doi:10.1006/jmbi.2001.5380 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 316, 1033–1040

JMB

Non-traditional *Alu* evolution and Primate Genomic Diversity

Astrid M. Roy-Engel[1]†, Marion L. Carroll[2,3]†, Mohamed El-Sawy[1] Abdel-Halim Salem[2,4], Randall K. Garber[2,4], Son V. Nguyen[2] Prescott L. Deininger[1,5] and Mark A. Batzer[2,4]*

[1]*Tulane Cancer Center SL-66, Department of Environmental Health Sciences, Tulane University - Medical Center 1430 Tulane Ave., New Orleans, LA 70112, USA*

[2]*Departments of Pathology, Biochemistry and Molecular Biology, Genetics, Stanley S. Scott Cancer Center, Neuroscience Center of Excellence, 1901 Perdido St. New Orleans, LA 70112, USA*

[3]*Department of Chemistry Xavier University of Louisiana 1 Drexel Dr., New Orleans LA 70125, USA*

[4]*Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge LA 70803, USA*

[5]*Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, 1516 Jefferson Highway, New Orleans, LA 70121, USA*

**Corresponding author*

Alu elements belonging to the previously identified ''young'' subfamilies are thought to have inserted in the human genome after the divergence of humans from non-human primates and therefore should not be present in non-human primate genomes. Polymerase chain reaction (PCR) based screening of over 500 *Alu* insertion loci resulted in the recovery of a few ''young'' *Alu* elements that also resided at orthologous positions in non-human primate genomes. Sequence analysis demonstrated these ''young'' *Alu* insertions represented gene conversion events of pre-existing ancient *Alu* elements or independent parallel insertions of older *Alu* elements in the same genomic region. The level of gene conversion between *Alu* elements suggests that it may have a significant influence on the single nucleotide diversity within the genome. All the instances of multiple independent *Alu* insertions within the same small genomic regions were recovered from the owl monkey genome, indicating a higher *Alu* amplification rate in owl monkeys relative to many other primates. This study suggests that the majority of *Alu* insertions in primate genomes are the products of unique evolutionary events.

Keywords: *Alu* insertion; SINE; parallel insertions; *Alu* evolution; gene conversion

Introduction

The amplification of *Alu* elements has greatly impacted the evolution of the human genome.[1] With over one million copies, *Alu* elements are the most abundant SINE (short interspersed element) in the human genome, contributing approximately 11% of its mass.[1] However, almost all of the individual *Alu* elements are incapable of amplification.[2] The generally accepted model of *Alu* amplification proposes that a few mobilization

†These authors contributed equally to this research.
P.L.D. and M.A.B. are equal senior authors.

Correspondence address: M. A. Batzer, Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA.

Abbreviations used: SINE, short interspersed element.

E-mail address of the corresponding author: mbatzer@lsu.edu

0022-2836/02/051033–8 $35.00/0

competent elements termed source or ''master'' genes generated the vast majority of the *Alu* elements.[3–5] Although SINEs are present in the genomes of practically all multicellular organisms, the *Alu* family of SINEs have amplified in the last 65 million years and are only found in primate genomes.[5–7] As time progressed, the sequence(s) of the active *Alu* element(s) mutated, thus generating copies carrying these mutations. These different *Alu* elements can be grouped into subfamilies that share common diagnostic mutations and are known to have amplified at different evolutionary time frames.[7,8] Most *Alu* elements (~90%) inserted in genomes early in primate evolution[7] (see Figure 1). Because there is no known mechanism that specifically removes *Alu* elements from primate genomes, members of older *Alu* subfamilies can be found in the human genome and also at the orthologous positions in other non-human primate genomes.

Consistent with this amplification model, the very young *Alu* subfamilies, such as Ya5, Ya5a2, Ya8, Yb8, and Yb9, are largely restricted to the human genome or orthologous loci of other African apes.[9–11] To date, only two *Alu* Ya5 elements have been identified that were also present in orthologous positions within non-human primate genomes.[12] One being our previously reported Ya5NBC42-containing locus,[13] that is characterized in detail here. The current amplification rate of human *Alu* elements of about one new insertion in every 100-200 births has decreased approximately two orders of magnitude since its peak activity.[2,14] In spite of the decreased amplification rate these *Alu* subfamilies have continued to generate insertion polymorphisms and *de novo* insertion mutations confirming their recent amplification within the human genome.[15]

