Alu fossil relics--distribution and insertion polymorphism.

S S Arcot, A W Adamson, J E Lamerdin, et al.

Genome Res. 1996 6: 1084-1092
Access the most recent version at doi:10.1101/gr.6.11.1084

LETTER

Alu Fossil Relics—Distribution and Insertion Polymorphism

Santosh S. Arcot,[1] Aaron W. Adamson,[1] Jane E. Lamerdin,[1] Brian Kanagy,[2,3] Prescott L. Deininger,[3,4] Anthony V. Carrano,[1] and Mark A. Batzer[2,5]

[1]Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94551; [2]Department of Pathology and [3]Department of Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Louisiana State University Medical Center, New Orleans, Louisiana 70112; [4]Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, Louisiana 70121

Screening of a human genomic library with an oligonucleotide probe specific for one of the young subfamilies of Alu repeats (Ya5/8) resulted in the identification of several hundred positive clones. Thirty-three of these clones were analyzed in detail by DNA sequencing. Oligonucleotide primers complementary to the unique sequence regions flanking each Alu repeat were used in PCR-based assays to perform phylogenetic analyses, chromosomal localization, and insertion polymorphism analyses within different human population groups. All 33 Alu repeats were present only in humans and absent from orthologous positions in several nonhuman primate genomes. Seven Alu repeats were polymorphic for their presence/absence in three different human population groups, making them novel identical-by-descent markers for the analysis of human genetic diversity and evolution. Nucleotide sequence analysis of the polymorphic Alu repeats showed an extremely low nucleotide diversity compared with the subfamily consensus sequence with an average age of 1.63 million years old. The young Alu insertions do not appear to accumulate preferentially on any individual human chromosome.

[The sequence data described in this paper have been submitted to GenBank under accession nos. U57004 (HS3.23), U57005 (HS2.43), U57006 (HS4.32), U57007 (HS4.65), U57008 (HS4.69), U57009 (HS4.75), X55925 (HS4.59 originally termed C3N3), U67208 (HS2.7), U67209 (HS2.10), U67210 (HS2.16), U67212 (HS2.29), U67213 (HS2.30), U67214 (HS2.36), U67215 (HS2.37), U67216 (HS3.13), U67217 (HS3.38), U67218 (HS3.71), U67219 (HS3.79), U67220 (HS4.3), U67222 (HS4.21), U67223 (HS4.38), U67224 (HS4.41), U67225 (HS4.48), U67226 (HS4.50), U67227 (HS4.52), U67228 (HS4.61), U67229 (HS4.66), U67230 (HS4.73), U67231 (HS4.74), U67232 (HS4.88), U67233 (HS5.24), U67234 (HS5.29), and U67235 (HS5.39).]

Repetitive DNA sequences comprise a significant portion of the genomes of eukaryotes. These repetitive sequences can be classified broadly into two principal components, tandem repeats and interspersed repeats. Based upon copy number and size, the bulk of interspersed repetitive sequences in primate genomes is comprised of Alu repeats (for review, see Deininger 1989; Schmid and Maraia 1992; Deininger and Batzer 1993). Alu repeats belong to a class of sequences defined as short interspersed elements (SINEs). Alu SINEs exist in copy numbers of 500,000–1,000,000 per haploid genome and make up ~10% of the mass of the human genome. Alu repeats are derived ancestrally from the 7SL RNA gene and mobilize through an RNA polymerase III-derived transcript in a process termed retroposition. A typical Alu element is ~300 nucleotides in length, dimeric in structure, and composed of two nearly identical halves joined by a middle A-rich region along with a 3′ oligo(dA)-rich tail and short flanking direct repeats.

Alu elements are the most successful of the mobile genetic elements within primate genomes, having amplified to large copy numbers over the past 65 million years of primate evolution. The amplification of Alu elements has been

[5]Corresponding author.
E-MAIL mbatze@lsumc.edu; FAX 504-568-6037.

6:1084–1092 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

dominated by a small number of master genes as well as a few fortuitous source genes (Deininger et al. 1992), resulting in the appearance of distinct subfamilies as defined by tightly linked diagnostic or subfamily-specific mutations (Slagel et al. 1987; Britten et al. 1988; Jurka and Smith 1988; Jurka and Milosavljevic 1991; Shen et al. 1991). The youngest of these subfamilies, defined as Ya5/8 and Yb8 (Batzer et al. 1996b), appeared around the time humans diverged from other primates; therefore, members of these subfamilies are largely restricted in their distribution to humans (Matera et al. 1990a,b; Batzer et al. 1990, 1995, 1996a; Batzer and Deininger 1991; Jurka 1993). A number of polymorphic Alu insertions have been identified (Batzer et al. 1991, 1994, 1995, 1996a; Blonden et al. 1994; Hammer 1994; Milewicz et al. 1996) including some that are unique to single individuals (Wallace et al. 1991; Vidaud et al. 1993) or families (Muratani et al. 1991). Members of the evolutionarily young Alu subfamilies are novel reagents for gaining insight into the mode and tempo of Alu repeat amplification. In an attempt to identify additional members of one of the young Alu subfamilies, a randomly sheared total genomic human library was constructed in bacteriophage lambda and screened using an oligonucleotide probe containing diagnostic mutations specific for the Ya5/8 subfamily (Batzer and Deininger 1991). In this study, we report the identification, analysis, and chromosomal distribution of members of the Ya5/8 subfamily of Alu repeats.

