8-1-2003

Recently integrated Alu elements and human genomic diversity

Abdel Halim Salem
Suez Canal University

Gail E. Kilroy

W. Scott Watkins

Lynn B. Jorde

Mark A. Batzer

Recommended Citation

Salem, A., Kilroy, G., Watkins, W., Jorde, L., & Batzer, M. (2003). Recently integrated Alu elements and human genomic diversity. *Molecular Biology and Evolution, 20* (8), 1349-1361. https://doi.org/10.1093/molbev/msg150

Recently Integrated Alu Elements and Human Genomic Diversity

Abdel-Halim Salem,† Gail E. Kilroy,* W. Scott Watkins,‡ Lynn B. Jorde,‡ and Mark A. Batzer**

*Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University; †Department of Anatomy, Faculty of Medicine, Suez Canal University, Ismailia, Egypt; and ‡Department of Human Genetics, University of Utah Health Sciences Center

A comprehensive analysis of two Alu Y lineage subfamilies was undertaken to assess Alu-associated genomic diversity and identify new Alu insertion polymorphisms for the study of human population genetics. Recently integrated Alu elements (283) from the Yg6 and Yi6 subfamilies were analyzed by polymerase chain reaction (PCR), and 25 of the loci analyzed were polymorphic for insertion presence/absence within the genomes of a diverse array of human populations. These newly identified Alu insertion polymorphisms will be useful tools for the study of human genomic diversity. Our screening of the Alu insertion loci also resulted in the recovery of several ''young'' Alu elements that resided at orthologous positions in nonhuman primate genomes. Sequence analysis demonstrated these ''young'' Alu insertions were the products of gene conversion events of older, preexisting Alu elements or independent parallel forward insertions of older Alu elements in the same short genomic region. The level of gene conversion between Alu elements suggests that it may have an influence on the single nucleotide polymorphism within Alu elements in the genome. We have also identified two genomic deletions associated with the retroposition and insertion of Alu Y lineage elements into the human genome. This type of Alu retroposition–mediated genomic deletion is a novel source of lineage-specific evolution within primate genomes.

Introduction

As short interspersed elements (SINEs), Alu repeats are the largest family of mobile genetic elements within the human genome, having reached a copy number of over 1,000,000 during the past 65 Myr (Batzer and Deininger 2002). Alu elements have achieved this copy number by duplicating via an RNA intermediate in a process termed retroposition (Weiner, Deininger, and Efstratiadis 1986). During retroposition, the RNA copy is reverse transcribed by target primed reverse transcription (TPRT) and subsequently integrated into the genome (Luan et al. 1993; Kazazian and Moran 1998; Kajikawa and Okada 2002). Although unable to retropose autonomously, Alu elements are thought to borrow the factors that are required for their amplification from the long interspersed element (LINE) retrotransposon family (Sinnett et al. 1992; Boeke 1997; Kajikawa and Okada 2002), which encodes a protein with endonuclease and reverse transcriptase activity (Feng et al. 1996; Jurka 1997). Full-length Alu elements are approximately 300 bp in length and commonly found in introns, untranslated regions of genes, and intergenic genomic regions (Deininger and Batzer 1993; Makalowski, Mitchell, and Labuda 1994).

Phylogenetic studies of Alu elements suggest that only a small number of Alu elements termed ''master'' or source genes are retropositionally competent (Deininger et al. 1992). Over time, the gradual accumulation of new mutations within these ''master'' or source genes created a hierarchy of Alu subfamilies (Deininger et al. 1992). Diagnostic mutation sites can be used to classify each individual element according to subfamily and to stratify Alu subfamily members based upon age from the oldest (designated J) to intermediate (S) and youngest (Y) (Batzer et al. 1996*b*). Some young Alu subfamilies have amplified so recently that they are virtually absent from the genomes of nonhuman primates (Batzer and Deininger 2002). As a result of the recent integration of some Alu elements into the human genome, individual humans may be polymorphic for the presence/absence of the ''young'' Alu elements at particular genomic loci (Batzer and Deininger 1991; Perna et al. 1992; Batzer et al. 1994). Since the likelihood of two Alu elements independently inserting into the same exact location of the genome is extremely small, and because there are no known biological mechanisms for the specific excision of Alu elements from the genome, Alu insertions can be considered identical by descent or homoplasy free characters for the study of human population genetics (Roy-Engel et al. 2002).

SINE insertion site homoplasy may occur across distantly related taxa as a function of evolutionary time and variable retroposition rates within various species and can limit the application of SINEs to deep evolutionary questions (Hillis 1999; Cantrell et al. 2001; Roy-Engel et al. 2002). Fortunately, the application of SINE elements to the study of human population genetics is thought to be homoplasy-free as a result of the short evolutionary time frame involved and the current relatively low rate of Alu retroposition within the human genome (Roy-Engel et al. 2002).

We have previously characterized a large number of recently integrated Alu elements found in the human genome that fall in four distinct lineages, termed Ya, Yb, Yc, and Yd based upon their diagnostic mutations (Carroll et al. 2001; Roy-Engel et al. 2001; Xing et al. 2003). Here, we have analyzed 283 members of two newly identified Alu subfamilies termed Yg6 and Yi6 (Jurka 2000; Jurka et al. 2002). We have identified several elements that have been subjected to gene conversion, some that have been involved in lineage-specific deletions, and several new Alu insertion polymorphisms that will be useful tools for the study of the human population genetics. This large data set allows us to begin to estimate the impact of these evolutionary processes on the architecture of primate genomes.

Key words: primates, deletion, insertion polymorphism, gene conversion.

E-mail: mbatzer@lsu.edu.

Mol. Biol. Evol. 20(8):1349–1361. 2003
DOI: 10.1093/molbev/msg150
Molecular Biology and Evolution, Vol. 20, No. 8,

Materials and Methods

Computational Analyses

Alu Yg and Yi elements were identified from the draft sequence of the human genome (August 6, 2001, UCSC GoldenPath assembly) using Basic Local Alignment Search Tool (Blast) (Altschul et al. 1990) queries of the draft sequence to identify exact complements to the oligonucleotides 5′-ATGGTGGCGCGCGCCTGTAGTCCCAGCTACA-3′and 5′-TGCGCCACTGCACTCCCGCCTGGGCC-3′ that are diagnostic for the Alu Yg and Yi lineages (respectively) as shown in figure 1. Using this approach, we identified 160 Yg insertion elements (141 of which were unique) that shared six diagnostic base positions and composed the Alu Yg6 subfamily. We also screened and found 123 elements (104 of which were unique) that shared six diagnostic base positions that compose the Alu Yi6 subfamily. All of the exact complements to the oligonucleotide queries, along with 1,000 bp of adjacent flanking unique DNA sequence, were excised and stored as unique files and subjected to additional analysis as outlined previously (Roy et al. 1999; Carroll et al. 2001; Roy-Engel et al. 2001). A complete list of all the Alu elements identified in the searches is available as online Supplementary Material at the journal's Web site and on our Web site (http://batzerlab.lsu.edu) under publications.

PCR primers for each Alu repeat were designed from flanking unique DNA sequences adjacent to individual Yg6 and Yi6 Alu elements using the Primer3 software (Whitehead Institute for Biomedical Research, Cambridge, Mass.) (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). The resultant PCR primers were screened against the GenBank nonredundant database for the presence of repetitive elements using the BLAST program, and primers that resided within known repetitive elements were discarded and new primers were designed. The sequences of the oligonucleotide primers, annealing temperatures, PCR product sizes, and chromosomal locations for all Yg6 and Yi6 elements are shown in tables 1 and 2 of Supplementary Material online and can be found on our Web site (http://batzerlab.lsu.edu).

DNA Samples

Diverse human DNA samples were available from previous studies (Roy et al. 1999; Carroll et al. 2001; Roy-Engel et al. 2001). Additional human DNA samples from South America (human diversity panels HD 17 and 18) that contained individuals from the Andes, Brazil, Guyana, and Venezuela were purchased from the Coriell Institute for Medical Research, Camden, N.J. The cell lines used to isolate DNA samples were as follows: chimpanzee (*Pan troglodytes*) WES (ATCC CRL1609); lowland gorilla (*Gorilla gorilla*) Coriell AG05251B, Ggo-1 (primary gorilla fibroblasts) provided by Dr. Stephen J. O'Brien, National Cancer Institute, Frederick, Md.; bonobo (*Pan pansicus*) Coriell AG05253A; orangutan (*Pongo pygmaeus*) ATCC CRL6301; green monkey (*Cercopithecus aethiops*) ATCC CCL70 (Old World monkey); and owl monkey (*Aotus trivirgatus*) ATCC CRL 1556 (New World monkey). Cell lines were maintained as directed by the source, and DNA isolations were performed using Wizard genomic DNA purification (Promega). Additional nonhuman primate DNA samples (*Pan troglodytes, Pan paniscus, Gorilla gorilla, Pongo pygmaeus, Macaca mulatta* [Old World monkey], *Macaca nemestrina* [Old World monkey], *Saguinus labiatus* [New World monkey], *Lagothrix lagotricha* [New World monkey], *Ateles geoffroyi* [New World monkey] and *Lemur catta* [prosimian]) were acquired as a primate phylogenetic panel (PRP00001) from the Coriell Institute for Medical Research.