We have previously identified a large number of ''young'' *Alu* containing loci in the human genome, both to understand their evolution and to develop them for use as markers in studies of human genetic diversity.[9,11,13] In this study, we performed a detailed primate phylogenetic analysis of over 500 recently integrated *Alu* elements originally ascertained within the human genome. Some of these elements have undergone gene conversions that modified the sequence architecture of the elements. In addition, we have identified a few insertions that have occurred in parallel in the orthologous

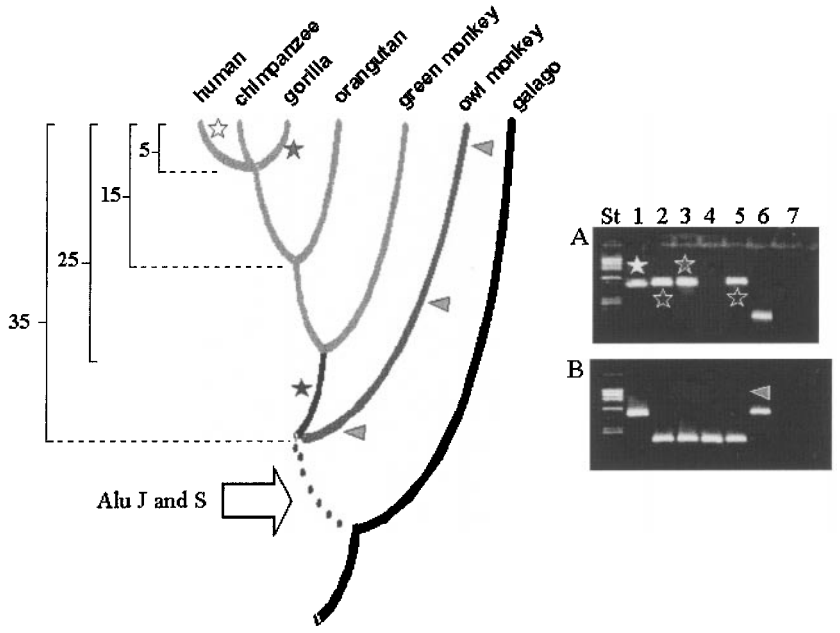

Figure 1. Gene conversion and parallel independent insertions of *Alu* elements throughout primate evolution. A schematic of the primate evolutionary tree is shown on the left. Estimated evolutionary time periods between the different speciation events are indicated on the left in millions of years. A blue star represents the time period of insertion of the *Alu* Ya5NBC42; the white and red stars represent time periods where the gene conversion of the pre-existing *Alu* element to a different subfamily occurred. Yellow arrowheads represent arbitrary examples of the time period when the parallel independent insertion events recovered from the owl monkey genome may have occurred. The green shaded branches represent the total time distances used for the calculation of the parallel independent rates (million insert site years) in Results. The large block arrow indicates the evolutionary time period where the vast majority of *Alu* elements (*Alu* S and J subfamilies) present in humans inserted into the genome. PCR amplification of selected loci in different primates to determine the presence/absence of *Alu* elements resulted in two types of events: gene conversion (top panel) and independent insertions within the same locus (bottom panel). Examples of a typical PCR amplification for each variant are shown. The lanes are St, standards; 1, human; 2, chimpanzee; 3, gorilla; 4, orangutan; 5, green monkey; 6, owl monkey; 7, negative control.

positions of non-human primate genome(s). This large dataset allows us to begin to estimate the impact of these evolutionary processes on the architecture of primate genome(s).

Results

Presence of *Alu* elements in primate orthologous loci

In previous studies, we utilized basic local alignment search tool (BLAST) searches of the human genome to retrieve over 4500 different loci containing *Alu* elements belonging to young subfamilies.[9–11,13] A subset of these loci were analyzed by PCR amplification to determine the presence or absence of the *Alu* element in different primate species. The loci amplified in all or almost all of the primate genomes, except for two cases (represented as 0 in Table 1). However, data from these two samples are not necessary for the evolutionary analysis of these particular loci and does not affect the current interpretation of the data. The absence of amplification is presumably the result of technical problems, where the primer (designed to perfectly match the human sequence) may not anneal properly to the non-human primate sequence as a result of normal evolutionary divergence between the two sequences. The analysis of different *Alu* containing loci included 231 Ya5, 14 Ya5a2, 14 Ya8, 244 Yb8, and 40 Yb9 elements, all from the recently integrated ''young'' *Alu* subfamilies. As predicted by the traditional model of *Alu* amplification, almost all of these *Alu* elements only resided in the human genome. However, some loci had PCR amplification patterns that were unanticipated (see Table 1). Some of the loci from non-human primates also contained *Alu* inserts (as observed by a larger PCR product) in many of the assayed primates, including orangutan, green monkey and owl monkey; suggesting potentially that these elements had retroposed much earlier in primate evolution than previously reported.[7] Other loci contained an *Alu* repeat in the human and owl monkey genomes only, suggesting either the selective loss of the *Alu* repeat in some of the other non-human primates or that parallel independent insertions occurred after the radiation of the owl monkeys from other old world primates and African apes. Sequence analysis of all of these unusual events indicated that two types of events had occurred: gene conversions of pre-existing *Alu* elements by an element belonging to a different *Alu* subfamily or parallel independent insertions of different *Alu* elements in very close, but non-identical genomic locations (see Figures 1 and 2).