RESULTS AND DISCUSSION

DNA Sequence Analysis

As a part of a directed strategy to identify additional members of the young subfamily of Alu repeats, a randomly sheared total genomic human library was screened with an oligonucleotide probe harboring the diagnostic mutations specific for the subfamily, which resulted in the identification of several hundred positive clones (Batzer et al. 1990). Clones were randomly selected and sequenced using internal Alu-specific primers to obtain the sequence of the regions flanking the Alu repeat (Batzer et al. 1990). Approximately 80% of the clones generated high-quality sequence. The clones that did not sequence well probably contained random mutations in the regions where the subfamily-specific sequencing primers anneal. The flanking sequences were verified by comparing the sequences of the direct repeats of each Alu element, because the majority of young Alu elements are flanked by perfect direct repeats as compared with older Alu elements in which the flanking direct repeats have been subjected to random mutations after the genomic integration of these elements (Batzer et al. 1990; Matera et al. 1990a; Arcot et al. 1995b).

Oligonucleotide primers complementary to the 5′ and 3′ unique sequence regions flanking each Alu repeat were designed and optimized for their specificity in PCR reactions using total human genomic DNA. The sequence of the primers and their optimal annealing temperatures are listed in Table 1. These primers were used in a series of PCR-based experiments to determine the phylogenetic distribution, chromosomal localization, and insertion polymorphism of each Alu repeat (outlined below). Initially each Alu element was tested for insertion polymorphism on a small sample of humans from three different population groups. Seven of the 33 Alu elements reported here were polymorphic for the presence or absence of the Alu repeat. Finished DNA sequences from all of the polymorphic Alu repeats were generated and compared with the Ya5 subfamily consensus sequence. An alignment of the sequences of these polymorphic Alu elements with the Ya5 subfamily consensus sequence is shown in Figure 1A.

Six of the polymorphic Alu elements had all five diagnostic mutations characteristic of the Ya5 subfamily (underlined in the consensus sequence; Fig. 1A). Alu element HS2.43 had an A instead of a C at position 238 in the consensus. Two of the elements shared 100% nucleotide identity with the subfamily consensus sequence (HS4.32 andHS4.59), two had single nucleotide substitutions (HS2.43 and HS4.69), and two elements (HS3.23 and HS4.65) contained single nucleotide insertions within the middle A-rich region of each element. HS3.23 had two substitutions compared with the consensus sequence at non-CpG nucleotides along with two changes at CpG dinucleotide positions. Mutations within Alu elements occur at a neutral rate, therefore the substitutions at non-CpG positions can be used as an indicator of the age of these elements (Labuda and Striker 1989; Batzer et al. 1990). The CpG positions are removed from the analysis because they mutate at a rate that is 9.2-fold higher than the non-CpG positions as a result of unidirectional decay from the spontaneous deamina-

Table 1. Oligonucleotides for PCR Amplification, Annealing Temperatures, Product Sizes, and Chromosomal Location of Young Alu Insertions