PCR Amplification

PCR amplification of 244 individual Alu Yg6 and Alu Yi6 subfamily members was carried out in 25 μl reactions containing 20 to 100 ng of template DNA, 40 pM of each oligonucleotide primer (shown in tables 1 and 2 of Supplementary Material online), 200 μM dNTPs, in 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.4), and Taq DNA polymerase (1.25 units). Each sample was subjected to the following amplification for 32 cycles: an initial denaturation of 150 s at 94°C, 1 min denaturation at 94°C, 1 min at the annealing temperature (specific for each locus), and extension at 72°C for one min. After the cycles, a final extension was performed at 72°C for 10 min. For analysis, 20 μl of each sample was fractionated on a 2% agarose gel with 0.05 μg/ml ethidium bromide. PCR products were directly visualized using UV fluorescence. Phylogenetic analysis of all the ascertained Alu elements was determined by PCR amplification of nonhuman primate DNA samples. The human genomic diversity associated with each Alu element was determined by the amplification of 20 individuals from each of four populations (African American, Asian, European, and South American).

Sequence Analysis

DNA sequencing was performed on gel-purified PCR products that had been cloned using the TOPO TA cloning vector (Invitrogen) and chain termination sequencing (Sanger, Nicklen, and Coulson 1977) on an Applied Biosystems 3100 automated DNA sequencer. The sequence of the nonhuman primate Yi6AH41, Yg6AH42, Yg6AH79, Yi6AH79, Yi6AH55, Yi6AH121, Yi6AH36, Yi6AH46, Yi6AH87, Yg6AH77, and Yg6AH134 ortholog loci have been assigned GenBank accession numbers AY190763 to AY190817 and AY219790 to AY219800. Sequence alignments for all of the Yg6 and Yi6 subfamily members were performed using MegAlign software (DNAStar version 3.1.7 for Windows 3.2). The ages of the Alu Yg6 and Yi6 subfamilies were calculated as previously described (Batzer et al. 1990; Batzer et al. 1995; Carroll et al. 2001; Roy-Engel et al. 2001). Multiple sequence alignments that contain all of the members of the Yg6 and Yi6 subfamilies can be found on the journal's Web site as Supplementary Material and on our Web site (http://batzerlab.lsu.edu) under publications.

Results

Alu Yg6 and Yi6 Sequence Analysis

To identify Alu Yg6 and Yi6 elements recently inserted into the human genome, we searched the draft

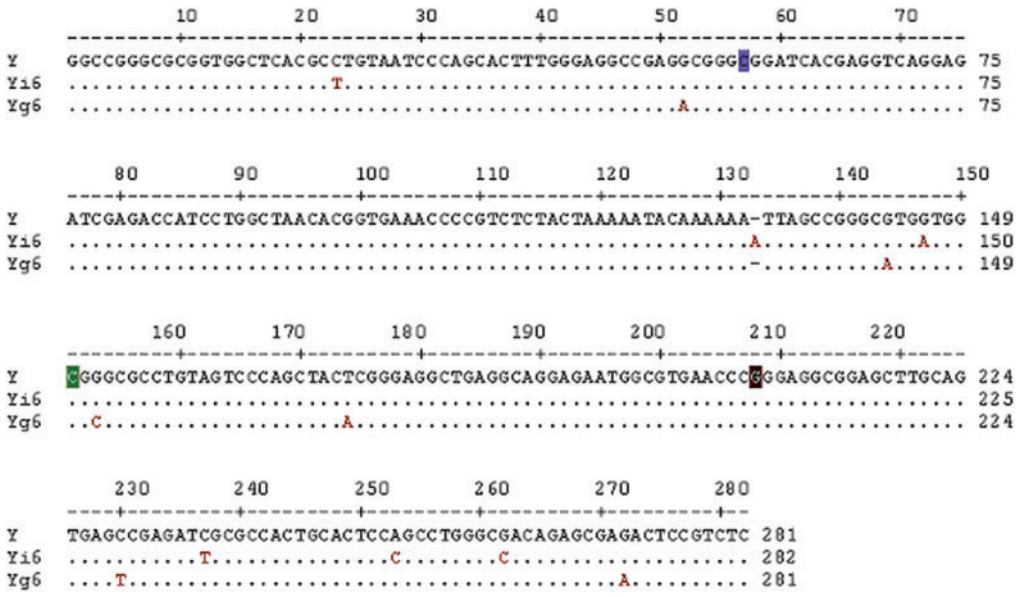

FIG. 1.—Sequence alignment of Alu Yg6 and AluYi6 subfamilies. The consensus sequence for the Alu Y subfamily is shown at the top. The sequences of the Alu Yg6 and Alu Yi6 subfamilies are shown below. The dots below represent the same nucleotides as the consensus sequence. Mutations are shown as the correct base (red) for each of the subfamilies. Each of the newer subfamilies such as Yg6 or Yi6 have all of the mutations of the ancestral Alu Y elements along with six additional mutations that are diagnostic for the particular Alu subfamily. The Yi6.1 subfamily has a T instead of C at the position denoted in purple, Yi6.2 has T instead of C (marked in green), and Yi6.3 has T instead of G (marked in brown) as compared with the Alu Y consensus.

sequence of the human genome (database version: BlastN 2.2.1 [April 13, 2001]) using Blast (Altschul et al. 1990) for the presence of Alu repeats using oligonucleotide sequences complementary to each of the subfamilies (outlined in *Materials and Methods*). We identified 160 Alu Yg6 elements (141 have all six diagnostic mutations, and the rest have five of the six diagnostic mutations of the subfamily) and 123 Alu Yi6 elements (104 have all six diagnostic mutations of the subfamily and 19 elements have either three, four, or five diagnostic mutations) from 2.868×10^9 bp of available human draft genomic sequence. Extrapolating this number to the actual size of the human genome (3.162×10^9 bp), we estimate that these Alu subfamilies contain about 176 and 136 elements for the Yg6 and Yi6 subfamilies, respectively.

To estimate the average ages for the Yg6 and Yi6 Alu subfamilies, we compared the individual Alu repeats with their respective subfamily consensus sequences and calculated the mutation density (from all the subfamily members) and used a neutral rate of evolution to estimate the average subfamily ages as previously described in detail (Carroll et al. 2001). For this analysis we divided the nucleotide substitutions within the elements in each family into those that occurred at CpG dinucleotides and those that occurred at non-CpG nucleotides. The distinction between types of mutations is made because the CpG dinucleotides mutate at a rate that is about 10 times faster than non-CpG positions (Labuda and Striker 1989; Batzer et al. 1990) as a result of the deamination of 5-methylcytosine (Bird 1980). In addition, all insertions, deletions, and 5′ truncations were excluded from our calculations. We also excluded Alu elements that did not have all the subfamily-specific diagnostic mutations from the analysis because these Alu elements are largely products of gene conversion events involving older preexisting Alu elements and their inclusion in the analysis would artificially inflate the subfamily age estimates. A total of 130 non-CpG and 161 CpG mutations occurred within the 134 Alu Yg6 subfamily members used in this analysis. For the 96 Alu Yi6 subfamily members analyzed, a total of 183 non-CpG and 149 CpG mutations were observed. Using a neutral rate of evolution for primate intervening DNA sequences of 0.15% per Myr (Miyamoto, Slightom, and Goodman 1987) and the non-CpG mutation density (number of non-CpG mutations divided by the total number of non-CpG bases in the analyzed sequences) of 0.42 % (130/30,820) within the 134 Yg6 Alu elements yields an estimated age of 2.81 Myr for the Yg6 subfamily members. Using only non-CpG mutations in the 96 AluYi6 sequences yields a mutation density of 0.81% (183/22,656) and age estimate of 5.39 Myr old for the Yi6 subfamily.