Alu gene conversion events

Two *Alu*-containing loci were involved in gene conversion events, Yb8NBC253 and Ya5NBC42. In this case, the Alu elements present in all the orthologous loci have the same flanking sequences and direct repeats (not identical due to mutations accumulated through time). DNA sequence analysis of these loci revealed that the element present in the Yb8NBC253-containing locus of all the non-human primates tested belonged to the *Alu* ''S''-like subfamily. This suggests that the gene conversion of the older *Alu* element to an *Alu* Yb8 subfamily member in the human genome took place after the radiation of humans from other African apes, which is thought to have occurred four to six million years ago[16] (Figure 1). In the case of the Ya5NBC42-containing locus the green monkey, pygmy chimpanzee and common chimpanzee contain an *Alu* element belonging to the Y subfamily (Figure 3), suggesting that the *Alu* Y element initially inserted in this locus. Two separate and independent gene conversions occurred, where the first changed the *Alu* sequence in the human genome so that it contains three of the five *Alu* Ya5 diagnostic nucleotides. The second gene conversion event occurred in the gorilla genome so that the sequence contains one clear *Alu* Sx diagnostic site (the dinucleotide that is deleted between S and Y subfamilies) (Figure 3). Through close inspection of the diagnostic sites and mutated nucleotides, we determined the putative region involved in the gene conversions (Figure 3). In both cases only a portion of the *Alu* element was gene converted. This observation is in good agreement with our previous hypothesis that the presence of ''mosaic'' elements in the genome probably occurred as a

Table 1. PCR analysis of orthologous loci for the presence or absence of *Alu* inserts

Alu element	Human	Chimpanzee	Pygmy chimpanzee	Gorilla	Orangutan	Green monkey	Owl monkey	Type
Ya5NBC42	+(Ya3)[a]	+(Y)	+(Y)	+(Sx)	0	+(Y)	-	GC × 2
Ya5NBC91	+(Ya5)	-	-	-	-	-	+(''Sg'')	Ind
Ya5NBC188	+(Ya5)	-	-	-	-	-	+(''Y'')	Ind[b]
Yb8NBC185	+(Yb8)	-	-	-	-	-	+($Y^?$)	Ind
Yb8NBC253	+(Yb8)	+(''S'')	+(''S'')	+(''S'')	+(''S'')	+(''S'')	0	GC
Total analyzed	543	399	217	268	177	202	139	

$Y^?$: very early intermediate in the evolution of Y; +, PCR product indicates presence of an *Alu* insert; -, small PCR product indicates absence of an *Alu* insert; 0, no PCR product of the locus was observed; GC, gene conversion; Ind, independent insertion.

[a] *Alu* subfamily indicated in parenthesis.

[b] One end identical.