Alu insertion[a]	Subfamily[b]	5′ Primer sequence (5′–3′)[c]	3′ Primer sequence (5′–3′)	Annealing[d] temperature	PCR product[e] filled site	PCR product empty site	Chromosomal[f] location
HS2.43*	Ya5	ACTCCCCACCAGGTAATGGT	AGGGCCTTCATCCAGTTTGT	67°C	482	184	1
HS3.23*	Ya5	GGTGAAGTTTCCAACGCTGT	CCCTCCTCTCCCTTTAGCAG	60°C	498	200	7
HS4.32*	Ya5	GTTTATTGGGCTAACCTGGG	TGACCTGCTAACTTGTACTTTAACC	60°C	601	289	12
HS4.59*	Ya5	TTTCTGAACCTGAAGACAGG	TCTCAATTTCTTCTCATAATGCA	60°C	461	143	9
HS4.65*	Ya5	TGAAGCCAATGGAAAGAGAG	ACAGGAGCATCTAAACCTTGG	61°C	650	329	9
HS4.69*	Ya5	GTCTGAATGTTCGTGTCCCC	GTCCAAGTTCAAGGCACCAG	70°C	572	262	6
HS4.75*	Ya5	CAGCATTACATACAATAGTTAGGAGC	GTGATATTTGTCTTTCTGTACCTGG	58°C	520	194	3
HS2.7	Ya	AAATCCTAACAGTGTCCCACA	ACAGAACTTCAATAGCAGACTAGATC	60°C	481	155	8
HS2.10	Ya	CCTTCCTAAAAACCTTCAGTAGC	TATATGTGTATGTAAATGTGGTGGG	63°C	457	144	3
HS2.16	Ya	TTGACTACGCTGTCCTCACC	CATGGATAGGCTTCAGTGGC	62°C	523	210	5
HS2.29	Ya	TACAGGAAAGTGATGGGGAG	GCCAAGGGATCCACTCTTAG	62°C	551	266	11
HS2.30	Ya	CCCTCAAATGAAATTTTGCA	TCCTTGAAAAAGATGGGTGG	55°C	388	75	14
HS2.36	Ya	TCTGCCCCTTCACCAGTATC	ACCCAACTTAGGCAACCCTC	60°C	498	179	2
HS2.37	Ya	GGATTGCTGCTCCATCTCAG	AAGGCAAGTGATGCTTTTCC	60°C	491	210	6
HS3.13	Ya	AGGAAGAACAGAGGGTGGGA	TGCTGCCTCTCTCCCAGTAG	68°C	566	284	6
HS3.38	Ya	ATACTTTGCACTCATAGACCTAACTG	CAAACTTAACTGATCTGAATCAACA	55°C	533	256	8
HS3.71	Ya	GACATCTAAGAGCTAAAGACTTCATG	TTAGCAAATAGAAAAGTGTTATGTCC	60°C	553	240	13
HS3.79	Ya	GGATCAATAGGTGTCAGGTCC	AGATTCCTCTTATTTCTTCTTCAGC	60°C	555	236	10
HS4.3	Ya	TTCATTAAGTGGATGGGGAA	AATACATGCAAAGAGGCTGC	57°C	582	275	9
HS4.21	Ya	GTGTCTGAATGAAGAGGCCA	TTACAGAAGCCTCTCCACTCA	67°C	462	147	9
HS4.38	Ya	ATGAATTATAGCATGGGTCACC	TATCAACATGCAGAGAATGCA	60°C	639	331	12
HS4.41	Ya	GGACCATAGATGATATTC	TGAGTAATGTCAACTTCTAGC	56°C	610	290	8
HS4.48	Ya	TGTGTGGGCAAGAAAAGAGA	CTCAGAAATGAGCCTGGAGC	65°C	418	198	3
HS4.50	Ya	AGCTAAGCACCTCACCCCTG	ACCCCTCCTCAGTTTTTGGA	60°C	481	167	X
HS4.52	Ya	GATTCATACCTACTTCCTAATCAGC	GAATGTAGACCCAGACAGATTTC	60°C	651	335	4
HS4.61	Ya	CCTTTGTATGGAGCCCTTCC	TTCCCCTTAGCAAGTCCTCC	60°C	528	218	11
HS4.66	Ya	ATAAAGGGAGTGGTAGTGGTTG	AGCAGGTAGTAAGTAGATAATGAAAGC	59°C	579	255	5
HS4.73	Ya	CCTGTTTGGAGAATGGAGGA	TCTCTTGGATTGCCGCTAAG	60°C	415	104	16
HS4.74	Ya	CAACAGTAGTGGAGAACTTGGG	TTGTTGTTGCTGTTGTGAGC	61°C	574	248	5
HS4.88	Ya	CAGGTGGCAATCTTAACTTCC	AGGGTTGGAAGTGAAAGTGG	68°C	531	185	5
HS5.24	Ya	TTCAGCACTACCAGGATGGA	CTTGCAGAGATTGGTGAGGA	57°C	469	119	1
HS5.29	Ya	AAAGATCCTTATACAAACATC	ATGTTTATGTGTTTTGAAAGG	56°C	614	310	6
HS5.39	Ya	GGCAGTGCTGGTGAAAATAG	TCAACTTCTCTCATCCAGTT	58°C	397	84	17

[a]Polymorphic Alu elements are denoted by an asterisk.
[b]Subfamily nomenclature described in Batzer et al. (1996b).
[c]Oligonucleotides were designed using the PRIMER software (Whitehead Institute for Biomedical Research, Cambridge, MA).
[d]PCR conditions described in Methods.
[e]Filled sites denote PCR products that contain an Alu element and empty sites represent PCR products with the preintegration site only. Fragment sizes are listed in bp.
[f]Chromosomal locations were determined by PCR amplifications of NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2 (Coriell Institute for Medical Research, Camden, NJ) as described in Methods.

tion of 5-methyl cytosine (Labuda and Striker 1989; Batzer et al. 1990). A total of 4 out of 1638 non-CpG nucleotides mutated within the seven dimorphic Alu repeats (excluding the two insertions) or an average of ~0.5714 ± 1.134 substitutions in each Alu element compared with the subfamily consensus sequence. Using a neutral rate of evolution of 0.15% per million years (Miyamoto et al. 1987), the average age of the seven polymorphic Alu elements is ~1.63 ± 3.23 million years. This estimate compares favorably to the previously reported average age of the Ya5 subfamily of 2.8 million years (Batzer et al. 1990) because the present age estimate only involves polymorphic elements that are presumably the youngest Ya5 subfamily members. Only two of the polymorphic Alu elements (HS4.65 and HS4.75) contained 3′ oligo(dA) tails comprised of nucleotides other than A's (a single G in both cases) (Fig. 1B). Homopolymeric A-rich tails are characteristic features of young Alu elements (Batzer et al. 1990). Mutations within the 3′ oligo(dA)-rich tail have been shown previously to serve as nuclei for the generation of Alu-associated microsatellite repeats (Arcot et al. 1995c). All seven of the elements were flanked by perfect direct repeats varying length from 14–17 nucleotides (Fig. 1B), yet another hallmark of evolutionarily young Alu repeats (Batzer et al. 1990, 1991, 1995; Arcot et al. 1995a,b).