We can also estimate the ages of each Alu subfamily using CpG-based mutations. The only difference in the estimate is to multiply the CpG mutation density by a mutation rate that is approximately 10 times the non-CpG rate, as previously described (Labuda and Striker 1989; Batzer et al. 1990). In this case, we calculate an average CpG mutation density for the Yg6 subfamily (161 mutations/ 6,700 total CpG bases) of 2.40% and an average CpG mutation density for the Yi6 subfamily (149 mutations/ 4,416 total CpG bases) of 3.49%. Using a neutral rate of evolution for CpG-based sequences of 1.5%/Myr yields average age estimates of 1.65 and 2.30 Myr old for the Yg6 and Yi6 Alu subfamilies, respectively. If we assume a linear rate of expansion for these Alu subfamilies, then the oldest elements would be approximately two times the average ages with an initial expansion of these Alu subfamilies 3.3 to 4.6 MYA. Thus, both estimates are

Table 1
Alu Yg6 and Yi6 Endonuclease Sites

Sites	Alu Yg6	Alu Yi6	Total
TTTT/A	61	34	95
TCTT/A	21	29	50
TTTT/G	11	14	25
CTTT/A	10	10	20
TTCT/A	9	10	19
TCTT/G	3	5	8
TTTT/C	6	1	7
TTTC/A	3	4	7
ATTT/C	3	1	4
TGTT/A	2	2	4
TTTT/T	3	1	4
CTTT/C	2	0	2
GATT/A	2	0	2
TATT/A	1	1	2
TCCT/A	1	1	2
TTCT/G	1	1	2
TTCC/C, TCTG/T, GTTT/G, CTTT/T, ATTC/A, AGTT/A, TCTC/A, CTCT/A	0	One each	8
TTTA/A, TTCT/C, GCTA/G, GCTT/C, TTTA/T, CTTT/G, AAAA/T, ATTC/T, ATTT/A	One each	0	9

consistent with the initiation of the expansion of the Yg6 and Yi6 Alu subfamilies that is roughly coincident with the divergence of humans and African apes, which is thought to have occurred 4 to 6 MYA. The average age estimates for mobile elements based upon CpG mutation density are typically more accurate than the non-CpG–based estimates because they are less likely to be influenced by sequencing errors as a result of the smaller number of total bases that are sequenced and utilized to generate the CpG-based age estimates (Roy-Engel et al. 2002).

The Yi6 subfamily gave rise to three new derivative Alu subfamilies termed Yi6.1 (21 members), Yi6.2 (57 members), and Yi6.3 (16 members) that have the six diagnostic mutations of Yi6 subfamily in addition to new subfamily-specific mutations (fig. 1). The estimated age of these newly identified Alu subfamilies are 5.02, 2.56, and 2.26 Myr using the non-CpG mutation density, and 2.39, 2.04, and 1.46 Myr based upon the CpG mutation density, respectively.

One hallmark of the integration of an Alu repeat into the genome is the generation of target site duplications flanking newly integrated elements. Of the 283 elements examined, we were able to identify clear target site duplications for 270 elements. The direct repeats of the individual elements range in size from 9 to 21 nt. These types of direct repeats are fairly typical of recently integrated Alu family members (Batzer et al. 1990; Jurka 1997).

We also predicted the endonuclease cleavage sites for the 270 Alu insertions that had clear target site duplications. A complete list of endonuclease cleavage sites is shown in table 1. All but four of the predicted endonuclease sites matched cleavage sites previously reported (Feng et al. 1996; Jurka 1997; Cost and Boeke 1998). The four previously undefined sites may be attributed to a nonstringent or ''relaxed'' human endonuclease with less specificity (Kajikawa and Okada 2002). Alternatively, these four Alu insertions may be the products of endonuclease independent insertion as part of double stranded DNA break repair similar to that previously reported for LINE elements (Morrish et al. 2002).

The appearance of 5′ truncations within a number of the Alu elements (24 elements, which is about 8.5% of the total) presumably occurred as a result of incomplete reverse transcription or improper integration into the genome rather than by postintegration instability. All of the Yi6 and Yg6 Alu family members analyzed have oligo-dA–rich tails, except one element Yg6AH116 that has both a 5′ and 3′ truncation, of 5 to 50 nt in length. The 3′ oligo-dA–rich tails of many of the elements have accumulated random mutations, beginning the process of the formation of simple sequence repeats of varied complexity. The oligo-dA–rich tails and middle A-rich regions of Alu elements have previously been shown to serve as nuclei for the genesis of simple sequence repeats (Arcot et al. 1995).

Phylogenetic Origin

To determine the phylogenetic time of origin of each Alu subfamily member (Yi6 and Yg6) in the primate lineage, we amplified a series of human and nonhuman primate DNA samples using the polymerase chain reaction (PCR) and the oligonucleotide primers shown in tables 1 and 2 of Supplementary Material online. Most of the 160 Yg6 Alu family members were absent from nonhuman primate genomes. However, four Alu elements (Yg6AH42, Yg6AH77, Yg6AH79, and Yg6AH134) had PCR amplification patterns that were unanticipated (PCR products about the size of an Alu-filled site in the nonhuman primate genomes), suggesting that these elements had retroposed much earlier in primate evolution than we suspected (fig. 2). In the Yi6 subfamily, 123 elements were assayed and only seven loci had larger PCR products in humans and Old World monkeys (Yi6AH41), owl monkey (Yi6AH79), great apes (Yi6AH36, Yi6AH46, and Yi6AH87), or all the nonhuman primates tested (Yi6AH55 and Yi6AH121), suggesting either the selective loss of the Alu repeat in some nonhuman primates, parallel independent insertion of Alu elements in multiple primate genomes, or that the insertion of some of the Alu elements predated the radiation of humans and nonhuman primates. Interestingly, the Alu YiAH36 element was present in the human and chimpanzee genomes and absent from the genomes of gorillas and other more evolutionarily distant primates. The results of the PCR-based phylogenetic analysis of orthologous loci are shown in table 2. Detailed sequence analysis of all of these unusual Alu elements indicated that three types of events had occurred: (1) gene conversions of older preexisting Alu elements by an element belonging to a different Alu subfamily, (2) parallel independent insertion of different Alu elements in very close, but not identical, genomic locations, or (3) Alu-mediated deletions of the human genomic sequence during retroposition (as outlined below).

Alu Gene Conversion Events

Gene conversion between Alu elements exerts a significant influence on the accumulation of single

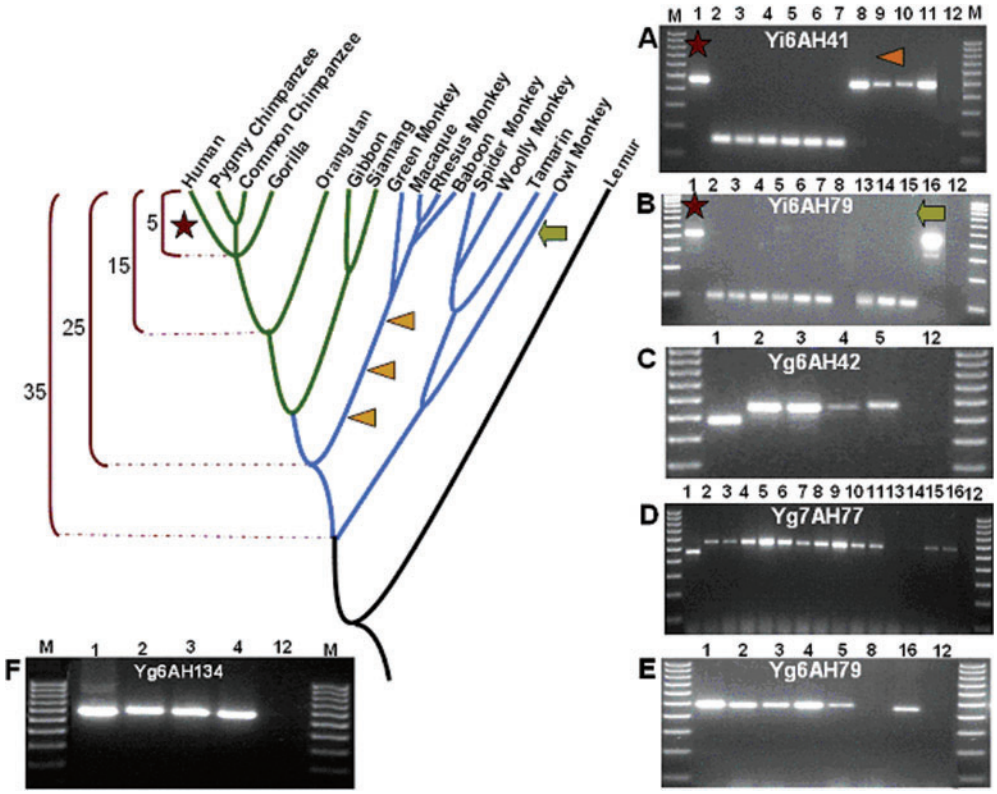

FIG. 2.—Gene conversion and parallel independent insertion of Alu repeats. A schematic of the primate evolutionary tree is shown on the left. Estimated evolutionary time periods between the different primate speciation events are indicated on the left in Myr (Goodman et al. 1998). A dark red star denotes the potential time period for the insertion of Alu Yi6AH41 and Alu Yi6AH79. Yellow arrowheads denote arbitrary examples of the time period between the parallel independent insertion events recovered from Old World monkey genomes. Green arrows represent arbitrary time periods for the parallel independent insertion of Alu elements in the owl monkey genome. PCR amplification of selected loci in different nonhuman primates resulted in the recovery of three types of events: gene conversion shown in C, E, and F, independent insertion within the same locus shown in A and B, and Alu retroposition–mediated deletions shown in C and D. The lanes are 1, human; 2, pygmy chimpanzee; 3, common chimpanzee; 4, gorilla; 5, orangutan; 6, gibbon; 7, siamang; 8, green monkey; 9, macaque; 10, rhesus monkey; 11, baboon; 12, negative control; 13, spider monkey; 14, woolly monkey; 15, tamarin; 16, owl monkey; M, molecular marker.