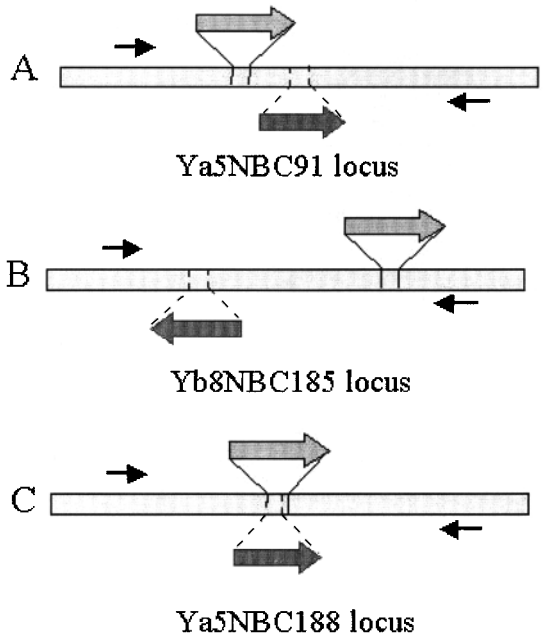

Figure 2. Schematic of the three parallel independent *Alu* insertions observed within the orthologous loci of human and owl monkey. The region of each locus amplified is approximately 200 bp (without the *Alu* insert) located between the two selected primers (continuous black arrows). The *Alu* elements present in the different genomes are represented as yellow (human) or red block arrows (owl monkey). The line extensions (continuous for human and dotted for owl monkey) correspond to the genomic cleavage sites that subsequently get duplicated to form the direct repeats once the *Alu* element inserts. The three owl monkey loci characterized contain *Alu* elements that inserted in different sites in both the same (a) or inverted orientation (b) as the human *Alu*. The third loci the owl monkey *Alu* element shared one of the cleavage sites as the human *Alu* but not the other (c). Note: the Figure is not drawn to scale.

result of frequent gene conversion events of small regions within *Alu* elements.[10]

Independent *Alu* insertions and *Alu* amplification rates

We also detected three parallel independent insertions that occurred in a new world monkey (owl monkey) within the same loci where an *Alu* Ya5 or Yb8 repeat was located in the human genome (Table 1). The elements present in owl monkey belong to different *Alu* subfamilies similar to the Sg and Y subfamilies. All of the insertion sites were very similar, but not identical to the human insertion site and were localized within the same 200 bp genomic region (Figure 2). Although the sites are not identical, we will continue to refer to them as ''parallel'' insertions. One was shifted over a few bases, while a second was shifted over and inverted relative to the human *Alu* element (see Table 1). The third shared one end of the insertion site, but had a different length direct repeat, consistent with a parallel insertion at the same site which made the second genomic nick at a slightly shifted position (Figure 2). All of these independent insertions must have occurred sometime between 35 million years (after the radiation of new and old world primates) and the present (see Figure 1). However, no other independent insertions were observed in any of the other primates tested, suggesting that the amplification rate of *Alu* repeats in the owl monkey genome was significantly higher than the amplification rate in the other primate genomes. To estimate the parallel insertion rate of *Alu* elements in owl monkeys we used the number of loci analyzed in owl monkey (139) and multiplied it by the time elapsed after the radiation of owl monkeys (35 million years; see Figure 1), giving us a rate of three events per 4865 million insert site years. To compare it to the other primates with no independent insertions detected, we added all of their individual rates. We used the age indicated in Figure 1 and the successful PCR amplifications in Table 1 for each of the different primates: $(399 \times 5) + (217 \times 5) + (268 \times 5) + (177 \times 15) + (202 \times 25) = 12{,}125$ million insert site years. Therefore, the owl monkey rate after radiation is much more than 2.5 (12,125/4865) times faster than the sum of all the rates (representing the green shaded branches in Figure 1) of the other non-human primates.

The frequency of occurrence of independent insertions in owl monkey can be estimated as three events out of 139 successful PCR amplifications (Table 1) or 2.2%. The size of the target site tested in our PCR assay is approximately 200 bp, making the amount of DNA screened 27,800 bp. Assuming the target site for integration is random we expect to detect one new insertion in every ~9000 bases. However, *Alu* elements do not insert completely randomly, but rather appear to have a site preference for locally A + T-rich regions,[17] adding a degree of uncertainly to the estimate.

Discussion

Our data have several implications for *Alu* insertions and post-integration sequence evolution. First, it supports the ''master'' or limited amplification model for *Alu* amplification. This model proposes that most *Alu* copies present in the human genome arose from a few active copies and that different subfamilies were active at different evolutionary periods. Therefore, *Alu* subfamilies that are active after a radiation of two species should yield newly inserted copies at specific loci that would not be shared between the two primate species. In our analysis, no *Alu* elements belonging to the younger *Alu* subfamilies were recovered in any of the non-human primate genomes, as all of our ''PCR positives'' from non-human primates were either gene conversion events or the products