Phylogenetic Analyses

Phylogenetic analyses were performed using DNA samples from humans as well as 15 nonhuman primates as templates for the PCR. The Ya5 subfamily of Alu repeats began to amplify around the time humans diverged from nonhuman primates (~5 million years ago) (Batzer et al. 1990; Batzer and Deininger 1991); therefore, members of this subfamily are restricted primarily in their distribution to humans. However, a few Ya lineage Alu repeats have been detected in nonhuman primates (Leeflang et al. 1992, 1993). All 33 Alu repeats reported here were found exclusively in humans, and were absent from the genomes of 15 nonhuman primates (data not shown). These data are consistent with our previous observations that Ya5 subfamily members are predominantly human-specific (Batzer and Deininger 1991; Batzer et al. 1991, 1995, 1996a) and suggest that at most 4–5% (three out of 70 reported to date) of these subfamily members reside in nonhuman primate genomes.

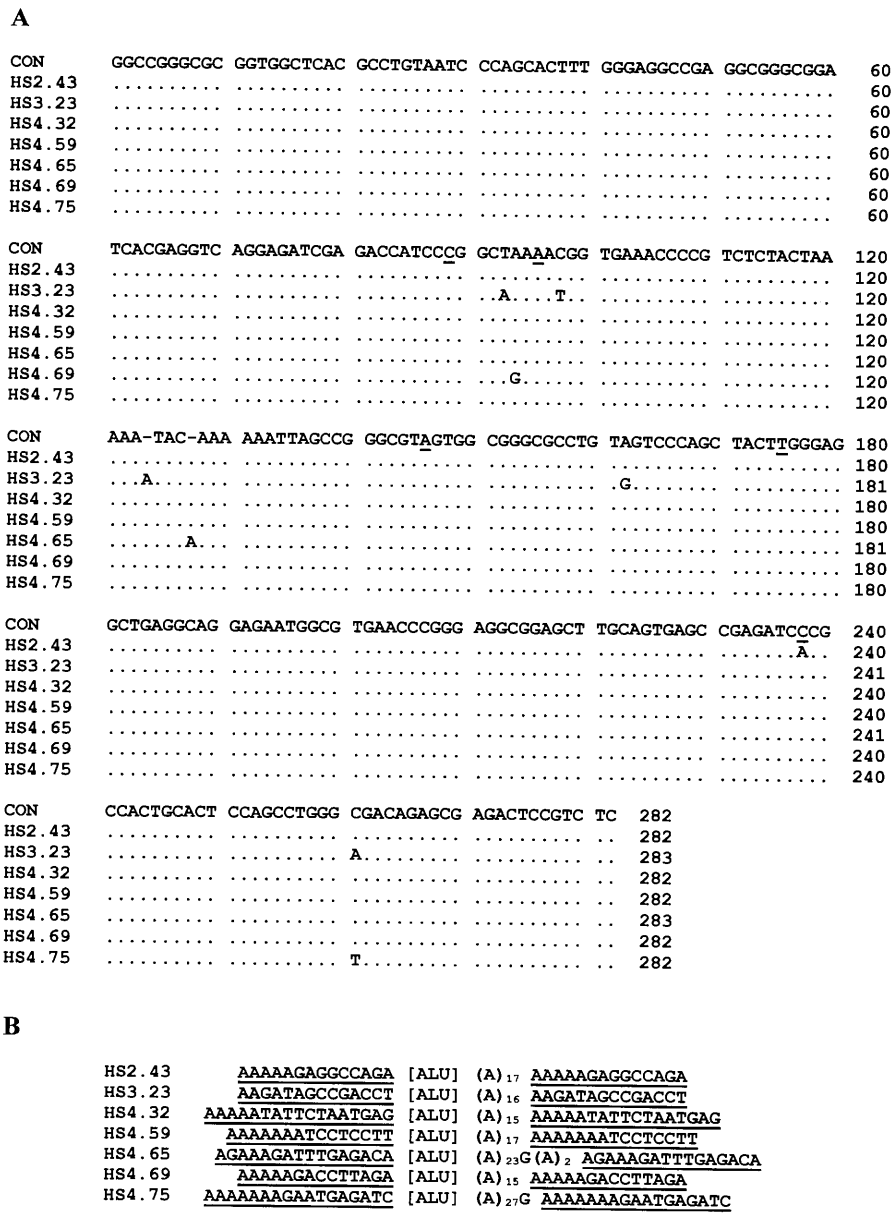

Figure 1 Sequences of the polymorphic Alu elements and their flanking direct repeats. (*A*) Alignment of the sequences of the seven polymorphic Alu elements with the consensus sequence (CON) for the Ya5 subfamily. The five diagnostic nucleotides characteristic of the Ya5 subfamily are underlined in the consensus sequence. Nucleotide identities are indicated by dots, whereas substitutions are indicated by appropriate nucleotides. (*B*) Flanking nucleotide sequences. The composition of the 3′ oligo(dA) tail is denoted by the appropriate nucleotide in parenthesis followed by the length in subscripts. The nucleotide sequences comprising the flanking direct repeats of each Alu element are underlined.

