Under the genomic radar: The Stealth model of Alu amplification

Kyudong Han  
*Louisiana State University*

Jinchuan Xing  
*Louisiana State University*

Hui Wang  
*Louisiana State University*

Dale J. Hedges  
*Louisiana State University*

Randall K. Garber  
*Louisiana State University*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.lsu.edu/biosci_pubs](https://digitalcommons.lsu.edu/biosci_pubs)

**Recommended Citation**


This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Digital Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Digital Commons. For more information, please contact ir@lsu.edu.
Authors
Kyudong Han, Jinchuan Xing, Hui Wang, Dale J. Hedges, Randall K. Garber, Richard Cordaux, and Mark A. Batzer

This article is available at LSU Digital Commons: https://digitalcommons.lsu.edu/biosci_pubs/214
Under the genomic radar: The Stealth model of Alu amplification

Kyudong Han,1 Jinchuan Xing,1 Hui Wang, Dale J. Hedges, Randall K. Garber, Richard Cordaux, and Mark A. Batzer2

Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, Baton Rouge, Louisiana 70803, USA

Alu elements are the most successful SINEs (Short INterspersed Elements) in primate genomes and have reached more than 1,000,000 copies in the human genome. The amplification of most Alu elements is thought to occur through a limited number of hyperactive “master” genes that produce a high number of copies during long evolutionary periods of time. However, the existence of long-lived, low-activity Alu lineages in the human genome suggests a more complex propagation mechanism. Using both computational and wet-bench approaches, we reconstructed the evolutionary history of the AluYb lineage, one of the most active Alu lineages in the human genome. We show that the major AluYb lineage expansion in humans is a species-specific event, as nonhuman primates possess only a handful of AluYb elements. However, the oldest existing AluYb element resided in an orthologous position in all hominoid primate genomes examined, demonstrating that the AluYb lineage originated 18–25 million years ago. Thus, the history of the AluYb lineage is characterized by ~20 million years of retrotranspositional quiescence preceding a major expansion in the human genome within the past few million years. We suggest that the evolutionary success of the Alu family may be driven at least in part by “stealth-driver” elements that maintain low retrotranspositional activity over extended periods of time and occasionally produce short-lived hyperactive copies responsible for the formation and remarkable expansion of Alu elements within the genome.

[The sequence data from this study have been submitted to GenBank under accession nos. AY791249–AY791290.]

As the most successful SINEs (Short INterspersed Elements) in primate genomes, Alu elements have enjoyed remarkable proliferation during the primate radiation and have expanded to more than 1,000,000 copies in the human genome (Lander et al. 2001; Batzer and Deininger 2002). Despite considerable progress in the understanding of their biology and distribution throughout primate taxa (Salem et al. 2003b; Singer et al. 2003; Carter et al. 2004; Hedges et al. 2004; Otieno et al. 2004; Roos et al. 2004; Ray et al. 2005), a great deal of uncertainty still remains concerning their strategy for survival. It has been generally accepted that the amplification of most Alu elements occurs through a small number of long-lived high activity “master” genes (Deininger et al. 1992; Deininger and Batzer 1993), although there is considerable debate as to the details of this amplification strategy (Materia et al. 1990; Schmid 1993; Batzer et al. 1995; Cordaux et al. 2004; Price et al. 2004). In this model, the mutations accumulated in the master genes are inherited by the copies they produced, and consequently, a series of hierarchical Alu subfamilies that share the novel diagnostic mutation(s) are generated (Slagel et al. 1987; Willard et al. 1987; Britten et al. 1988; Deininger and Slagel 1988; Jurka and Smith 1988; Deininger et al. 1992; Batzer and Deininger 2002). On the other hand, the evolutionary history of the AluYa5 lineage, one of the most active human Alu lineages, suggests that the founder gene of this Alu lineage existed long before the major expansion of the lineage within the human genome (Leeflang et al. 1993). Contrary to the prediction of the master gene model, it has maintained low retrotranspositional activity, and this founder gene itself may not be directly responsible for the propagation of the recent human AluYa5 elements (Shaikh and Deininger 1996). Thus, these studies suggest that the expansion of Alu elements may follow a more complex propagation mechanism. Unfortunately, aside from the AluYa5 lineage, little data exist concerning the evolutionary origin of other Alu subfamilies in humans, making it difficult to assess the evolutionary significance of the results reported in the original AluYa5 lineage studies.