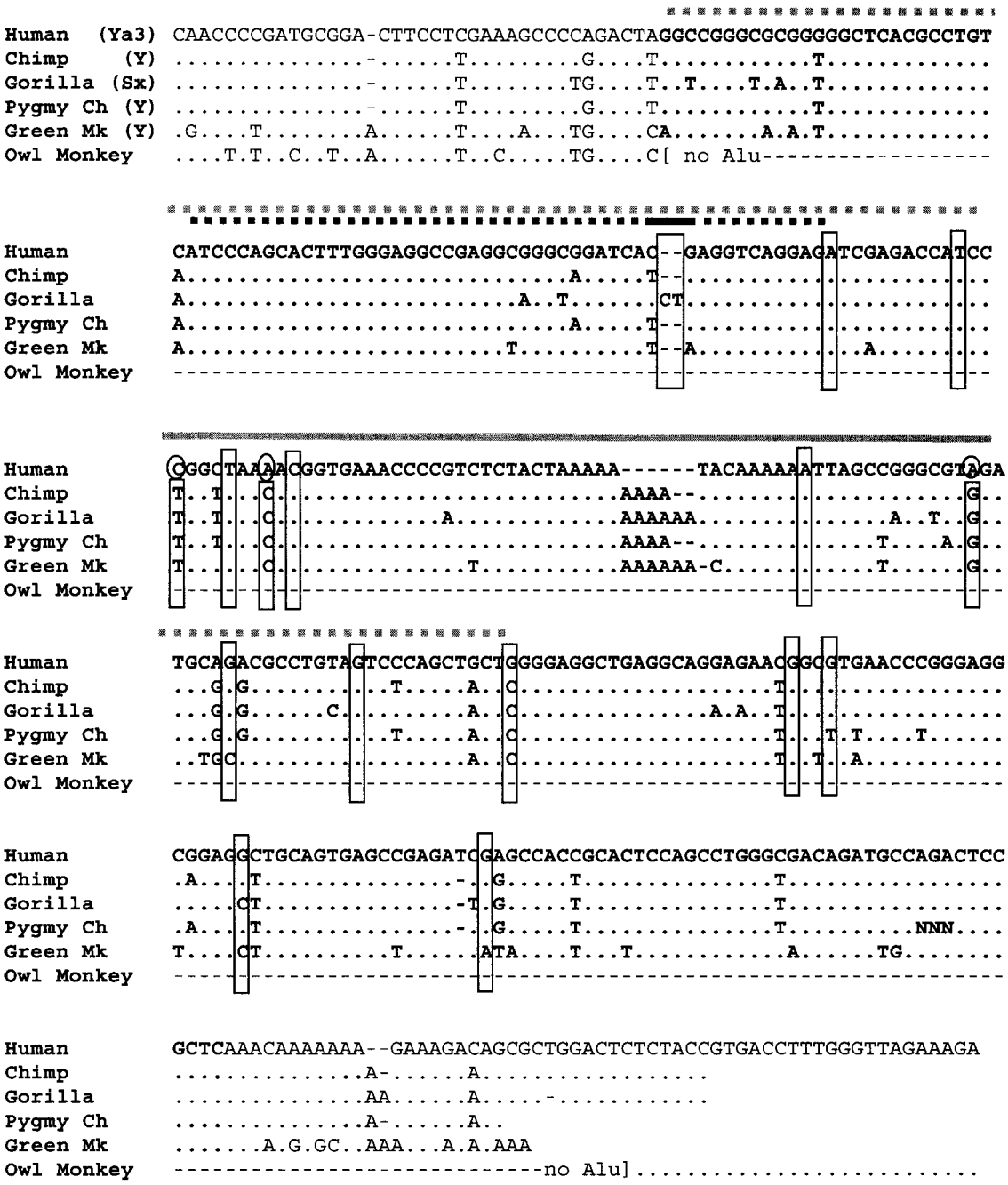

Figure 3. Sequence alignment of the Ya5NBC42 *Alu* containing loci in different primates. The human sequence of the Ya5NBC42 *Alu* containing locus (accession number AL078621) is shown on the top line. Nucleotide substitutions at each position are indicated with the appropriate nucleotide. Deletions are marked by dashes (-). The subfamily of the *Alu* element present in each species is indicated in parenthesis. The sequence involved in the gene conversion to a Ya3 in human (gray line) or Sx in gorilla (black line) is indicated above the sequence. The extended dotted lines represent the uncertainty of the boundaries of the region involved in the gene conversion. Rectangles enclose the diagnostic sites used to elucidate the gene-converted region. Circles enclose the three Ya5 diagnostic sites present in the human sequence.

of parallel independent *Alu* insertions. To date, only one authentic *Alu* Ya5 element that has been characterized is also found in orthologous positions within non-human primate genomes.[12]