nucleotide polymorphism within the human genome (Roy et al. 2000; Batzer and Deininger 2002). We identified two Alu Yi6 subfamily members (Yi6AH121 and Yi6AH55) that appeared to have been subjected to partial gene conversion at their 3′ ends. Alu Yi6AH121 contains four mutations that are diagnostic for the Yi6 subfamily and six mutations that are diagnostic for the Sx subfamily. Several nonhuman primates that were sequenced had diagnostic mutations for the Alu Sx subfamily only. Interestingly, in the owl and woolly monkey genomes, we found an insertion of another Alu Sg element within an Alu Sx subfamily member. AluYi6AH55 has four mutations that are diagnostic for theYi6 subfamily, whereas none of the nonhuman primates that were sequenced contained these mutations. Each of these gene conversions involved sequence exchanges in short contiguous sequences, suggesting that they were products of gene conversion rather than a series of homoplasic point mutations. Another three Alu-containing loci were involved in full gene conversion events, (Yg6AH42, Yg6AH79, and Yg6AH134). In these cases, the orthologous Alu elements have similar flanking sequences and direct repeats, although they are not precisely identical due to the random mutations that accumulated over time. DNA sequence analysis of these loci showed that the Alu element present in the Yg6AH42-containing and Yg6AH79-containing loci of all the nonhuman primates belonged to the Alu Sq subfamily and to the Alu Y subfamily in the case of Yg6AH134. This suggests that the gene conversion of older preexisting Alu elements from the Sq or Y subfamilies to Alu Yg6 subfamily members in the human genome took place after the radiation of humans from other African Apes, which is thought to have occurred 4 to 6 MYA (Miyamoto, Slightom, and Goodman 1987).

Alu Retroposition–Mediated Genomic Deletions

We have also identified two deletions of part of the human genome associated with an Alu retroposition. These deletions were identified in loci Yg6AH42 and Yg6AH77. In the case of Yg6AH42, the deletion was also associated with a gene conversion and involved 68 bp of the 3′ flanking region (fig. 3). For Alu Yg6AH77, the Alu element replaced about 300 bp of the genomic sequence that was identified in the nonhuman primate genomes. Based on our data, we estimate the frequency of Alu retroposition mediated deletions of approximately 0.82% (2/244).

Table 2
PCR Analysis of Orthologous Loci for the Presence or Absence of Alu Inserts

Alu Element	Human	Pygmy Chimpanzee	Common Chimpanzee	Gorilla	Orangutan	Green Monkey	Owl Monkey	Types
Yg6AH42	+(Yg6)	+(Sq)	+(Sq)	+(Sq)	+(Sq)	0	0	GC+deletion
Yg6AH79	+(Yg6)	+(Sq)	+(Sq)	+(Sq)	+(Sq)	0	Sq	GC
Yg6AH134	+(Yg6)	+(Y)	+(Y)	+(Y)	0	0	0	GC
Yi6AH41	+(Yi6)	—	—	—	—	+(Y)	0	Ind
Yi6AH79	+(Yi6)	—	—	—	—	0	+(Sc)	Ind
Total analyzed	283	176	159	158	170	80	60	

NOTE.— +, 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 indicates gene conversion; Ind indicates independent insertion.

Parallel Alu Insertions and Amplification Rates

We have also identified two parallel independent Alu insertion events into the same genomic region. One parallel Alu insertion was present in all the Old World monkey genomes tested (green monkey, macaque, rhesus, and baboon), within the same locus where an Alu Yi6 element was located in the human genome. This suggests that the parallel insertion occurred sometime after the divergence of humans from Old World monkeys, but before the radiation of the Old World monkeys. The second parallel Alu insertion involved the Yi6AH79 locus and was only found in a single nonhuman primate, the owl monkey genome. The Alu elements present in the Old World monkey and owl monkey genomes belong to AluY and Alu Sc subfamilies, respectively. The insertion sites for each of these events were not identical to the human insertion site and were localized 3 bp upstream of the human insertion site in the case of Alu Yi6AH41 (fig. 4) and 4 bp upstream of the human insertion site and inverted orientation for Alu Yi6AH79. Although the integration sites were not identical, we will continue to refer to them as parallel insertions since the Alu elements independently integrated within the same 100-bp region of the nonhuman primate genomes. Previously, we have reported four cases of independent Alu insertion in the owl monkey genome (Roy-Engel et al. 2002; Xing et al. 2003). These are the first reported cases of the independent insertion of an Alu repeat in the Old World monkey lineage.

To estimate the parallel insertion rate in Old World monkey genomes, we used the number of loci analyzed in Old World monkeys (80) and multiplied it by the time elapsed after the radiation of Old World monkeys (25 Myr), giving us a rate of one event per 2,000 million insert site years. To compare it with other primates with no independent insertions detected, we added all of their individual rates. We used the ages indicated in figure 2 and the successful orthologous PCR amplifications in table 2 for each of the different primates: $(283 \times 5) + (176 \times 5) + (159 \times 5) + (158 \times 5) + (170 \times 5) = 4{,}730$ million insert site years. Therefore, the Old World monkey rate of parallel Alu insertion is about 2.4 (4,730/2,000) times faster than the sum of all the rates of other nonhuman primates. The frequency of occurrence of independent insertions in Old World monkeys can be estimated as one event out of 80 successful PCR amplifications (table 2) or 1.25%. The size of the target site tested in our PCR assay is approximately 200 bp; making the total amount of genomic DNA screened 16,000 bp. If we assume that the target site for integration of Alu elements is random, we expect to detect one new insertion in every 16,000 bases. However, Alu elements do not insert completely random, but rather appear to have a site preference for locally A + T rich regions (Jurka 1997), adding a degree of uncertainty to the estimate.

Human Genomic Diversity

To determine the human genomic variation associated with each of the Yg6 and Yi6 Alu subfamily members, each element was subjected to PCR amplification using a panel of human DNA samples as templates. The panel was composed of 20 individuals of European origin, 20 African Americans, 20 Asians and 20 South Americans for a total of 80 individuals (160 chromosomes). Using this approach, 125 Alu Yg6 and 94 AluYi6 subfamily members were monomorphic for the presence of the Alu element, suggesting that these elements integrated in the genome before the radiation of humans. A total of eight Yg6 and Yi6 Alu family members were inserted in other previously unidentified repeated sequences and were not amenable to PCR analysis as a result of paralogous amplification. An additional 31 elements were located in other repetitive regions of the genome that were identified computationally and discarded from further analysis. The remaining elements were polymorphic for the presence of an Alu repeat within the genomes of the test panel individuals (summarized in table 3). Autosomal loci that were polymorphic for the presence/absence of individual Alu insertions were subsequently classified as high, low, or intermediate frequency insertion polymorphisms (tables 4 and 5) with sex-linked polymorphisms shown in table 6. The unbiased heterozygosity values for these Alu insertion polymorphisms were variable and approached the theoretical maximum of 50% in several cases. This suggests that many of these Alu insertion polymorphisms will make excellent markers for the study of human population genetics. Approximately 10.7% (15/140) of the Yg6 and 9.6 % (10/104) of the Yi6 Alu family members were polymorphic for insertion presence/absence within diverse human genomes. In addition, we identified three X chromosome Alu elements that were polymorphic (table 6).