To gain additional insight into Alu subfamily propagation, we reconstructed the evolutionary history of the AluYb lineage, one of the largest and most active Alu lineages in the human genome (Jurka 1993; Carter et al. 2004) that comprises ~40% of the human-specific Alu elements (Hedges et al. 2004). This lineage, originally termed Sb2, is characterized by a 7-n tandem duplication involving positions 246 through 252 of the AluY consensus sequence (Jurka 1993; Batzer et al. 1996). The AluYb lineage in the human genome is subdivided into three main subfamilies as follows: AluYb7, AluYb8, and AluYb9 (Batzer et al. 1996; Roy-Engel et al. 2001; Jurka et al. 2002; Carter et al. 2004) based on diagnostic mutations following the standardized nomenclature for Alu repeats (Batzer et al. 1996). The majority of the human AluYb elements integrated into the genome during the last 3 to 4 million years (Myr) and reached a total copy number of about 2000 elements (Carter et al. 2004). The human diseases caused by de novo AluYb insertions suggest that this subfamily is currently actively retrotransposing, and a comprehensive analysis of the Yb lineage indicated that about 20% of AluYb elements are poly-
morphic in human genome (Muratan et al. 1991; Oldridge et al. 1999; Carter et al. 2004). Previous studies suggested that the evolutionary history of the AluYb lineage is much older than its period of major expansion in the human genome, and AluYb elements have also been identified in other nonhuman primates (Zietkiewicz et al. 1994; Gibbons et al. 2004). Nevertheless, the extent to which these elements are distributed among nonhuman primate species remains undetermined.

Using both a computational approach and polymerase chain reaction (PCR) display methodology, we have determined the distribution of AluYb elements in different hominoid genomes. We find that the long-term evolutionary history of the AluYb lineage exhibits a pattern that is remarkably similar to that of the AluYa lineage. Thus, the evolution of the AluYa and AluYb lineages illustrates a common strategy for Alu element proliferation. We propose a model for the expansion and survival of Alu elements in the primate order.

Results

AluYb elements in the common chimpanzee genome

To determine the evolutionary history of the AluYb lineage, we first computationally retrieved all of the AluYb elements from the first draft of the common chimpanzee (Pan troglodytes) genomic sequence (panTro1 Nov. 2003 assembly). A total of 12 AluYb elements were identified and subjected to PCR amplification on a common chimpanzee population panel and a separate primate panel composed of human and eight additional nonhuman primates (see Methods). The number of AluYb elements recovered from the chimpanzee draft sequence is in good agreement with a previous study (Gibbons et al. 2004). Detailed information on each locus, including primer sequences, annealing temperature, PCR product sizes, and chromosomal locations are shown in Table 1. Among the 12 elements identified, 10 belong to the AluYb8 subfamily, while the other two are non-AluYb7/8/9 elements (Table 2). Of the 10 AluYb8 elements, four loci were specific to the common chimpanzee lineage, including three elements that were polymorphic within the chimpanzee population panel (Figs. 1, 2B). Five elements were shared between the pygmy chimpanzee and common chimpanzee lineage (Fig. 2C) and one was present within human, chimpanzee, and gorilla genomes. These results suggest that the initial expansion of the AluYb8 subfamily predates the divergence of gorillas and humans/chimpanzees, which is thought to have occurred ~7 million years ago (Mya) (Goodman et al. 1998). In addition, two non-AluYb7/8/9 elements displayed PCR amplicon sizes consistent with the presence of an Alu element in all of the hominoid primates we tested.

To confirm that the PCR products were derived from authentic AluYb elements, all of the filled amplicons were cloned and sequenced. This sequencing effort revealed additional insight into the evolutionary history of the following four loci: Pan1, Pan2, Pan3, and Pan4. At the Pan1 locus, an AluYb8 element was inserted into the chimpanzee lineage (pygmy and common) after the divergence of humans and chimps. However, the pygmy chimpanzee and common chimpanzee showed both filled and empty size amplicons in their PCR amplifications, suggesting the presence of an Alu insertion polymorphism in both species (Fig. 3A). Sequence analysis of the locus showed that the AluYb8 element inserted immediately upstream of an AluX element in the same orientation, and that the integration site is partially shared (three copies of direct repeats). More recently, a nonhomologous recombination or intrachromosomal recombination event in the common chimpanzee genome generated a hybrid Alu element that is composed of the first half of the older AluX element and the second half of the newly inserted AluYb8 element. In contrast, the smaller allele amplified in other primates appeared to be the reintegaration site of the AluYb8 element (Fig. 3B).