Secondly, these data indicate that newly integrated *Alu* elements are stable integrations within primate genomes and that they are identical by descent. In our screening, the few cases of discordant amplification patterns (presence of an *Alu* element in a locus of one primate, but not in the other evolutionarily younger primates) resulted from parallel independent insertions. Out of

approximately 500 loci analyzed only three contained parallel independent *Alu* insertions. The parallel rate of *Alu* insertion events is extremely low when considering the number of loci analyzed and the full length of the evolutionary tree, since the different primates shared a common ancestor resulting in only three events recovered in a total of 16,990 million insert site years. Therefore, we conclude that these represent relatively unique events within primate genomes. Within the old world primates, we have assayed hundreds of sites with a combined total of over 12,000 million years of site evolution without detecting a single parallel insertion. This represents having sampled across 500 genomic sites analyzed with an average of 25 million years of evolution per site. Based on this number, if we assume humans diverged from one another as far back as one million years, we would expect to see less than one parallel insertion event per locus in a diverse population of over 12,000 individuals. This demonstrates the very low probability of detecting parallel independent *Alu* insertions in the human population. Thus, parallel insertion events represent a trivial percentage of the total retroposition events in the human genome when using *Alu* insertion polymorphisms to study human population genetics. Therefore, *Alu* insertion polymorphisms are largely homoplasy free characters for the study of human evolution.

Only a few examples of parallel independent SINE insertions between species have been reported. One example reports the presence of the SINE B1 present in the same insertion site of the orthologous locus from two different species of rodents.[18] Two others report the insertion of two different retroposons in the same locus but at different insertion sites.[19,20] In addition, the Mys family of retrotransposons contains several parallel SINE insertions including two that occurred at identical positions in a genome.[21] Thus, the parallel insertion of SINE elements is a very rare event within a population and a relatively rare event even when different species and longer evolutionary time periods are considered.

Gene conversion between *Alu* repeats has been observed,[22–24] where the authors suggest that it may have contributed to between 10 % and 20 % of the *Alu* Ya5 SNP (single nucleotide polymorphism) diversity involving extensive levels of short gene conversions.[10] Also, the previous report suggested that most of the gene conversions involving Ya5 elements gene converted ''backwards'' to an older *Alu* subfamily.[10] Here, we have identified and characterized three gene conversion events (two within the same locus) after screening 543 independent *Alu*-containing loci within the human genome. Because, our strategy consists of screening loci that already contain young *Alu* elements in humans, we would expect that any gene conversions detected would be in the ''forwards'' orientation, i.e. an older *Alu* element present in the other non-human primate genomes, which converted forward to a younger *Alu* subfamily member in the human genome. However, in the Ya5NBC42-containing loci the ancestral *Alu* Y element gene converted ''backwards'' to an *Alu* ''Sx''-like element in gorilla.

Previously, it has been suggested that the germline recombination machinery in mammals has evolved to prevent high levels of ectopic recombination between repetitive sequences.[25] The authors were unable to detect gene conversion events at significant levels when using 624 bp in a model breeding transgenic mice. However, in this case the copy number of the transgene in the mouse lines used to evaluate recombination varied from 1 to 12 copies. This may not appropriately reflect the million plus *Alu* copies that are dispersed throughout the human genome. Although, the *Alu* elements are non-identical with an average of 15 % sequence divergence,[26] they share sequence identity in small regions within their sequence. The examples reported here and previously[10] suggest that only a portion of each *Alu* element was typically involved in the gene conversion. It is quite possible that the high copy number of *Alu* elements allows for pairing between the homologous regions of different *Alu* elements initiating the start of gene conversion before cellular control systems terminate the process resulting in small gene conversion tracts.

Our identification of gene conversion events may represent the tip of the genomic iceberg. As previously mentioned, we need to consider that we are introducing a bias when using the younger *Alu* Ya5/Yb8 subfamilies that represent only about 0.5 % of all the *Alu* elements for the initial screening. In addition, any gene conversions changing the *Alu* to another type of subfamily would not be included in the screening. We also need to take into account that the majority of gene conversions between *Alu* elements probably encompass small genomic regions (50 bp or less), and if the *Alu* subfamily specific diagnostic nucleotides were unaffected the event would not be recognized as such. In addition, gene conversion may not be equal throughout the genome and hot spots may exist. All these factors suggest that the impact of gene conversion on *Alu* evolution may be grossly underestimated. Furthermore, these observations support the growing literature suggesting that gene conversion has played a critical role in sculpting the genetic diversity within many different eukaryotic genomes.[27,28]