Discussion

The Alu Yg6 and Yi6 subfamilies are characterized by a series of distinct diagnostic mutations, and they both have relatively small copy numbers within the human genome. Some members of each of the subfamilies are

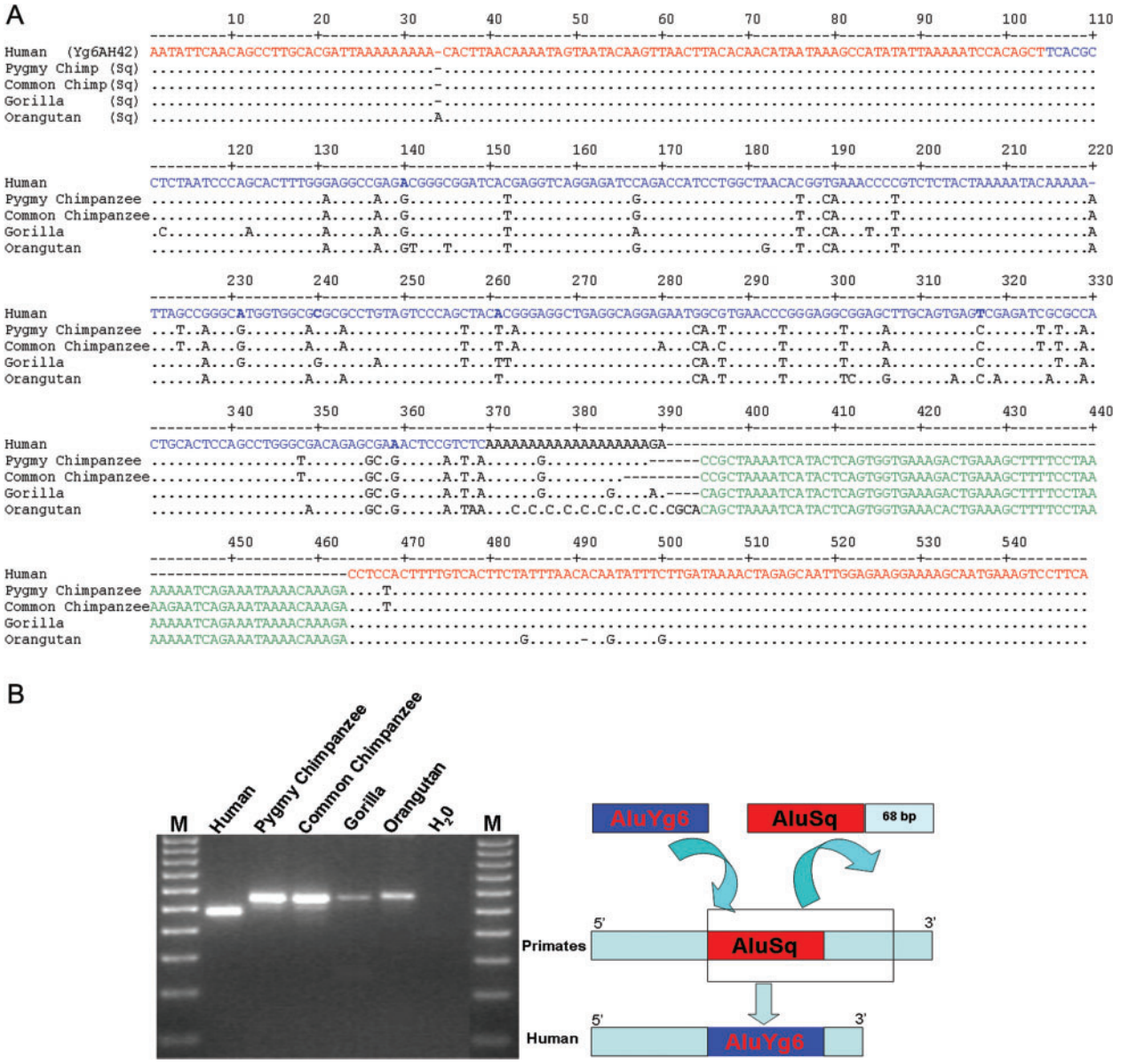

FIG. 3.—Sequence alignment of the Yg6AH42 loci in different primates. (*A*) The human sequence of the Yi6AH42 locus is shown on the top line. Nucleotide substitutions at each position are indicated with the appropriate nucleotide. Deletions are marked by dashes (–). The subfamily of the Alu element present in each nonhuman primate species is indicated in parenthesis. The flanking unique sequence regions are colored in red, and the Alu element is shown in blue. The corresponding sequence of the region deleted in humans is shown in green. (*B*) The bottom part of the figure shows an agarose gel chromatograph of a phylogenetic PCR analysis with schematic diagram depicting the 68-bp deletion of human sequence during an Alu-mediated gene conversion.

polymorphic with respect to insertion presence/absence in the human genome. This suggests that each subfamily has been generated by ''master'' or source Alu elements that were capable of retroposition within the human lineage over the past 4 to 6 Myr since the divergence of humans and African apes. However, the proportion of polymorphic elements within each of the subfamilies is quite low, with only 9.6% of the Yi6 elements and 10.7% of the Yg6 elements being polymorphic. By contrast, many other young Alu subfamilies have levels of insertion polymorphism in excess of 20% (Batzer and Deininger 2002). Therefore, the amplification of these Alu subfamilies within the human genome has occurred at a very low rate and may have recently ceased entirely. The reason for the low level of retroposition within these Alu subfamilies is unknown, although the current amplification rate of Alu elements has decreased by several orders of magnitude from its peak 35 to 60 MYA (Shen, Batzer, and Deininger 1991). Several reasons have been proposed for reduction of Alu retroposition, including altered transcription, Alu RNA secondary structure, or reduced TPRT ability (Deininger and Batzer 1999).

The estimated average ages of 1.65 and 2.30 Myr for the Alu Yg6 and Yi6 subfamilies are consistent with their relatively recent origin in primate genomes. Assuming a linear rate of amplification the oldest members of the Alu