Chromosomal Distribution

Based on the number of Alu elements present within the genome, it is estimated that there is one Alu element every 4–6 kb (Deininger 1989). However, many Alu elements cluster together (Slagel et al. 1987; Moyzis et al. 1989), with some global preference for gene-rich regions of the genome (Korenberg and Rykowski 1988). This clustering has been attributed to the integration site preference of Alu repeats for A + T-rich regions of the genome. Because each Alu repeat contains a middle as well as a 3′ oligo(dA)-rich region, they serve as targets for the subsequent insertion of new Alu repeats (for re-

view, see Deininger 1989; Deininger and Batzer 1993). The clustering Alu repeats occurs over a period of millions of years (Sawada and Schmid 1986; Slagel et al. 1987); however, new Alu insertions have no known chromosomal integration site preferences.

Previously, we performed a statistical analysis on the distribution of 34 young Alu repeats within the human genome (Arcot et al. 1995b). Using the NIGMS somatic cell hybrid mapping panels 1 and 2, we mapped the chromosomal location of 34 additional young Alu repeats (Table 1; Arcot et al. 1995b). With a total of 68 young Alu elements comprising Ya5/8 and Yb8 subfamilies, we reevaluated the distribution pattern of the Alu repeats using Pearson's chi-square test, a goodness-of-fit statistic. The results of the analysis show the contribution of each chromosome to the overall chi-square score (Table 2). Using the Pearson's chi-square test, the distribution of young Alu repeats within the various chromosomes cannot be distinguished from what would occur randomly (chi-square = 27.35, $0.75 < P < 0.9$, df = 23). Therefore, the young Alu repeats do not appear to accumulate preferentially on any particular human chromosome.

Table 2. Statistical Analysis of the Chromosomal Distribution of Young Alu Insertions

Chromosome	Observed[a]	Expected[b]	χ^2
1	5	5.4	0.03
2	1	5.4	3.58
3	5	4.4	0.08
4	1	4.2	2.43
5	6	4.0	1.00
6	6	3.8	1.27
7	2	3.6	0.71
8	5	3.2	1.01
9	7	3.0	5.33
10	1	3.0	1.33
11	6	3.0	3.00
12	5	3.0	1.33
13	1	2.4	0.82
14	3	2.2	0.29
15	0	2.2	2.20
16	3	1.9	0.64
17	2	1.8	0.02
18	0	1.8	1.80
19	2	1.4	0.26
20	1	1.4	0.11
21	1	1.0	0.00
22	1	1.2	0.03
X	3	3.4	0.05
Y	1	1.2	0.03
Total	68	68.0	27.35

[a]Includes all young Alu repeats described in Table 1 and in Arcot et al. (1995a,b).
[b]Expected number of Alu repeats proportional to the length of the chromosome as described in Mayall et al. (1984).

Distribution of the Polymorphic Alu Repeats

Polymorphic Alu elements are novel nuclear markers that can be used in studies pertaining to relationships between different human population groups and human genetic diversity (Perna et al. 1992; Batzer et al. 1994, 1996a; Hammer 1994; Tishkoff et al. 1996b). The insertion of an Alu element within the human genome is a unique event that occurs in a single individual, and therefore each polymorphic Alu element is identical by descent (Batzer et al. 1994). Other commonly used polymorphic markers such as restriction fragment length polymorphisms (RFLP) and variable number of tandem repeats (VNTR) are merely identical by state, making the Alu insertion polymorphisms unique nuclear markers.

Previous studies have utilized polymorphic Alu elements to study human genetic diversity and evolution (Perna et al. 1992; Batzer et al. 1994, 1996a; Hammer 1994; Tishkoff et al. 1996b). To ascertain the utility of these polymorphic Alu elements as tools for population studies, we determined the distribution of the seven polymorphic elements in individuals from three different population groups (U.S. Caucasians, African-Americans, and Hispanic-Americans) using PCR-based assays for the presence or absence of individual Alu repeats (Table 3). These three population groups represent a sampling of the genetic diversity present in Africa, Europe, and the Americas. Each locus was in Hardy-Weinberg equilibrium for all populations, as judged by independent chi-square tests for goodness of fit. All seven of the Alu elements were polymorphic within the three population groups except for HS4.75 that was fixed in Hispanic-Americans. These data suggest that these seven polymorphic Alu elements arose as genomic fossils prior to the radiation of modern humans from Africa, which is thought to have occurred within the last 15,000–200,000 years (Cann et al. 1987; Armour et al. 1996; Knight et al. 1996; Tishkoff et al.