DNA sequence analysis of the Pan2 and Pan3 loci resulted in the recovery of two gene conversion events. The Pan2 locus appears to be the oldest Alu element that we recovered. PCR analysis of the locus showed the presence of an Alu element or filled

<table>
<thead>
<tr>
<th>Name</th>
<th>Forward primer sequence (5'-3')</th>
<th>Reverse primer sequence (5'-3')</th>
<th>Chr. loc. in human</th>
<th>Direct repeat</th>
<th>A. T.</th>
<th>Filled</th>
<th>Empty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan1*</td>
<td>ACAGAAATCCGATGCTCTTCTGTTT</td>
<td>CTCAGTTTTCCTAATCCACTCTCTTCT</td>
<td>X:93270745</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan2*</td>
<td>CCGTGGCTAGTATGACTTTTCTTCT</td>
<td>GCTGTGAAGAATGCTTTCTTCTTTTCTT</td>
<td>4:75958275</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan3*</td>
<td>CGATGATCCTGGTCTTCTTCTTCTT</td>
<td>TGCCAGAGAAGAAGACTTTTCTTCTT</td>
<td>18:66254955</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan4*</td>
<td>AGGCGAGTACAAGTGGATCTTGTA</td>
<td>TTCTTCCCAATGTTTCTTCTTCTTCTT</td>
<td>3:155790154</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan5*</td>
<td>GCCCGTGTCGCTTCGTTCTTCTT</td>
<td>TCGTTCACTTTAGTTCTCTCTTCTT</td>
<td>5:17287078</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan6*</td>
<td>GCACTTTATGACCTTTCTTTCTTCTT</td>
<td>CATGCTTGACATGTTCTCTCTTCTT</td>
<td>10:97871078</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan7*</td>
<td>GATGGCTGGTTTCTTCTTCTTCTT</td>
<td>TCAGATGTGACTGTGTCTCTCTTCTT</td>
<td>2:183429337</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan8*</td>
<td>CGCGAATTCGCTTTCTCTTCTTCTT</td>
<td>AGCCAGATTTGCGTTCTCTCTTCTT</td>
<td>3:49401325</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan9*</td>
<td>CGCGAATTCGCTTTCTCTTCTTCTT</td>
<td>AGCCAGATTTGCGTTCTCTCTTCTT</td>
<td>6:101149947</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan10*</td>
<td>ATGAGTTGTTTCTTCTTCTTCTTCTT</td>
<td>TTGCTGCTCAAGATGTTCTCTCTTCTT</td>
<td>14:52808462</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan11*</td>
<td>Agagctggctcttttctttctttcttt</td>
<td>ttcagcttcggatgatggctttcttt</td>
<td>10:10509278</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Pan12*</td>
<td>ATGAGTTGTTTCTTCTTCTTCTTCTT</td>
<td>TTGCTGCTCAAGATGTTCTCTCTTCTT</td>
<td>6:101149947</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Gorilla7*</td>
<td>TGAGATGCTTTTCTTCTTCTTCTT</td>
<td>TCTGCTGCTATTGTTCTCTCTTCTT</td>
<td>3:49401325</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Gorilla8*</td>
<td>TGGCAGATTTGCGTTCTCTCTTCTT</td>
<td>AGCCAGATTTGCGTTCTCTCTTCTT</td>
<td>10:22965741</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Gorilla16*</td>
<td>ATGAGTTGTTTCTTCTTCTTCTTCTT</td>
<td>TTGCTGCTCAAGATGTTCTCTCTTCTT</td>
<td>14:52808462</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Gorilla19*</td>
<td>ATGAGTTGTTTCTTCTTCTTCTTCTT</td>
<td>TTGCTGCTCAAGATGTTCTCTCTTCTT</td>
<td>14:52808462</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
<tr>
<td>Gorilla21*</td>
<td>ATGAGTTGTTTCTTCTTCTTCTTCTT</td>
<td>TTGCTGCTCAAGATGTTCTCTCTTCTT</td>
<td>14:52808462</td>
<td>AAAACATCCC</td>
<td>55</td>
<td>892</td>
<td>561</td>
</tr>
</tbody>
</table>

(*) Retrieved from chimpanzee genome draft sequence; (#) identified using PCR display methodology.

*Chromosomal location.