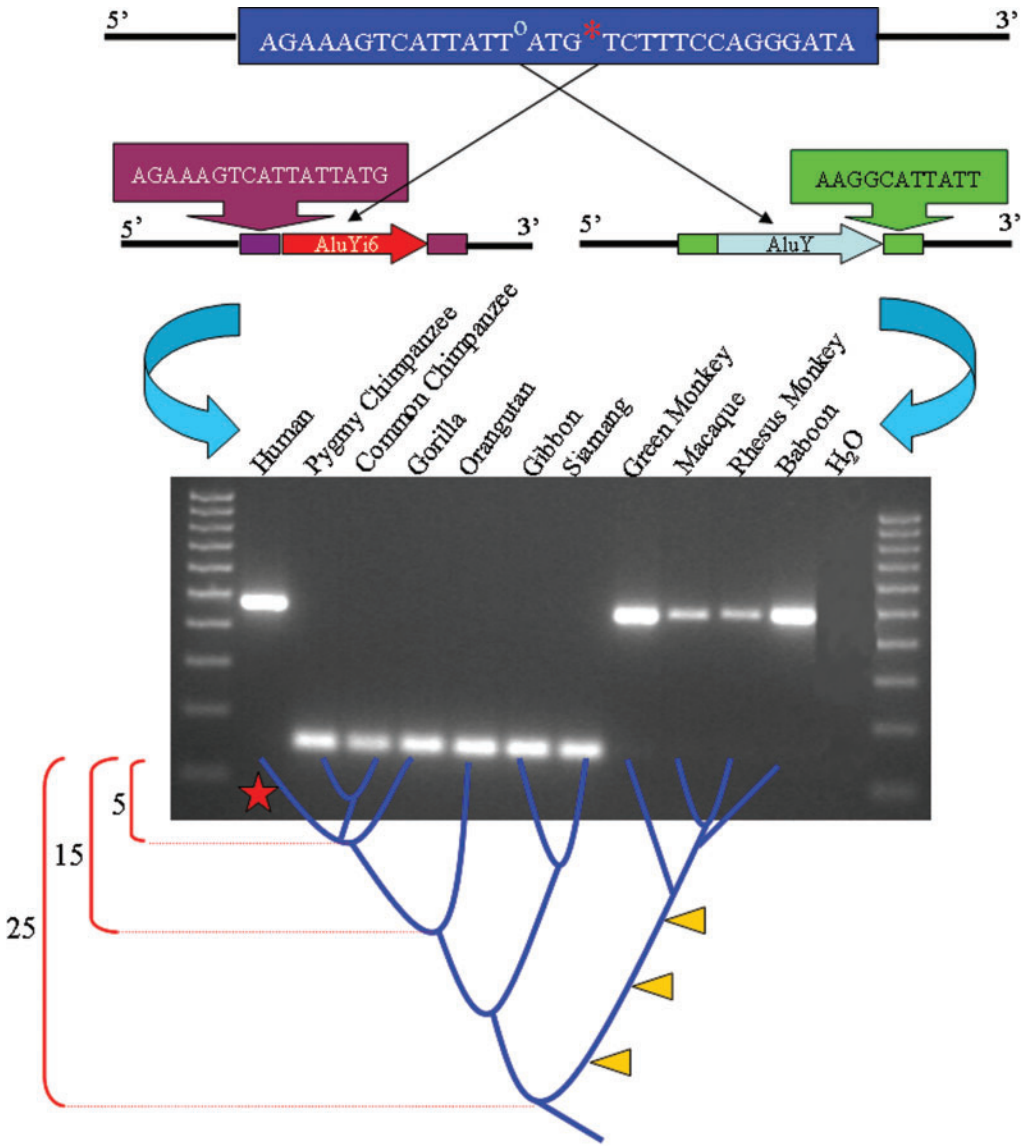

FIG. 4.—Parallel independent Alu insertions at the Yi6AH41 locus. The upper portion of this figure shows a schematic diagram of the parallel insertion of Alu Yi6AH41 in the human genome and an Alu Y subfamily member that occurred in the Old World monkey lineage 3 bp upstream of human insertion site. The schematic shows the insertion of the Yi6 Alu and an Alu Y element in the same orientation. The bottom part of the figure shows an agarose gel chromatograph of a phylogenetic analysis of the locus with a tree of primate evolution superimposed on it. The human genome that was assayed for the presence of Alu Yi6AH41 was fixed for the presence of this Alu element. All of the other nonhuman primate genomes do not contain an Alu element within this 150–bp region, aside from the Old World monkey genomes. The red star denotes the approximate time in primate evolution when the insertion of the Alu Yi6 element in this locus may have occurred. The parallel insertion of the Alu Y subfamily member may have occurred at any point in time since the divergence of humans from Old World monkeys as denoted by the yellow arrows.

Yg6 and Yi6 subfamilies would be twice the average age of each group or 3.3 and 4.6 Myr, respectively. Therefore, the estimated ages for the Alu Yg6 and Alu Yi6 subfamilies are in good agreement with what would be expected for groups of Alu repeats that are largely restricted to the human genome and absent from nonhuman primate genomes, since human and nonhuman primates are thought to have diverged from each other 4 to 6 MYA (Miyamoto, Slightom, and Goodman 1987; Stewart and Disotell 1998).

Several members of the Alu Yg6 and Yi6 subfamilies were polymorphic for insertion presence/absence in diverse human genomes. Alu insertion polymorphisms have proved useful in a number of studies of human population genetics (Perna et al. 1992; Batzer et al. 1994; Hammer 1994; Batzer et al. 1996*a*; Stoneking et al. 1997; Novick et al. 1998; Comas et al. 2000; Jorde et al. 2000; Bamshad et al. 2001; Nasidze et al. 2001; Watkins et al. 2001; Battilana et al. 2002; Romualdi et al. 2002; Bamshad et al. 2003). Individual Alu insertion polymorphisms are useful tools for the study of human population genetics since the Alu alleles are generally thought to be reliable, homoplasy-free characters (Roy-Engel et al. 2002) with a known ancestral state (Perna et al. 1992; Batzer et al. 1994). In addition, there is no known mechanism for the site-specific deletion of Alu insertions

from the genome (Perna et al. 1992; Batzer et al. 1994). Therefore, detailed studies of the human variation associated with the newly identified Alu insertion polymorphisms reported here should prove useful for human population genetics and forensic genomics.

Our data have several implications for Alu insertion and postintegration sequence evolution. First, they support the ''master'' or limited amplification model (Deininger et al. 1992). This model posits 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 the radiation of two species should generate new copies at specific loci that are not shared between primate species. In our analysis, only three elements from the AluYi6 subfamily were recovered in pygmy chimpanzee and common chimpanzee genomes, with two of these three elements also present in the gorilla genome. These data are also in good agreement with our age estimates for these Alu subfamilies. The rest of our ''PCR positives'' from nonhuman primate genomes were either gene conversion events (with or without a deletion), the products of parallel, independent Alu insertions or Alu retoposition associated genomic deletions. Secondly, these data suggest that newly integrated Alu elements are stable integrations within primate genomes and that they are identical by descent. In our study, only two out of 283 loci analyzed contained parallel independent Alu insertions. The rate of parallel Alu insertion events is extremely low when considering the number of loci analyzed and the full length of the evolutionary tree of 6,730 million insert years. Within great apes, we have assayed hundred of sites with a combined total of over 4,730 Myr of site evolution without detecting any parallel Alu insertion events. This represents having sampled across 315 genomic sites analyzed with an average of 15 Myr of evolution per site. Based on this number, if we assume humans diverged from one another as far back as 1 MYA, we would expect to see less than one parallel insertion event per locus in a diverse population of over 4,730 individuals. This estimate is somewhat larger than that published previously (Roy-Engel et al. 2002), however the probability of detecting parallel independent Alu insertions in the human population is still extremely low. Therefore, we conclude that Alu insertion polymorphisms are largely homoplasy-free characters for the study of human evolution.

Gene conversion between Alu repeats has been reported previously (Maeda et al. 1988; Kass, Batzer, and Deininger 1995; Roy-Engel et al. 2002). Here, we have identified and characterized three forward gene conversion events after screening 283 independent Alu-containing loci within the human genome. Based on an examination of low copy number transgenes in the mouse, it has been suggested that the germline recombination machinery in mammals has been evolved to prevent high levels of ectopic recombination between repetitive sequences (Cooper, Schimenti, and Schimenti 1998). 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 can terminate the process resulting in the production of small gene conversion tracts.

Table 3
Summary of Alu Yg6 and Yi6 Analyses

	Alu Yg6	Alu Yi6
Loci analyzed by PCR	140	104
Fixed	125	94
High frequency	2	0
Intermediate frequency	9	8
Low frequency	4	2
Total polymorphic	15	10
Paralog	2	6
Loci not analyzed by PCR		
Inserted in repeats	18	13
Total elements analyzed	160	123

Genomic deletions created upon LINE-1 retrotransposition using cell culture assays have been recently identified (Gilbert, Lutz-Prigge, and Moran 2002). The rate of LINE element deletion was estimated indirectly in the human genome to be about 3% (Kazazian and Goodier 2002; Myers et al. 2002). However, the precise molecular mechanism of the LINE-mediated genomic deletions is still unclear. Recently, an Alu-mediated deletion that resulted in the inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene has been identified (Hayakawa et al. 2001). The deletion of the human CMP-N-acetylneuraminic acid hydroxylase gene involved about 478 bp, including a 92-bp exon along with the replacement of an Alu Sq element in nonhuman primates with an AluY element in the human lineage (Hayakawa et al. 2001). Here we report two new examples of Alu retroposition–mediated deletions that may have been performed by a mechanism similar to that of the LINE element–mediated genomic deletions since Alu and L1 elements utilize a common mobilization pathway (Boeke 1997; Batzer and Deininger 2002; Kajikawa and Okada 2002).

In the first case, Alu Yg6AH42, the deletion appears to have occurred during the process of gene conversion similar to the lineage-specific Alu deletion reported previously (Hayakawa et al. 2001). In the second case, Alu Yg6AH77, two scenarios for the deletion can be envisioned. In the first scenario, the deletion would have occurred before the Alu insertion as a result of double-stranded DNA break repair since this element has no direct repeats (a hallmark of LINE element–mediated, endonuclease-independent, double-stranded DNA break repair) (Morrish et al. 2002). In the second scenario, the deletion would have occurred during the integration of the Alu element in the genome, possibly during TPRT, suggesting a new role for Alu elements in creating human genomic diversity.