Table 3. Distribution of Polymorphic Alu Elements

			U.S. Caucasians			African-Americans			Hispanic-Americans		
Alu insertion	Genotype[a]		observed individuals	expected[b] individuals	allele frequencies	observed individuals	expected individuals	allele frequencies	observed individuals	expected individuals	allele frequencies
HS2.43	+	+	1	0.58	0.12	0	0.03	0.03	0	0.10	0.05
	+	–	8	8.84		2	1.94		4	3.80	
	–	–	34	33.58	0.88	29	29.03	0.97	37	37.10	0.95
HS3.23	+	+	33	32.70	0.87	46	46.00	0.99	22	22.81	0.74
	+	–	9	9.60		1	0.99		16	16.37	
	–	–	1	0.70	0.13	0	0.01	0.01	3	2.81	0.26
HS4.32	+	+	18	15.39	0.63	10	11.51	0.51	8	7.07	0.46
	+	–	13	18.22		25	21.99		15	16.87	
	–	–	8	5.39	0.37	9	10.51	0.49	11	10.07	0.54
HS4.59	+	+	21	19.51	0.74	15	13.11	0.67	21	20.02	0.69
	+	–	11	13.97		9	12.78		16	17.95	
	–	–	4	2.51	0.26	5	3.11	0.33	5	4.03	0.31
HS4.65	+	+	0	0.01	0.01	1	1.22	0.16	1	0.34	0.09
	+	–	1	0.99		13	12.55		6	7.32	
	–	–	47	47.00	0.99	32	32.22	0.84	40	39.34	0.91
HS4.69	+	+	5	3.35	0.28	5	4.05	0.30	0	1.68	0.20
	+	–	14	17.30		17	18.9		17	13.64	
	–	–	24	22.35	0.72	23	22.05	0.70	26	27.68	0.80
HS4.75	+	+	15	17.78	0.66	33	31.54	0.82	46	46.00	1.00
	+	–	24	18.44		11	13.93		0	0.00	
	–	–	2	4.78	0.34	3	1.54	0.18	0	0.00	0.00

[a]The presence and absence of the Alu repeat are denoted by + and –, respectively.
[b]Expected values are based upon Hardy–Weinberg equilibrium.

1996a). The low frequency of Alu element HS2.43 in African-Americans (3.0%) and Hispanic-Americans (5.0%) as well as HS4.65 in U.S. Caucasians (1.0%) and Hispanic-Americans (9.0%) suggests that these two Alu elements are of more recent origin than the other five polymorphic elements. Alu elements HS2.43 and HS4.65 should be more informative for resolving recent human migrations, whereas the other five polymorphic Alu elements will provide insight into the origin of modern humans.

Based on the number of Ya5/8 and Yb8 Alu repeats that have been characterized in detail, it appears that ~24% of the Alu elements from these subfamilies are polymorphic for insertion presence or absence within diverse human population groups (16 polymorphic elements out of 68 characterized; Arcot et al. 1995a,b and the present study). Approximately 1000–4000 Ya5/8 and Yb8 Alu repeats exist within the human genome (Batzer et al. 1995), therefore 240–960 of these Alu elements should be polymorphic for insertion presence or absence. The polymorphic Alu elements provide unique insight into the genetic structure and evolutionary history of diverse human population groups.

SUMMARY

There are an estimated 1000–4000 Alu elements that belong to the Ya5/8 and Yb8 Alu subfamilies. Whereas a few of these elements have been identified randomly, directed approaches involving screening genomic libraries for members of specific Alu subfamilies have facilitated the identification and analysis of a number of these elements. Evolutionarily young Alu elements provide valuable information about the biological properties of this class of interspersed repetitive DNA sequences, particularly with regard to their amplification dynamics and impact on the evolutionary architecture of the human genome. Alu (Ya5/8 and Yb8) subfamily members are largely restricted in their distribution to humans and absent from nonhuman primates, making them unique sequence tagged sites (STSs) for physical mapping and anchored reference loci for comparative genome mapping studies in primates. Polymorphic Alu elements represent novel identical by descent nuclear genomic fossils for the study of human genetic diversity as well as diagnostic tools for forensics and human identification.

METHODS

Cell Lines and DNA Samples

Cell lines used to isolate DNA samples were as follows: human (*Homo sapiens*), HeLa (ATCC CCL2); chimpanzee (*Pan troglodytes*), Wes (ATCC CRL1609); gorilla (*Gorilla gorilla*), Ggo-1 (primary gorilla fibroblasts) provided by Dr. Stephen J. O'Brien (National Cancer Institute, Frederick, MD); African green monkey (*Cercopithecus aethiops*), CV1 (ATCC CCL70); and owl monkey (*Aotus trivirgatus*), OMK (637-69 ATCC CRL1556). Cell lines were maintained as directed by the source and DNA isolations were performed as described previously (Ausabel et al. 1987). Additional DNA samples from five individual chimpanzees (*Pan troglodytes*), one gorilla (*Gorilla gorilla*), three orangutans (*Pongo pygmaeus*), one macaque (*Macaca fascicularis*), and one tamarin (*Saguinus oedipus*) were also obtained from BIOS laboratories (New Haven, CT). Human DNA samples from U.S. Caucasian (northern-European ancestry), African-American, and Hispanic-American groups were isolated from peripheral lymphocytes (Ausabel et al. 1987) available from previous studies (Batzer et al. 1996a).

Library Construction and Screening

Construction of a randomly sheared total genomic library in bacteriophage lambda ZAP II (Stratagene, La Jolla, CA), screening with an oligonucleotide probe specific for the Ya5 subfamily of Alu elements, plaque purification, and plasmid rescue have been described previously (Batzer and Deininger 1991).