Annealing temperature.
Table 2. Primate phylogenetic PCR panel results for chimpanzee AluYb elements

<table>
<thead>
<tr>
<th>Name</th>
<th>Chimpanzee panel</th>
<th>Human (HeLa)</th>
<th>Pygmy chimpanzee</th>
<th>Gorilla</th>
<th>Orangutan</th>
<th>Gibbon</th>
<th>Siamang</th>
<th>Green monkey</th>
<th>Owl monkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan1</td>
<td>Fixed present</td>
<td>–</td>
<td>Polymorphic</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td>(AluYb8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Pan2</td>
<td>Fixed present</td>
<td></td>
<td>+ (AluYb6)</td>
<td>+ (AluYb6)</td>
<td>+ (AluYb6)</td>
<td>+ (AluYb6)</td>
<td>+ (AluYb6)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan3</td>
<td>Fixed present</td>
<td></td>
<td>+ (AluYb8)</td>
<td>+ (AluYb8)</td>
<td>+ (AluYb8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan4</td>
<td>Fixed present</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan5</td>
<td>Polymorphic</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan6</td>
<td>Fixed present</td>
<td></td>
<td>–</td>
<td>+ (AluYb8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan7</td>
<td>Fixed present</td>
<td>–</td>
<td>+ (AluYb8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan8</td>
<td>Fixed present</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan9</td>
<td>Fixed present</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan10</td>
<td>Fixed present</td>
<td></td>
<td>+ (AluYb8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan11</td>
<td>Polymorphic</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan12</td>
<td>Polymorphic</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(AluYb8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(+/-) Presence/Absence of Alu element; (x) no amplification.

The Pan11 locus also contains a gene conversion event (Fig. 4B). Siamang, gibbon, and orangutan possess an Alu element that does not contain the AluYb lineage diagnostic duplication and has the highest sequence identity to the AluY consensus. However, the Alu elements in the gorilla, chimpanzee, and human genomes belong to the AluYb lineage, as indicated by the presence of the AluY diagnostic duplication in these elements. This indicates that a gene conversion event occurred after the divergence of gorillas and orangutans, but prior to the divergence of gorillas, humans, and chimpanzees. During this process, a small proportion of the ancestral AluY element has been converted to an AluYb-like sequence, including the AluYb diagnostic duplication and another adjacent tightly linked diagnostic mutation. The gene-converted region could be as small as 8 bp (251–259) or as long as 30 bp (237–267) in this case. Thus, the overall gene conversion rate (2/12) we observed here seems to be much higher than in previous studies (Maeda et al. 1988; Kass et al. 1995; Roy-Engel et al. 2002; Salem et al. 2003a). However, this is not surprising, since the difference may be due to the small sample size in the current study or reflect the longer evolutionary time period that each element may have been subjected to gene conversion.
The Pan4 locus contains the oldest AluYb8 element we identified. Interestingly, the Pan4 AluYb8 element in the human genome only has one CpG mutation (G→A) at position #5, compared with the AluYb8 consensus sequence, while the chimpanzee and gorilla AluYb8 elements at the orthologous Pan4 locus each have accumulated five species-specific mutations compared with the AluYb8 consensus sequence. Since no mutations are shared by the human, chimpanzee, and gorilla AluYb8 elements at the Pan4 locus, we believe that the ancestral AluYb8 element at the Pan4 locus was a canonical AluYb8 element. Using the BLAT program (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start), we identified three additional AluYb8 elements in the human genome with 100% sequence identity to the Pan4 element. This result suggests that the Pan4 element may still be retrotranspositionally active in the human genome, although we cannot completely rule out the possibility that these four elements independently mutated at position #5. Multiple alignments of the Pan1, Pan2, Pan3, and Pan4 loci are available on our Web site under publications (http://batzerlab.lsu.edu).

**AluYb insertions in other primate genomes**

To further investigate the propagation of the AluYb lineage, we next searched for the presence of AluYb elements in additional nonhuman primate genomes. Since no complete draft genomic sequences are available other than human and common chimpanzee, we used a modified PCR display method to identify AluYb elements from other primate genomes (see Methods). The display procedure was performed for four hominoid primates: pygmy chimpanzee, gorilla, orangutan, and gibbon; three Old World monkeys: green monkey, rhesus monkey, and pig-tailed macaque. In addition, this approach was also applied to the common chimpanzee to identify additional AluYb elements absent from the common chimpanzee genomic sequence. Two restriction enzymes were used for every template, and a minimum of 72 colonies were sequenced for each species. For the common chimpanzee, a total of seven AluYb elements were retrieved, all of which had previously been identified from the draft sequence. This suggested that our method involving two restriction enzymes yielded a ~60% coverage (7/12) of the AluYb elements in the genome. One pygmy chimpanzee-specific and five gorilla-specific AluYb8 insertions were identified and confirmed by a PCR analysis of nonhuman primates (Fig. 2A,D) and DNA sequencing. No AluYb element was recovered from the orangutan genome using the PCR display approach. The only locus identified within the gibbon genome was Pan2, which had been previously identified in the chimpanzee genome using the computational approach. Despite at least two trials for each species, no AluYb elements were identified in the three Old World monkeys (green monkey, rhesus monkey, and pig-tailed macaque) examined. All of the new AluYb loci identified by PCR display are listed in Table 2.

