Genetica **107:** 149–161, 1999.

Recently integrated human Alu repeats: finding needles in the haystack

Astrid M. Roy[1], Marion L. Carroll[2], David H. Kass[3], Son V. Nguyen[2], Abdel-Halim Salem[2,*], Mark A. Batzer[2] & Prescott L. Deininger[1,4,**]

[1]*Tulane Cancer Center, SL-66, Department of Environmental Health Sciences, Tulane University - Medical Center, 1430 Tulane Ave., SL-66, New Orleans, LA 70112, USA;* [2]*Departments of Pathology, Biometry and Genetics, Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Neuroscience Center of Excellence, Louisiana State University Health Sciences Center, 1901 Perdido Street, New Orleans, LA 70112, USA;* [3]*Department of Biology, 316 Mark Jefferson, Eastern Michigan University, Ypsilanti, MI 48197, USA;* [*]*Present address: Department of Anatomy, Faculty of Medicine, Suez Canal University, Ismailia, Egypt;* [4]*Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, 1516 Jefferson Highway, New Orleans, 70121;* [**]*Author and address for correspondence: Tulane Cancer Center, Tulane University – Medical Center, 1430 Tulane Ave., SL-66, New Orleans, LA 70112, USA (Phone: (504) 988-6385; Fax: (504) 588-5516; E-mail: pdeinin@tcs.tulane.edu)*

Accepted 25 January 2000

Key words: Alu insertion polymorphisms, anchored-PCR, comparative genomics, computational biology

Abstract

Alu elements undergo amplification through retroposition and integration into new locations throughout primate genomes. Over 500,000 Alu elements reside in the human genome, making the identification of newly inserted Alu repeats the genomic equivalent of finding needles in the haystack. Here, we present two complementary methods for rapid detection of newly integrated Alu elements. In the first approach we employ computational biology to mine the human genomic DNA sequence databases in order to identify recently integrated Alu elements. The second method is based on an anchor-PCR technique which we term Allele-Specific Alu PCR (ASAP). In this approach, Alu elements are selectively amplified from anchored DNA generating a display or 'fingerprint' of recently integrated Alu elements. Alu insertion polymorphisms are then detected by comparison of the DNA fingerprints generated from different samples. Here, we explore the utility of these methods by applying them to the identification of members of the smallest previously identified subfamily of Alu repeats in the human genome termed Ya8. This subfamily of Alu repeats is composed of about 50 elements within the human genome. Approximately 50% of the Ya8 Alu family members have inserted in the human genome so recently that they are polymorphic, making them useful markers for the study of human evolution.

Introduction

Alu repeats are the most successful class of mobile elements in the human genome. Alu elements spread through the genome via an RNA mediated amplification mechanism termed retroposition and reviewed in Deininger and Batzer, 1993. There are over 500,000 Alu elements in the human genome, which have clearly played a major role in sculpting and/or damaging the genome. Alu elements have contributed to genetic disease, both by the disruption of genes through the insertion of newly retroposed elements and by recombination between Alu elements (reviewed in Deininger & Batzer, 1999). Previous estimates indicate that retroposition of Alu elements contributes to approximately 0.1% of human genetic diseases and recombination between Alu repeats contributes to another 0.3% of genetic diseases (Deininger & Batzer, 1999). Therefore, the spread of the Alu family of mobile elements has generated a significant amount of human genomic variation as well as diseases through recombination-based fluidity as well as insertional mutagenesis.

Alu repeats are distributed rather haphazardly throughout the human genome. Alu elements began expanding in the ancestral primate genomes about 65 mya (Shen, Batzer & Deninger, 1991) reaching a peak amplification between 35 and 60 mya. Presently, Alu elements amplify at a rate that is 100 fold lower than their peak rate, with an estimate of one new Alu insert in every 100–200 births (Deininger & Batzer, 1993, 1995). Evolutionary studies have demonstrated that the majority of evolutionarily recent Alu inserts have specific diagnostic sequence mutations (Deininger & Batzer, 1993, 1995). These mutations have accumulated in Alu elements throughout primate evolution resulting in a hierarchical subfamily structure, or lineage, of Alu repeats. The mutations facilitate the classification of Alu elements into different subfamilies, or clades, of related elements that share common diagnostic mutations (reviewed in Batzer, Schmid & Deninger, 1993; Batzer & Deininger, 1991; Batzer et al., 1996a). Almost all of the recently integrated Alu elements within the human genome belong to one of four closely related subfamilies: Y, Ya5, Ya8, and Yb8, with the majority being Ya5 and Yb8 subfamily members. Collectively, these subfamilies of Alu elements comprise less than 10% of the Alu elements present within the human genome with the Ya5/8 and Yb8 subfamilies collectively accounting for less than half of a percent of all Alu elements. These evolutionarily recent Alu insertions are useful for human population studies, since there appears to be no specific mechanism to remove newly inserted Alu repeats, and the Alu elements are identical by descent with a known ancestral state (Batzer et al., 1991, 1994a, 1996a; Stoneking et al., 1997; Perna et al., 1992).