DNA Sequence Analysis

Plasmid templates for sequencing were prepared by the Qiagen mini-alkaline lysis method according to the manufacturer's instructions (Qiagen, Chatsworth, CA). Double-stranded plasmid templates corresponding to the positive clones from the HeLa genomic library were sequenced with internal Alu-specific primers (Batzer et al. 1990) and fluorescently labeled dye terminators using either AB *Taq* cycle sequencing (Applied Biosystems Division, Foster City, CA) or Sequenase kits. Sequencing of the PCR products corresponding to the dimorphic Alu elements cloned into *p*CR II TA-cloning vector (Invitrogen, San Diego, CA) was accomplished using fluorescently labeled M13 forward and reverse primers and AB *Taq* cycle sequencing kits. Sequencing reactions were fractionated on a 6% polyacrylamide gel, followed by data collection and analysis on an ABI 373A DNA sequencer.

Primer Design, PCR Amplification, and Chromosomal Localization

Sequences flanking individual Alu elements were screened for the presence of human repetitive sequences by comparison with a database of Alu, L1, THE, and MER elements using fastdb (Intelligenetics, Mountain View, CA). PCR primers were designed using the software PRIMER (Whitehead Institute for Biomedical Research, Cambridge, MA). PCR amplification was carried out in 50 µl reactions using 50–100 ng of target DNA, 40 pM of each oligonucleotide primer, 200 µM dNTPs in 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.4), and AmpliTaq DNA polymerase (2.5 u) as recommended by the supplier (Roche Molecular Systems, Branchburg, NJ). Each sample was subjected to the following amplification cycle: 1 min of denaturation at 94°C, 1 min at the annealing temperature, and 1 min of extension at 72°C, and repeated for 32 cycles. Twenty microliters of each sample was fractionated on a 2% agarose gel with 0.25 µg/ml ethidium bromide. PCR products were directly visualized using UV fluorescence. The sequences of the oligonucleotide primers and their annealing temperatures are shown in Table 1. Phylogenetic analysis of all the Alu elements listed in Table 1 was determined by PCR amplification of human and 15 nonhuman primate DNA samples. The chromosomal location of each Alu repeat was determined by PCR amplification of NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2 (Coriell Institute for Medical Research, Camden, NJ).

Data Analysis

Analysis of the distribution of the young Alu elements throughout the genome was assessed by Pearson's chi-square statistic [chi-square = $\Sigma i\ (O_i - Np_i)^2 / Np_i$], where O_i is the observed number of Alu elements on chromosome *i*, p_i is the probability of finding an Alu element on chromosome *i*, and N = 68 (Table 1; Arcot et al. 1995a,b) is the number of Alu elements in the data set. The probability of assignment to a chromosome was proportional to the length of that chromosome, as described in Mayall et al. (1984).

ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Institutes of Health (HG 00770) to P.L.D. Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy under contract no. W-7405-ENG-48.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Arcot, S.S., J.J. Fontius, P.L. Deininger, and M.A. Batzer. 1995a. Identification and analysis of a "young" polymorphic Alu element. *Biochem. Biophys. Acta* **1263:** 99–102.

Arcot, S.S., T.H. Shaikh, J. Kim, L. Bennett, M. Alegria-Hartman, D.O. Nelson, P.L. Deininger, and M.A. Batzer. 1995b. Sequence diversity and chromosomal distribution of "young" Alu repeats. *Gene* **163:** 273–278.

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

Armour, J.A.L., T. Anttinen, C.A. May, E.E. Vega, A.

Sajantila, J.R. Kidd, K.K. Kidd, J. Bertranpetit, S. Paabo, and A.J. Jeffreys. 1996. Minisatellite diversity supports a recent African origin for modern humans. *Nature Genet.* **13:** 154–160.

Ausabel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl. 1987. *Current protocols in molecular biology*. John Wiley and Sons, New York, NY.

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

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., V.A. Gudi, J.C. Mena, D.W. Foltz, R.J. Herrera, and P.L. Deininger. 1991. Amplification dynamics of human-specific Alu family members. *Nucleic Acids Res.* **19:** 3619–3623.

Batzer, M.A., M. Stoneking, M. Alegria-Hartman, H. Bazan, D.H. Kass, T.H. Shaikh, G.E. Novick, P.A. Ioannou, W.D. Scheer, R.J. Herrera, and P.L. Deininger. 1994. African origin of human-specific polymorphic Alu insertions. *Proc. Natl. Acad. Sci.* **91:** 12288–12292.

Batzer, M.A., C.M. Rubin, U. Hellman-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., S.S. Arcot, J.W. Phinney, M. Alegria-Hartman, D.H. Kass, S.M. Milligan, C. Kimpton, P. Gill, M. Hochmeister, P.A. Ioannou, R.J. Herrera, D.A. Boudreau, W.D. Scheer, B.J.B. Keats, P.L. Deininger, and M. Stoneking. 1996a. Genetic variation of recent Alu insertions in human populations. *J. Mol. Evol.* **42:** 22–29.

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

Blonden, L.A.J., G.M. Terwindt, J.T. Den Dunnen, and G-J.B. Van Ommen. 1994. A polymorphic STS in intron 44 of the dystrophin gene. *Hum. Genet.* **93:** 479–480.

Britten, R.J., W.F. Baron, D.B. Stout, and E.H. Davidson. 1988. Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci.* **88:** 839–843.

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* **325:** 31–36.

Deininger, P.L. 1989. SINEs: Short interspersed repeated DNA elements in higher eukaryotes. In *Mobile DNA* (ed. D.E. Berg and M.M. Howe), pp. 619–636. American Society for Microbiology, Washington, DC.