**Age estimates for the AluYb8 insertions in chimpanzee and gorilla**

To estimate the average age of AluYb8 elements in the chimpanzee genome, CpG and non-CpG mutation densities were calculated for 10 chimpanzee-specific AluYb8 elements as reported previously (Xing et al. 2004) using the chimpanzee AluYb8 consensus. The 10 elements contained a total of nine non-CpG mutations of 2420 nucleotides and 10 CpG mutations of 460 CpG nucleotides. The mutation densities were 0.37% ± 0.30% (average ± standard deviation) and 2.17% ± 2.29% for the non-CpG nucleotides and CpG nucleotides, respectively. Using a neutral mutation rate of 0.0015/site/Myr for non-CpG sites and a mutation rate of 0.0090/site/Myr for CpG sites (Xing et al. 2004), the average non-CpG and CpG mutation densities yield age estimates of 2.48 ± 2.03 and 2.42 ± 2.55 Myr, respectively. For the gorilla-specific AluYb8 elements, a total of five elements were analyzed. The mutation densities were 0.17% ± 0.23% and 3.48% ± 3.30% for the non-CpG and CpG sites, yielding age estimates of 1.10 ± 1.51 and 3.86 ± 3.66 Myr, respectively.

**Phylogenetic analysis of AluYb elements in the human genome**

The presence of AluYb8 elements in different primate genomes suggests that the origin of the AluYb8 subfamily may be much older than its major expansion in humans. In addition, the pres-
ence of \textit{Alu}Yb8, but not \textit{Alu}Yb7 or \textit{Alu}Yb9 elements in the chimpanzee genome suggests that the \textit{Alu}Yb8 subfamily may be the ancestral component of the human \textit{Alu}Yb lineage and would, therefore, predate the other two major human \textit{Alu}Yb subfamilies (i.e., \textit{Alu}Yb7 and \textit{Alu}Yb9). To test this hypothesis, we examined the phylogenetic relationships of the different components of the \textit{Alu}Yb lineage in the human genome, using a median-joining network approach (Cordaux et al. 2004) as implemented in the software NETWORK 4.1 (Bandelt et al. 1999) available at http://www.fluxus-engineering.com/sharenet.htm. From the \textit{Alu}Yb elements identified by Carter et al. (2004) in the human genome, we removed truncated elements and members of the \textit{Alu}Yb7, \textit{Alu}Yb8, and \textit{Alu}Yb9 subfamilies, leaving 36 previously unclassified \textit{Alu}Yb elements. After deleting the middle A-rich region and poly-A tail of the elements, a network of the \textit{Alu}Yb lineage was reconstructed using the 36 non-\textit{Alu}Yb7/8/9 elements and the consensus sequences of the \textit{Alu}Yb7, \textit{Alu}Yb8, and \textit{Alu}Yb9 subfamilies. Collectively, these three subfamilies comprise \sim 2000 copies. A preliminary analysis suggested that several nucleotide positions may have mutated more than once. Thus, the \textit{Alu}Yb network was calculated with these putative hypervariable positions down-weighted to 1 (positions #64, #98, and #144) or 5 (position #174 and position #211), while other positions were given a weight of 10. The resulting network (Fig. 5) shows that the \textit{Alu}Yb8 subfamily occupies a central position in the network. In addition, the \textit{Alu}Yb8 node is associated with the highest number of direct branches (nine), as compared with the \textit{Alu}Yb7 and \textit{Alu}Yb9 nodes (four branches each). Finally, 72% of the non-\textit{Alu}Yb7/8/9 elements are more closely related to \textit{Alu}Yb8 than to \textit{Alu}Yb7 or \textit{Alu}Yb9. Taken together, these results are strongly suggestive that the \textit{Alu}Yb8 elements are ancestral to the other \textit{Alu}Yb subfamilies in the human genome (Posada and Crandall 2001; Cordaux et al. 2004).