Here, we have estimated the frequency of Alu retroposition associated genomic deletions of approximately 0.82%. New Alu integrations have been estimated to occur in vivo at a frequency of one new event in every 10 to 200 births (Deininger and Batzer 1999). If sizable deletions accompany one in every 100 new Alu retroposition events in vivo, the impact on genomic evolution could be substantial. This is not a trivial number of deletions when extrapolated to the copy number of Alu

Table 4
AluYg6 Subfamily Autosomal Allele Frequency and Heterozygosity

	African American					Asian					European/German Caucasian					South American					
	Genotypes					Genotypes					Genotypes					Genotypes					
Locus Name	+/+	+/−	−/−	*f*Yg6	Het[a]	+/+	+/−	−/−	*f*Yg6	Het[a]	+/+	+/−	−/−	*f*Yg6	Het[a]	+/+	+/−	−/−	*f*Yg6	Het[a]	Avg Het[b]
High frequency																					
Yg6AH138	19	1	0	0.98	0.05	20	0	0	1	0	20	0	0	1	0	19	0	0	1	0	0.01
Intermediate frequency																					
Yg6AH28	5	1	14	0.28	0.41	14	6	0	0.85	0.26	3	3	14	0.23	0.36	11	3	6	0.63	0.48	0.38
Yg6AH35	14	6	0	0.85	0.26	7	13	0	0.68	0.45	19	1	0	0.98	0.05	19	0	0	1	0	0.19
Yg6AH55	12	2	1	0.87	0.24	17	2	0	0.95	0.1	18	1	0	0.97	0.05	18	1	0	0.97	0.05	0.11
Yg6AH98	17	2	1	0.9	0.18	4	6	10	0.35	0.47	13	6	1	0.8	0.33	11	9	0	0.78	0.36	0.33
Yg6AH99	5	10	5	0.5	0.51	19	1	0	0.98	0.05	8	6	5	0.58	0.5	9	7	4	0.63	0.48	0.39
Yg6AH108	5	11	4	0.53	0.51	12	7	1	0.78	0.36	13	6	0	0.84	0.27	13	6	1	0.8	0.33	0.37
Yg6AH133	7	6	6	0.53	0.51	9	4	5	0.61	0.49	15	5	0	0.88	0.22	14	3	2	0.82	0.31	0.38
Yg6AH147	13	5	2	0.78	0.36	20	0	0	1	0	14	6	0	0.85	0.26	18	2	0	0.95	0.1	0.18
Low frequency																					
Yg6AH30	1	2	16	0.11	0.19	0	0	20	0	0	0	0	20	0	0	0	0	20	0	0	0.05
Yg6AH44	0	2	18	0.05	0.1	0	0	20	0	0	0	0	20	0	0	0	0	20	0	0	0.02
Yg6AH86	0	3	17	0.08	0.14	0	0	19	0	0	0	0	20	0	0	0	0	20	0	0	0.04
Yg6AH160	0	0	20	0	0	0	1	19	0.03	0.05	0	4	14	0.11	0.2	1	2	16	0.11	0.19	0.11

[a] Unbiased heterozygosity.
[b] Average heterozygosity for all populations.

Table 5
AluYi6 Subfamily Autosomal Allele Frequency and Heterozygosity

	African American					Asian					European/German Caucasian					South American					
	Genotypes					Genotypes					Genotypes					Genotypes					
Locus Name	+/+	+/−	−/−	*f*Yi6	Het[a]	+/+	+/−	−/−	*f*Yi6	Het[a]	+/+	+/−	−/−	*f*Yi6	Het[a]	+/+	+/−	−/−	*f*Yi6	Het[a]	Avg Het[b]
Intermediate frequency																					
Yi6AH16	2	7	11	0.28	0.41	14	5	1	0.83	0.3	2	15	3	0.48	0.51	7	9	3	0.61	0.49	0.43
Yi6AH29	0	11	9	0.28	0.41	2	9	8	0.34	0.5	4	11	5	0.48	0.51	8	10	2	0.65	0.47	0.46
Yi6AH65	3	10	7	0.4	0.49	3	10	7	0.4	0.5	9	6	5	0.6	0.49	8	7	5	0.58	0.5	0.49
Yi6AH86	0	5	15	0.13	0.22	3	9	8	0.38	0.5	8	7	5	0.58	0.5	5	5	10	0.38	0.48	0.42
Yi6AH97	3	13	4	0.48	0.51	0	6	14	0.15	0.3	2	9	9	0.33	0.45	3	13	4	0.48	0.51	0.43
Yi6AH110	6	12	2	0.6	0.49	19	0	0	1	0	18	1	0	0.97	0.05	20	0	0	1	0	0.14
Yi6AH116	1	6	12	0.21	0.34	18	1	1	0.93	0.1	14	3	3	0.78	0.36	13	1	6	0.68	0.45	0.32
Yi6AH125	10	9	1	0.73	0.41	19	1	0	0.98	0.1	19	1	0	0.98	0.05	14	5	1	0.83	0.3	0.2
Low frequency																					
Yi6AH63[c]	0	0	20	0	0	0	0	19	0	0	0	0	20	0	0	0	0	20	0	0	0

[a] Unbiased heterozygosity.
[b] Average heterozygosity for all populations.
[c] This Alu element is absent from every individual tested in our diverse population panel and is present only in the draft sequence of the human genome, making it a low frequency insertion polymorphism.

Table 6
Sex-Linked AluYg6 and AluYi6 Insertion Polymorphism, Genotypes, and Heterozygosity

	African American							Asian							European/German Caucasian							South American							
	Genotype							Genotype							Genotype							Genotype							
	Female		Male					Female		Male					Female		Male					Female		Male					
Name	+/+	+/−	−/−	+	−	fAlu	Het[a]	+/+	+/−	−/−	+	−	fAlu	Het[a]	+/+	+/−	−/−	+	−	fAlu	Het[a]	+/+	+/−	−/−	+	−	fAlu	Het[a]	Avg Het[b]
Intermediate frequency																												
Yg6AH38	1	4	0	14	1	0.8	0.4	6	4	1	6	3	0.71	0.44	11	0	1	8	0	0.94	0.08	8	4	0	6	2	0.81	0.34	0.317
High frequency																												
Yg6AH118	2	3	0	8	2	0.42	0.42	11	0	0	7	0	1	0	11	0	0	7	0	1	0	11	0	0	7	0	1	0	0.104
Low frequency																												
Yi6AH124[c]	0	0	5	0	15	0	0	0	0	11	0	9	0	0	0	0	12	0	8	0	0	0	0	12	0	8	0	0	0

[a] Unbiased heterozygosity, which takes into account sex differences within the calculation.
[b] Average heterozygosity is the average of the population heterozygosity across all four populations. The level of insertion polymorphism was determined as low frequency, the absence of the element from all individuals tested, except one or two heterozygous individuals; intermediate frequency, the Alu element is variable as to its presence or absence in at least one population; and high frequency, the element is present in all individuals in all populations tested, except for one or two heterozygous or absent individuals.
[c] This Alu element is absent from every individual tested in our diverse population panel and is present only in the draft sequence of the human genome, making it a low frequency insertion polymorphism.

elements in the human genome, which is over 1,000,000 (Batzer and Deininger 2002). About 8,000 Alu elements may have been involved in retroposition–mediated deletion events within primate genomes. If each of these deletion events removes 150 bp of genomic sequence, this would mean that Alu retroposition may have been responsible for the deletion of over 1.2 Mb of the primate genomic sequence. If the Alu-associated deletions have involved larger sequences similar to those recently reported for LINE elements (Gilbert, Lutz-Prigge, and Moran 2002), then the impact of these events may be 12 to 120 Mb of lineage-specific deletions. In either case, these types of events represent a novel mechanism of lineage-specific deletion within the primate order. Detailed studies of the orthologous regions of primate genomes deleted in this manner may prove instructive for understanding the genetic basis of the difference between humans and nonhuman primates.

Acknowledgments

This research was supported by Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000–05)-05, (2000–05)-01, and (2001–06)-02 (M.A.B.), National Science Foundation grants BCS-0218338 (M.A.B.) and BCS-0218370 (L.B.J.), and National Institutes of Health RO1 GM59290 (M.A.B. and L.B.J.).

Literature Cited

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

Arcot, S. S., Z. Wang, J. L. Weber, P. L. Deininger, and M. A. Batzer. 1995. Alu repeats: a source for the genesis of primate microsatellites. Genomics **29**:136–144.

Bamshad, M., T. Kivisild, W. S. Watkins et al. (18 co-authors). 2001. Genetic evidence on the origins of Indian caste populations. Genome Res. **11**:994–1004.

Bamshad, M. J., S. Wooding, W. S. Watkins, C. T. Ostler, M. A. Batzer, and L. B. Jorde. 2003. Human population genetic structure and inference of group membership. Am. J. Hum. Genet. **72**:578–589.

Battilana, J., S. L. Bonatto, L. B. Freitas et al. (12 co-authors). 2002. Alu insertions versus blood group plus protein genetic variability in four Amerindian populations. Ann. Hum. Biol. **29**:334–47.

Batzer, M. A., S. S. Arcot, J. W. Phinney et al. (15 co-authors). 1996*a*. Genetic variation of recent Alu insertions in human populations. J. Mol. Evol. **42**:22–29.

Batzer, M. A., and P. L. Deininger. 1991. A human-specific subfamily of Alu sequences. Genomics **9**:481–487.

———. 2002. Alu repeats and human genomic diversity. Nat. Rev. Genet. **3**:370–379.