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

———. 1995. SINE master genes and population biology. In *The impact of short interspersed elements (SINEs) on the host genome* (ed. R.J. Maraia), pp. 43–60. R.G. Landes, Georgetown, TX.

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

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.

Jurka, J. 1993. A new subfamily of recently retroposed Alu repeats. *Nucleic Acids Res.* **21:** 2252.

Jurka, J. and A. Milosavljevic. 1991. Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* **32:** 105–121.

Jurka, J. and T. Smith. 1988. A fundamental division in the Alu family of repeated sequences. *Proc. Natl. Acad. Sci.* **85:** 4775–4778.

Knight, A., M.A. Batzer, M. Stoneking, H.K. Tiwari, W.D. Scheer, R.J. Herrera, and P.L. Deininger. 1996. DNA sequences of Alu elements support a recent replacement hypothesis for the human nuclear genetic component. *Proc. Natl. Acad. Sci.* **93:** 4360–4364.

Korenberg, J.R. and M.C. Rykowski. 1988. Human genome organization: Alu, lines and the molecular structure of metaphase chromosome bands. *Cell* **53:** 391–400.

Labuda, D. and G. Striker. 1989. Sequence conservation in Alu evolution. *Nucleic Acids Res.* **17:** 2477–2491.

Leeflang, E.P., W-M. Liu, C. Hashimoto, P.V. Choudary, and C.W. Schmid. 1992. Phylogenetic evidence for multiple Alu source genes. *J. Mol. Evol.* **35:** 7–16.

Leeflang, E.P., I. Chesnokov, and C.W. Schmid. 1993. Mobility of short interspersed repeats within the chimpanzee lineage. *J. Mol. Evol.* **37:** 566–572.

Matera, A.G., U. Hellmann, M.F. Hintz, and C.W. Schmid. 1990a. Recently transposed Alu repeats result from multiple source genes. *Nucleic Acids Res.* **18:** 6019–6023.

Matera, A.G., U. Hellmann, and C.W. Schmid. 1990b. A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.* **10:** 5424–5432.

Mayall, B.H., A.V. Carrano, D.H. Moore II, L.K. Ashworth, D.E. Bennett, and M.L. Mendelsohn. 1984. The DNA-based human karyotype. *Cytometry* **5:** 376–385.

Milewicz, D.M., P.H. Byers, J. Reveille, A.L. Hughes, and M. Duvic. 1996. A dimorphic Alu Sb-like insertion is ethnic-specific. *J. Mol. Evol.* **42:** 117–123.

Miyamoto, M.M., J.L. Slightom, and M. Goodman. 1987. Phylogenetic relations of humans and African apes from DNA sequences in the pseudo-ηt-globin region. *Science* **238:** 369–373.

Moyzis, R.K., D.C. Torney, J. Meyne, J.M. Buckingham, J.-R. Wu, C. Burks, K.M. Sirotkin, and W.B. Goad. 1989. The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4:** 273–289.

Muratani, K., T. Hada, Y. Yamamoto, T. Kaneko, Y. Shigeto, T. Ohue, J. Furuyama, and K. Higashino. 1991. Inactivation of the cholinesterase gene by Alu insertion: Possible mechanism for human gene transposition. *Proc. Natl. Acad. Sci.* **88:** 11315–11319.

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.

Sawada, I. and C.W. Schmid. 1986. Primate evolution of the α-globin gene cluster and its Alu-like repeats. *J. Mol. Biol.* **192:** 693–709.

Schmid, C.W. and R. Maraia. 1992. Transcriptional and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2:** 874–882.

Shen, M.R., M.A. Batzer, and P.L. Deininger. 1991. Evolution of master Alu gene(s). *J. Mol. Evol.* **33:** 311–320.

Slagel, V., E. Flemington, V. Traina-Dorge, H. Bradshaw, and P. Deininger. 1987. Clustering and subfamily relationship of the Alu family in the human genome. *Mol. Biol. Evol.* **4:** 19–29.

Tishkoff, S.A., E. Dietzsch, W. Speed, A.J. Pakstis, J.R. Kidd, K. Cheung, B. Bonne-Tamir, A.S. Santachiara-Benerecetti, P. Moral, M. Krings, S. Paabo, E. Watson, N. Risch, T. Jenkins, and K.K. Kidd. 1996a. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271:** 1380–1397.

Tishkoff, S.A., G. Ruano, J.R. Kidd, and K.K. Kidd. 1996b. Distribution and frequency of a polymorphic Alu insertion at the plasminogen activator locus in humans. *Hum. Genet.* **97:** 759–764.

Vidaud, D., M. Vidaud, B.R. Bahnak, V. Siguret, S.G. Sanchez, Y. Laurian, D. Meyer, M. Goossens, and J.M. Lavergne. 1993. Hemophilia B due to a *de novo* insertion of human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* **1:** 30–36.

Wallace, M.R., L.B. Andersen, A.M. Saulino, P.E. Gregory, T.W. Glover, and F.S. Collins. 1991. A *de novo* Alu insertion results in neurofibromatosis type I. *Nature* **353:** 864–866.

Received June 25, 1996; accepted in revised form August 30, 1996.