**Discussion**

The origin of the \textit{Alu}Yb lineage

The \textit{Alu}Yb lineage is one of the most active \textit{Alu} lineages in the human genome, with an estimated copy number of \sim 2000 (Carter et al. 2004). To obtain further insight into the origin of the \textit{Alu}Yb lineage in the primate order, we analyzed the draft sequence of the common chimpanzee genome, and identified 12 \textit{Alu}Yb insertions, 10 of which are members of the \textit{Alu}Yb8 subfamily, while the other two are non-Yb7/8/9 elements. The presence of an \textit{Alu}Yb element at the Pan2 locus within siamang and gibbon genomes suggested the \textit{Alu}Yb lineage originated before the divergence of all hominoid primates. However, no \textit{Alu}Yb elements have been identified in multiple Old World monkey genomes; thus, the origin of the \textit{Alu}Yb lineage was after the divergence of Old World monkeys and apes. These results place the origin of the \textit{Alu}Yb lineage at the early stage of the
hominoid evolution, about 18–25 Mya (Goodman et al. 1998). It is worth noting that although the Alu\textsubscript{Yb}6 insertion at the Pan2 locus is the oldest Alu\textsubscript{Yb} element we identified, it was most likely generated via a gene conversion event. Therefore, it may not be the founder gene of the Alu\textsubscript{Yb} lineage, but rather an early offspring of the Alu\textsubscript{Yb} founder gene, which was subsequently lost in extant primates.

The Pan4 Alu\textsubscript{Yb} element appears to have been fixed in the human, chimpanzee, and gorilla genomes, and contains all eight diagnostic mutations that characterize the Alu\textsubscript{Yb}8 consensus sequence. This suggests that within the 10 Myr or so after the Alu\textsubscript{Yb}8 lineage initially arose, it was not very active in terms of retrotransposition, if at all. However, this lineage retained its retrotranspositional potential during this extended period of time. Furthermore, our results suggest that in gorilla and chimpanzee genomes, the copy number of Alu\textsubscript{Yb} elements is two orders of magnitude lower than that in human. Therefore, over several million years following the insertion of the Alu\textsubscript{Yb}8 at the Pan4 locus, the Alu\textsubscript{Yb} lineage still retained a very low retrotransposition activity (∼1.5 fixed copies/Myr in the chimpanzee genome) until the major expansion of the Alu\textsubscript{Yb} lineage in the human genome within the past 3–4 Myr (Carter et al. 2004; Xing et al. 2004).

The “stealth driver” model of Alu evolution

The long evolutionary history of Alu\textsubscript{Yb} lineage leads to the conclusion that the Alu\textsubscript{Yb} lineage has remained in the genome with little or no retrotransposition activity for an extended period of time, while retaining the ability to generate an appreciable number of new copies later in a species-specific manner (Fig. 6). This scenario is different from the classic “master” gene model, in which a master gene is defined as an element that is highly active over long periods of time (Deininger et al. 1992). In general, the amplification dynamics of the Alu\textsubscript{Yb}8 lineage show a striking similarity to that of the Alu\textsubscript{Ya}5 lineage (Leelang et al. 1993; Shaikh and Deininger 1996). Although the existence of low-activity Alu source genes has previously been suggested for the Alu\textsubscript{Ya}5 subfamily (Shaikh and Deininger 1996), here we provide evidence that low retrotransposition activity Alu source genes should be recognized as a major factor driving Alu expansion and evolution.

We propose a model of Alu evolution and retrotransposition, in which the low-activity Alu elements are termed “stealth drivers.” In contrast to master genes, stealth drivers are not responsible for generating the majority of new Alu copies, but rather for maintaining genomic retrotransposition capacity over extended periods of time. By generating new Alu copies at a slow rate, a stealth driver may generate some daughter elements that are capable of much higher retrotransposition rates. These hyperactive daughter elements may act as master genes for the amplification of Alu subfamilies and are responsible for producing the majority of the subfamily members. In addition, the new master genes may also give birth to additional retrotranspositionally active Alu copies that serve as secondary master genes or sprouts and also contribute to the expansion of the subfamily (Deininger and Batzer 1993). This view is supported by the fact that recently integrated human Alu subfamilies typically contain ∼15% of such secondary master genes, in addition to the primary
subfamily master gene (Cordaux et al. 2004). While the highly active master genes that are produced from the stealth driver would be deleterious and generally be subject to negative selection, the low-activity stealth driver itself will not be subject to such selection and would allow the lineage to persist for extended periods of time. Furthermore, both stealth driver and its daughter elements will also generate new elements that exhibit low levels of retrotransposition activity, effectively becoming new stealth-driver elements. In fact, it is possible that the continuation of a lineage over extended periods of time may depend on the production of new driver elements.

But the question remains as to why, after persisting relatively quietly for millions of years, a lineage can show a sudden increase in numbers, as appears to be the case for AluYa and AluYb lineages. Under the stealth-driver model, the master or stealth locus does not need to be “turned on,” as one or more such drivers has been active during the entire history of the lineage. There are multiple scenarios that may account for the sudden expansion of the Alu lineage. In the scenario that we favor, periods of rapid expansion may be related to the ability of highly active daughter elements to escape selection at the population level and, consequently, produce more progeny. The change in the efficiency of natural selection in weeding out overactive elements may be related to population bottlenecks or other demographic factors (Hedges et al. 2004). Alternatively, periodic increases in element numbers may simply be due to the stochastic nature by which active daughter elements are produced. Yet another possibility is that, contingent on the relative abundance of available L1 retrotransposition machinery, there may be fluctuations in Alu expansion rates (Dewannieux et al. 2003). Less likely is the possibility that molecular host defense mechanisms, which were previously suppressing the activity of Alu elements, failed for some reason during these periods.