Batzer, M. A., P. L. Deininger, U. Hellmann-Blumberg, J. Jurka, D. Labuda, C. M. Rubin, C. W. Schmid, E. Zietkiewicz, and E. Zuckerkandl. 1996*b*. Standardized nomenclature for Alu repeats. J. Mol. Evol. **42**:3–6.

Batzer, M. A., G. E. Kilroy, P. E. Richard, T. H. Shaikh, T. D. Desselle, C. L. Hoppens, and P. L. Deininger. 1990. Structure and variability of recently inserted Alu family members. Nucleic Acids Res. **18**:6793–6798.

Batzer, M. A., C. M. Rubin, U. Hellmann-Blumberg, M. Alegria-Hartman, E. P. Leeflang, J. D. Stern, H. A. Bazan, T. H. Shaikh, P. L. Deininger, and C. W. Schmid. 1995. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. J. Mol. Biol. **247**:418–427.

Batzer, M. A., M. Stoneking, M. Alegria-Hartman et al. (13 co-authors). 1994. African origin of human-specific polymorphic Alu insertions. Proc. Natl. Acad. Sci. U S A **91**: 12288–12292.

Bird, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. **8**:1499–504.

Boeke, J. D. 1997. LINEs and Alus—the polyA connection. Nat. Genet. **16**:6–7.

Cantrell, M. A., B. J. Filanoski, A. R. Ingermann, K. Olsson, N. DiLuglio, Z. Lister, and H. A. Wichman. 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. Genetics **158**:769–777.

Carroll, M. L., A. M. Roy-Engel, S. V. Nguyen et al. (16 co-authors). 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J. Mol. Biol. **311**:17–40.

Comas, D., F. Calafell, N. Benchemsi, A. Helal, G. Lefranc, M. Stoneking, M. A. Batzer, J. Bertranpetit, and A. Sajantila. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. Hum. Genet. **107**:312–319.

Cooper, D. M., K. J. Schimenti, and J. C. Schimenti. 1998. Factors affecting ectopic gene conversion in mice. Mamm. Genome **9**:355–360.

Cost, G. J., and J. D. Boeke. 1998. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry **37**:18081–18093.

Deininger, P. L., and M. A. Batzer. 1993. Evolution of retroposons. Evol. Biol. **27**:157–196.

———. 1999. Alu repeats and human disease. Mol. Genet. Metab. **67**:183–193.

Deininger, P. L., M. A. Batzer, C. A. Hutchison, 3rd, and M. H. Edgell. 1992. Master genes in mammalian repetitive DNA amplification. Trends Genet. **8**:307–311.

Feng, Q., J. V. Moran, H. H. Kazazian, Jr., and J. D. Boeke. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **87**:905–916.

Gilbert, N., S. Lutz-Prigge, and J. V. Moran. 2002. Genomic deletions created upon LINE-1 retrotransposition. Cell **110**:315–325.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, and C. P. Groves. 1998. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9**:585–598.

Hammer, M. F. 1994. A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. Mol. Biol. Evol. **11**:749–761.

Hayakawa, T., Y. Satta, P. Gagneux, A. Varki, and N. Takahata. 2001. Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. Proc. Natl. Acad. Sci. USA **98**:11399–11404.

Hillis, D. M. 1999. SINEs of the perfect character. Proc. Natl. Acad. Sci. USA **96**:9979–9981.

Jorde, L. B., W. S. Watkins, M. J. Bamshad, M. E. Dixon, C. E. Ricker, M. T. Seielstad, and M. A. Batzer. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am. J. Hum. Genet. **66**:979–988.

Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. USA **94**:1872–1877.

———. 2000. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. **16**:418–420.
Jurka, J., M. Krnjajic, V. V. Kapitonov, J. E. Stenger, and O. Kokhanyy. 2002. Active Alu elements are passed primarily through paternal germlines. Theor. Popul. Biol. **61**:519–530.
Kajikawa, M., and N. Okada. 2002. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell **111**:433–444.
Kass, D. H., M. A. Batzer, and P. L. Deininger. 1995. Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. Mol. Cell Biol. **15**:19–25.
Kazazian, H. H., Jr., and J. L. Goodier. 2002. LINE drive. retrotransposition and genome instability. Cell **110**:277–280.
Kazazian, H. H., Jr., and J. V. Moran. 1998. The impact of L1 retrotransposons on the human genome. Nat. Genet. **19**: 19–24.
Labuda, D., and G. Striker. 1989. Sequence conservation in Alu evolution. Nucleic Acids Res. **17**:2477–2491.
Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72**:595–605.
Maeda, N., C. I. Wu, J. Bliska, and J. Reneke. 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.
Makalowski, W., G. A. Mitchell, and D. Labuda. 1994. Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet. **10**:188–193.
Miyamoto, M. M., J. L. Slightom, and M. Goodman. 1987. Phylogenetic relations of humans and African apes from DNA sequences in the psi eta-globin region. Science **238**:369–373.
Morrish, T. A., N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato, G. E. Taccioli, M. A. Batzer, and J. V. Moran. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat. Genet. **31**:159–165.
Myers, J. S., B. J. Vincent, H. Udall et al. (12 co-authors). 2002. A comprehensive analysis of recently integrated human Ta L1 elements. Am. J. Hum. Genet. **71**:312–326.
Nasidze, I., G. M. Risch, M. Robichaux, S. T. Sherry, M. A. Batzer, and M. Stoneking. 2001. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. Eur. J. Hum. Genet. **9**:267–272.
Novick, G. E., C. C. Novick, J. Yunis et al. (11 co-authors). 1998. Polymorphic Alu insertions and the Asian origin of Native American populations. Hum. Biol. **70**:23–39.
Perna, N. T., M. A. Batzer, P. L. Deininger, and M. Stoneking. 1992. Alu insertion polymorphism: a new type of marker for human population studies. Hum. Biol. **64**:641–648.
Romualdi, C., D. Balding, I. S. Nasidze, G. Risch, M. Robichaux, S. T. Sherry, M. Stoneking, M. A. Batzer, and G. Barbujani. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res. **12**:602–612.
Roy, A. M., M. L. Carroll, D. H. Kass, S. V. Nguyen, A. H. Salem, M. A. Batzer, and P. L. Deininger. 1999. Recently integrated human Alu repeats: finding needles in the haystack. Genetica **107**:149–161.
Roy, A. M., M. L. Carroll, S. V. Nguyen, A. H. Salem, M. Oldridge, A. O. Wilkie, M. A. Batzer, and P. L. Deininger. 2000. Potential gene conversion and source genes for recently integrated Alu elements. Genome Res. **10**:1485–1495.
Roy-Engel, A. M., M. L. Carroll, M. El-Sawy, A. H. Salem, R. K. Garber, S. V. Nguyen, P. L. Deininger, and M. A. Batzer. 2002. Non-traditional Alu evolution and primate genomic diversity. J. Mol. Biol. **316**:1033–1040.
Roy-Engel, A. M., M. L. Carroll, E. Vogel, R. K. Garber, S. V. Nguyen, A. H. Salem, M. A. Batzer, and P. L. Deininger. 2001. Alu insertion polymorphisms for the study of human genomic diversity. Genetics **159**:279–290.
Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
Shen, M. R., M. A. Batzer, and P. L. Deininger. 1991. Evolution of the master Alu gene(s). J. Mol. Evol. **33**:311–320.
Sinnett, D., C. Richer, J. M. Deragon, and D. Labuda. 1992. Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. J. Mol. Biol. **226**:689–706.
Stewart, C. B., and T. R. Disotell. 1998. Primate evolution—in and out of Africa. Curr. Biol. **8**:R582–R588.
Stoneking, M., J. J. Fontius, S. L. Clifford, H. Soodyall, S. S. Arcot, N. Saha, T. Jenkins, M. A. Tahir, P. L. Deininger, and M. A. Batzer. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. **7**:1061–1071.
Watkins, W. S., C. E. Ricker, M. J. Bamshad, M. L. Carroll, S. V. Nguyen, M. A. Batzer, H. C. Harpending, A. R. Rogers, and L. B. Jorde. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. Am. J. Hum. Genet. **68**:738–752.
Weiner, A. M., P. L. Deininger, and A. Efstratiadis. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**:631–661.
Xing, J., A. H. Salem, D. J. Hedges, G. E. Kilroy, W. S. Watkins, J. E. Schienman, C. B. Stewart, J. Jurka, L. B. Jorde, and M. A. Batzer. 2003. Comprehensive analysis of two Alu Yd subfamilies. J. Mol. Evol. (in press).

David Goldstein, Associate Editor

Accepted April 11, 2003