Materials and Methods

DNA samples and PCR amplification

PCR primers and reactions for each locus analyzed were performed as described.[13] Human DNA was available from previous studies.[9] The cell lines used to isolate DNA samples were as follows: chimpanzee (*Pan troglodytes*), WES (ATCC CRL1609); gorilla (*Gorilla gorilla*) Lowland Gorilla Coriell AG05251B, Ggo-1 (primary gorilla fibroblasts) provided by Dr Stephen J. O'Brien, National Cancer Institute, Frederick, MD, USA; pygmy

chimpanzee (*Pan paniscus*) Coriell AG05253A; orangutan (*Pongo pygmaeus*) ATCC CRL6301; green monkey (*Cercopithecus aethiops*) ATCC CCL70; and owl monkey (*Aotus trivirgatus*) OMK (OMKidney) ATCC CRL 1556. Cell lines were maintained as directed by the source and DNA isolations were performed using Wizard genomic DNA purification (Promega).

Sequence analysis

DNA sequencing was performed on gel-purified PCR products utilizing the original amplification primers to obtain sequence from both strands. Initial sequence data were obtained using the Thermo Sequenase Cycle Sequencing kit (USB Corporation). Confirmation of the sequence was performed at the Sequencing Analysis Core at the Tulane Health Sciences Center with an ABI Prism 3100 sequencer. Sequence alignments were performed using MegAlign software (DNAStar version 3.1.7 for Windows 3.2). Sequences can be retrieved from GenBank using the following accession numbers: (1) for the Ya5NBC188-containing locus-human: AY055347; owl monkey: AY055348; gorilla: AY055349; chimpanzee: AY055350; pygmy chimpanzee: AY055351; green monkey: AY055352; and Orangutan: AY055353; (2) for the Ya5NBC42-containing locus- human chromosome 2: AY055354; chimpanzee: AY055355; gorilla: AY055356; pygmy chimpanzee: AY055357; green monkey: AY055358; owl monkey: AY055359, and two additional human chromosomes containing a duplication of the region on chromosome 20: AY055360 and chromosome 22: AY055361; (3) for the Yb8NBC253-containing locus human AY055362; chimpanzee: AY055363; gorilla: AY055364; pygmy chimpanzee: AY055365; green monkey: AY055366; and orangutan: AY055367; (4) for the owl monkey Ya5NBC91-containing locus: AY055463 and Yb8NBC185-containing locus: AY055464. In addition, sequence alignments can be found on our website (http://batzerlab.lsu.edu).

Acknowledgments

A.M.R. was supported by a Brown Foundation fellowship from the Tulane Cancer Center. This research was supported by National Institutes of Health RO1 GM45668 (to P.L.D.), Department of the Army DAMD17-98-1-8119 to (to P.L.D. and M.A.B.), Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05)-05 and (2000-05)-01 (to M.A.B. and P.L.D.), and award number 2001-IJ-CX-K004 from the Office of Justice Programs, National Institute of Justice, Department of Justice (to M.A.B.). Points of view in this document are those of the authors and do not necessarily represent the official position of the US Department of Justice.

