by Jan-Fang Cheng’s laboratory at Lawrence Berkeley National Laboratory. Large genomic sequences (>50 kb in length) from chimpanzee (RP43), baboon (RP41), marmoset (CH259), and lemur (LB2) were retrieved from GenBank. Orthologous sequence relationships were identified, and optimal global alignments were constructed and validated as
Figure 4. Phylogenetic trees of primate lineage-specific Alu consensus elements. (A) Neighbor-joining tree: All branches are labeled with the bootstrap values (>50%) with n = 1000 replicates. (B) Minimum spanning tree. The color and label schemes are as described in Figure 2.
described previously (Liu et al. 2003). In total, we examined 51 loci (5.0 Mb) for human–chimpanzee, 42 loci (5.0 Mb) for human–baboon, 45 loci (4.0 Mb) for human–marmoset, and 29 loci (2.8 Mb) for human–lemur genomic sequence alignments (She et al. 2006). Large gaps (>100 bp) in these pairwise alignments were subdivided into one of two categories based on their association with a repeat sequence as described previously (Liu et al. 2003). Briefly, we classified an indel as a retrotransposition if at least 80% of the indel contained one predominant repeat (LINE, SINE, LTR). We considered the known interspersed repeat phylogeny based on the established repeat subfamilies (Smit 1999). For L1 and Alu elements, insertion sequences were examined for the presence of target-site duplications and a polyadenylation tail at the site of integration. The directionality of these retrotransposition events were unambiguously assigned to a specific lineage.

**BAC end sequencing**

We generated 24,513 BAC end sequences from 12,200 randomly sampled clones as part of an effort to randomly sample sequence from a diversity panel of primate genomes (BES originally generated at The Institute for Genomic Research, Supplemental Table S1; sequence and quality data are downloadable at http://bfgl.anri.barc.usda.gov/Alusite/). DNA sequence was isolated from single-colony-derived templates and prepared as described previously (Zhao et al. 2000). With the exception of the marmoset, the average Q20 length was 433.5 bp (Supplemental Table S1). Marmoset BES of higher quality were produced with improved sequencing techniques, as described previously (Zhao et al. 2001). Table 2 includes extra chimpanzee BES from the Riken Institute (Fujyama et al. 2002) and extra human BES (Lander et al. 2001). For Figure 6, besides the BES generated in this study, we also included 10,101 lemur (BES and whole-genome shotgun) reads and 43 lemur accessions assembled from 116,761 shotgun reads.

**Alu-element identification and phylogenetic analyses**

We initially detected Alu repeat elements using the slow search option (-s of RepeatMasker version 2002/07/13) with Repbase (http://www.girinst.org/, version 9.04). Owing to the variable lengths of poly(A) tails (Batzer and Deininger 2002), the default human consensus sequences were trimmed at their 3' poly(A) until only five bases of adenine remained. We selected all Alu repeats with at least 80% length of the consensus repeat. We then examined those indels that were not captured by RepeatMasker. None of these indels displayed any grouping or any Alu distinct features based on either length (>300 bp) or sequence identity (including diagnostic mutations). Therefore, we were convinced that the default human consensus library is sufficiently robust to identify Alu elements in other primates.

**Pairwise sequence alignments and divergences of Alu elements**

We computed the sequence divergences of Alu elements from the consensus sequences provided by RepeatMasker. Divergence levels reported by RepeatMasker were corrected for the CpG content of each repeat by $D_{CpG} = D/(1 + 9F_{CpG})$. Distribution histograms were plotted using a 0.01 bin size. For major branches within phylogenetic trees, multiple sequence alignments were performed with ClustalW at the default setting. The consensus sequences were derived using the simple majority rule. Degenerated
nucleotides were defined according to the standard IUPAC codes. MEGA (Kumar et al. 2001) was used to construct NJ trees using the Kimura two-parameter model. The minimum spanning trees of primate Alu subfamilies, that is, the trees with Alu subfamilies as nodes that minimize the sum of edge distances, were constructed using Alucode. Under the null hypothesis of uniformity, the P-value for the linkage was calculated using the nonparametric computation as described by Price et al. (2004). Since Alucode can run on a wide range of resolutions, it can split a small dataset into large numbers of subfamilies. Based on the size of our data, we chose MINCOUNT = 15 with all other default parameters. Under this setting, Alucode created similar numbers of Alu subfamilies as the conventional NJ method.

Acknowledgments

This work was supported in part by CRIS Project No. 1265-31000-090-00D from USDA and by National Institutes of Health grant GM058815 to E.E.E. E.E.E. is an investigator of the Howard Hughes Medical Institute.

References


shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. *Genome Res.* **16**: 576–583.


Received July 31, 2008; accepted in revised form September 30, 2008.