In the AluYb lineage, multiple lines of evidence suggest that the AluYb8 element at the Pan4 locus may be a recent stealth-driver gene in human-specific Alu retrotransposition. First, the Pan4 element is the oldest AluYb8 element we identified; second, it only accumulated one point mutation over the last 7 Myr; third, the presence of three identical human AluYb8 elements is consistent with low levels of retrotransposition of this element. On the other hand, since the Pan4 AluYb8 element has accumulated five mutations in the chimpanzee and gorilla genome, and there are no other AluYb elements in these genomes with the same mutations, the AluYb8 element at Pan4 locus is unlikely to be the current driver gene in the chimpanzee and gorilla genomes. The reasons why the Pan4 AluYb8 element may be a stealth driver in human, but not chimpanzee and gorilla, are unclear. However, there is a striking correlation between sequence similarity to the AluYb8 consensus sequence and total number of AluYb copies in these species, raising the possibility of mutational inactivation of the Pan4 AluYb8 element in chimpanzee and gorilla, but not in the human lineage.

Previous studies have shown that the amplification of the Alu family reached its peak about 30 Mya and subsequently underwent retrotranspositional quiescence (Shen et al. 1991; Britten 1994). Although we still do not know the underlying mechanisms for the retrotranspositional quiescence, the stealth-driver model may explain why the Alu lineage has been subjected to periods of retrotranspositional quiescence interspersed with episodic bursts of amplification, as suggested by the accumulation of at least 5000 human-specific Alu elements since the human-chimpanzee divergence (Carroll et al. 2001; Batzer and Deininger 2002; Xing et al. 2003; Carter et al. 2004; Otieno et al. 2004). The AluYa and AluYb lineages that comprise >60% of the human-specific Alu elements may be just two successful examples of this strategy. Similar patterns of amplification have also been observed in the retrotransposition of rodent SINE family, ID, (Kim et al. 1994) and the rodent LINE (Long INterspersed Element) family, Lx (Pascale et al. 1993). In addition to these evolutionary observations, there is also experimental evidence that indicates
that some varieties of mobile elements evolve strategies to attenuate their own activity. In Alu, the acquisition of a second monomer to form its dimeric structure has been linked to decreased retrotranspositional activity (Li and Schmid 2004). In vitro modifications to currently active L1 elements can produce orders of magnitude increases in the L1 amplification rate (Han and Boeke 2004). In addition, cryptic polyadenylation sites throughout the L1 sequence may serve to quell the number of full-length, retrotransposition-competent L1 copies generated (Peretelitsa-Belancio and Deininger 2003). Taken together, this evidence suggests that the stealth-driver model may not be a unique feature of the Alu family itself, but rather may be one variant of a common survival strategy for SINE and LINE elements. More generally, the ability of mobile elements to maintain low to moderate levels of amplification activity, rather than more rapid duplication rates, may be a common feature of long-lived, successful families of transposons.

Methods

Computational identification of Alu/Yb elements

A 31-bp (TGGCCCACTGAGTCCGAGTCGCAGTGCTGGG) oligonucleotide that included the Alu/Yb lineage-specific duplication was used to screen the common chimpanzee genome draft sequences (panTro1 Nov. 2003 assembly, http://www.ncbi.nlm.nih.gov; R. Waterston, pers. comm.) using the Basic Local Alignment Search Tool (BLAST) program available at http://www.ensembl.org/multi/blastview (Altschul et al. 1990). All Alu elements that have the diagnostic 7-bp duplication were selected and extracted along with 1000 bp of unique DNA sequence adjacent to both ends of the elements. The program RepeatMasker (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) was then used to annotate all known repeat elements within the DNA sequence. Flanking oligonucleotide primers for the PCR amplification of each Alu element were then designed using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). The primers were subsequently screened against the GenBank NR database using BLAST queries to determine whether they resided in unique DNA sequences and would only amplify the Alu elements of interest.