References

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. *et al*. (2001). Initial sequencing and analysis of the human genome. *Nature,* **409**, 860-921.
2. Deininger, P. L. & Batzer, M. A. (1993). Evolution of Retroposons. In *Evolutionary Biology* (Heckht, M. K. *et al*., eds), pp. 157-196, Plenum Publishing, New York.
3. Batzer, M. A., Kilroy, G. E., Richard, P. E., Shaikh, T. H., Desselle, T. D., Hoppens, C. L. *et al*. (1990). Structure and variability of recently inserted *Alu* family members. **19**, 698-699). *Nucl. Acids Res.* **18**, 6793-6798.
4. Deininger, P. L., Batzer, M. A., Hutchison, C. A. & Edgell, M. H. (1992). Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**, 307-311.
5. Deininger, P. L. & Daniels, G. (1986). The recent evolution of mammalian repetitive DNA elements. *Trends Genet.* **2**, 76-80.
6. Sawada, I., Willard, C., Shen, C.-K., Chapman, B., Wilson, A. C. & Schmid, C. W. (1985). Evolution of *Alu* family repeats since the divergence of human and chimpanzee. *J. Mol. Evol.* **22**, 316-322.
7. Shen, M. R., Batzer, M. A. & Deininger, P. L. (1991). Evolution of the master *Alu* gene(s). *J Mol. Evol.* **33**, 311-320.
8. Batzer, M. A., Schmid, C. W. & Deininger, P. L. (1993). Evolutionary analyses of repetitive DNA sequences. *Methods Enzymol.* **224**, 213-232.
9. Roy, A. M., Carroll, M. L., Kass, D. H., Nguyen, S. V., Salem, A.-H., Batzer, M. A. *et al*. (1999). Recently integrated human *Alu* repeats: finding needles in the haystack. *Genetica,* **107**, 149-161.
10. Roy, A. M., Carroll, M. L., Nguyen, S. V., Salem, A. H., Oldridge, M., Wilkie, A. O. *et al*. (2000). Potential gene conversion and source genes for recently integrated *Alu* elements. *Genome Res.* **10**, 1485-1495.
11. Roy-Engel, A. M., Carroll, M. L., Vogel, E., Garber, R. K., Nguyen, S. V., Salem, A.-H. *et al*. (2001). *Alu* insertion polymorphisms for the study of human genomic diversity. *Genetics,* **159**, 279-290.
12. Leeflang, E. P., Liu, W. M., Chesnokov, I. N. & Schmid, C. W. (1993). Phylogenetic isolation of a human *Alu* founder gene: drift to new subfamily identity. *J. Mol. Evol.* **37**, 559-565.
13. Carroll, M. L., Roy-Engel, A. M., Nguyen, S. V., Salem, A.-H., Vogel, E., Vincent, B. *et al*. (2001). Large-scale analysis of the *Alu* Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J. Mol. Biol.* **311**, 17-40.
14. Deininger, P. L. & Batzer, M. A. (1995). SINE master genes and population biology. In *The Impact of Short, Interspersed Elements (SINEs) on the Host Genome* (Maraia, R., ed.), pp. 43-60, R. G. Landes, Georgetown, TX.
15. Deininger, P. L. & Batzer, M. A. (1999). *Alu* repeats and human disease. *Mol. Genet. Metab.* **67**, 183-193.
16. Miyamoto, M. M., Slightom, J. L. & Goodman, M. (1987). Phylogenetic relations of humans and African apes from DNA sequences in the psi eta-globin region. *Science,* **238**, 369-373.
17. Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl Acad. Sci. USA,* **94**, 1872-1877.
18. Kass, D. H., Raynor, M. E. & Williams, T. M. (2000). Evolutionary history of B1 retroposons in the genus Mus. *J. Mol. Evol.* **51**, 256-264.
19. Nikaido, M., Rooney, A. P. & Okada, N. (1999). Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl Acad. Sci. USA,* **96**, 10261-10266.
20. Arcot, S. S., Adamson, A. W., Risch, G. W., LaFleur, J., Robichaux, M. B., Lamerdin, J. E. *et al*. (1998).

High-resolution cartography of recently integrated human chromosome 19- specific *Alu* fossils. *J. Mol. Biol.* **281**, 843-856.
21. Cantrell, M. A., Filanoski, B. J., Ingermann, A. R., Olsson, K., DiLuglio, N., Lister, Z. *et al*. (2001). An ancient retrovirus-like element contains hot spots for SINE insertion. *Genetics,* **158**, 769-777.
22. Kass, D. H., Batzer, M. A. & Deininger, P. L. (1995). Gene conversion as a secondary mechanism in SINE evolution. *Mol. Cell. Biol.* **15**, 19-25.
23. Batzer, M. A., Rubin, C. M., Hellmann-Blumberg, U., Alegria-Hartman, M., Leeflang, E., Stern, J. D. *et al*. (1995). Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats. *J. Mol. Biol.* **247**, 418-427.
24. Maeda, N., Wu, C. I., Bliska, J. & Reneke, J. (1988). Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. *Mol. Biol. Evol.* **5**, 1-20.
25. Cooper, D. M., Schimenti, K. J. & Schimenti, J. C. (1998). Factors affecting ectopic gene conversion in mice. *Mamm. Genome,* **9**, 355-360.
26. Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H. & Deininger, P. L. (1987). Clustering and sub-family relationships of the *Alu* family in the human genome. *Mol. Biol. Evol.* **4**, 19-29.
27. Ardlie, K., Liu-Cordero, S. N., Eberle, M. A., Daly, M., Barrett, J., Winchester, E. *et al*. (2001). Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am. J. Hum. Genet.* **69**, 582-589.
28. Frisse, L., Hudson, R. R., Bartoszewicz, A., Wall, J. D., Donfack, J. & Di Rienzo, A. (2001). Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831-843.

Edited by J. Karn

(Received 19 November 2001; accepted 3 January 2002)