Identification of Alu/Yb elements with PCR display

The Alu element PCR display methodology has been reported previously (Ray et al. 2005). Using this approach, 500 ng of genomic DNA was partially digested using restriction endonucleases NdeI or MseI as recommended by the manufacturer (New England Biolabs) in 120 µL reactions. Digestion products were then ligated with double-stranded linkers and amplified and “displayed” using the primer LNP (5′-GAATTCGTAAACATAGCATTCT-3′) and an Alu/Yb-specific primer (5′-GGCGCCAATGTAGCAGTGG-3′) to acquire partial Alu sequences and the accompanying flanking unique sequences from each template. Since the Alu sequence in the amplicon is about 300 bp long, the PCR products were then purified by BD CHROMAS SPIN-400 columns (BD Biosciences) to select the fragments larger than 400 bp, so that enough unique flanking sequence can be obtained to locate the orthologous sequences in the draft sequence of the human genome. A second-round amplification was performed using the LNP oligonucleotide and a second nested Alu/Yb-specific primer (5′-AATCTCGAGTCGTAACATGCAGTCGGCCTGGG-3′) to increase the specificity of the amplicons. The second-round PCR products were separated on a 2% agarose gel, and fragments larger than 400 bp were excised and extracted from the gel using the Wizard gel-purification kit (Promega). The purified products were then cloned into the TOPO-TA cloning vector (Invitrogen). At least 72 clones were randomly isolated from each template, and DNA sequences were determined from both strands using chain-termination sequencing (Sanger et al. 1977) on an ABI 3100 automated DNA sequencer.

After obtaining the sequences, the BLAST-Like Alignment Tool (BLAT) program (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) (Kent 2002) was used to compare the resulting partial Alu/Yb element and its adjacent flanking sequence with human draft sequences to identify orthologous sequences. Using a combination of the unique flanking sequence and the orthologous human DNA sequences, oligonucleotide primers were designed around each newly identified Alu/Yb element as outlined above.

PCR analysis of Alu/Yb elements

All of the Alu/Yb loci were screened on a panel composed of human genomic DNA (HeLa cell line ATCC CCL-2) and DNA samples from the following nonhuman primate species: Pan troglodytes (common chimpanzee), Pan paniscus (bonobo or pygmy chimpanzee), Gorilla gorilla (lowland gorilla), Pongo pygmaeus (orangutan), Hylobates syndactylus (siamang), Hylobates lar (white handed gibbon), Chlorocebus aethiops sabaens (green monkey), and Aotus trivirgatus (owl monkey). The nonhuman primate DNA is available as a primate phylogenetic panel (PRP00001) from the Coriell Institute for Medical Research. The chimpanse-specific Alu/Yb loci were also screened for insertion presence/absence using a common chimpanzee population panel composed of 12 unrelated individuals of unknown geographic origin, which was provided by the Southwest Foundation for Biomedical Research.

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 dNTP’s in 50 mM KCl, 1.5 mM MgCl2, 10 mM Tris-HCl (pH 8.4), and 2.5 units Taq DNA polymerase. Each sample was subjected to an initial denaturation step of 94°C for 150 sec, followed by 32 cycles of 1 min of denaturation at 94°C, 1 min of annealing at optimal 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 fractionated on a 2% agarose gel with 0.25 µg of ethidium bromide and visualized using UV fluorescence.

DNA sequence analysis

To confirm the presence of Alu/Yb elements, all PCR products suggesting the presence of an Alu element were gel purified using the Wizard gel-purification kit (Promega). Purified PCR products were then cloned into vectors using the TOPO TA cloning kit (Invitrogen) and sequenced using chain-termination sequencing (Sanger et al. 1977) on an Applied Biosystems 3100 automated DNA sequencer. All clones were sequenced in both directions to confirm the sequence. The DNA sequences generated in this study are available in the GenBank under accession numbersAY791249–AY791290.

Acknowledgments

We thank Dr. David Ray for critical reading and suggestions during the preparation of this manuscript. We also thank Anthony Carter for providing the human Alu/Yb sequences. This research was supported by the Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05-01) (M.A.B.), Na-


Web site references

http://batzerlab.lsu.edu; Batzer laboratory.
http://www.ensembl.org/multi/blastview; Basic Local Alignment Search Tool (BLAST).
http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker; RepeatMasker.
http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_web.cgi; Primer3.
http://genome.ucsc.edu/cgi-bin/hgBlat?command=start; BLAT.

Received November 19, 2004; accepted in revised form January 28, 2005.
Under the genomic radar: The Stealth model of Alu amplification

Kyudong Han, Jinchuan Xing, Hui Wang, et al.

Genome Res. 2005 15: 655-664
Access the most recent version at doi:10.1101/gr.3492605

References
This article cites 49 articles, 14 of which can be accessed free at:
http://genome.cshlp.org/content/15/5/655.full.html#ref-list-1

License

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.