Previously, it has been technically impossible to determine the full impact of mobile elements on the human genome. The identification of newly inserted Alu elements has been very difficult due to the complexity of detecting one new Alu insertion in a cell that already has 500,000 pre-existing Alu elements. We have previously utilized laborious library screening and sequencing strategies to isolate relatively small numbers of Alu insertion polymorphisms (Arcot et al., 1995a, b, c; Batzer & Deininger 1991a; Batzer et al., 1990, 1991b; 1995), as well as investigating rare 300 bp restriction fragment length polymorphisms (Kass et al., 1994). This makes these studies the genomic equivalent of the search for needles in the haystack. In this paper, we discuss two alternative methods that overcome the inherent difficulties in these experiments, making these studies manageable. First, the availability of large quantities of human genomic DNA sequence provided by the Human Genome Project facilitates genomic database mining for recently integrated Alu elements. This approach should prove useful in determining the chromosome-specific and genome wide dispersal patterns of mobile elements, as well as for the identification of polymorphic mobile element fossils to apply to the study of human population genetics and primate comparative genomics. Secondly, we have developed a PCR-based method that we term Allele-Specific Alu PCR (ASAP). This technique allows us to take advantage of the subfamily-specific diagnostic mutations within Alu mobile elements to isolate and display recently integrated Alu repeats from different DNA samples, allowing for direct comparisons of the Alu content of different genomes or different cells from an individual.

Materials and methods

Cell lines and DNA samples

The cell lines used to isolate human 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, USA. Cell lines were maintained as directed by the source and DNA isolations were performed using Wizard genomic DNA purification (Promega). Human DNA samples from the European, African American and Greenland native population groups were isolated from peripheral blood lymphocytes (Ausubel et al., 1996) that were available from previous studies (Stoneking et al., 1997). Egyptian samples were collected from throughout the Nile river valley region and DNA from peripheral lymphocytes was prepared using Wizard genomic DNA purification kits (Promega). Human DNA used for ASAP was isolated from peripheral lymphocytes utilizing the super-quick gene method (Analytical Genetic Testing Center).

Computational analyses

A schematic overview summarizing the computational analyses of recently integrated Alu elements is shown in Figure 1. Initial screening of the GenBank non-redundant and high throughput genomic sequence (HTGS) databases was performed using the basic local

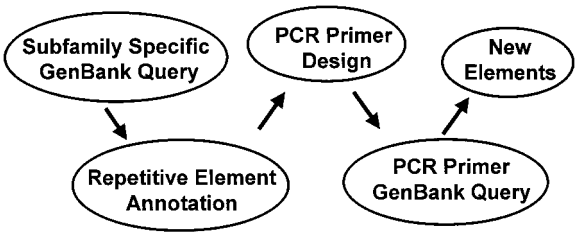

Figure 1. Computational analysis of repetitive elements. The flow chart shows the computational tools utilized for the identification and analysis of recently integrated Ya8 Alu family members. The process begins with BLAST searches of the non-redundant and high-throughput genomic sequence databases. Subsequently sequences (about 1000 nucleotides) adjacent to the matches with 100% identity to the query sequence are annotated using the RepeatMasker2 or Censor server. Following sequence annotation, oligonucleotide primers complementary to the unique DNA sequences adjacent to each element are designed using the Primer3 web server. The oligonucleotides designed using Primer3 are then subjected to a second BLAST search to determine if they reside in other repetitive elements, and subsequently they are used for PCR based analyses of individual mobile elements.

alignment search tool (BLAST) (Altschul et al., 1990) available from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). The database was searched for exact complements to the oligonucleotide 5′-ACTAAAACTACAAAAAATAG-3′ that is an exact match to a portion of the Alu Ya8 subfamily consensus sequence containing unique diagnostic mutations. Sequences that were exact complements to the oligonucleotide were then subjected to more detailed annotation. A region composed of 1000 bases of flanking DNA sequence directly adjacent to the sequences identified from the databases that matched the initial GenBank BLAST query were subjected to annotation using either RepeatMasker2 from the University of Washington Genome Center server (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) or Censor from the Genetic Information Research Institute (http://www.girinst.org/Censor_Server-Data_Entry_Form_s.html) (Jurka et al., 1996). These programs annotate the repeat sequence content of DNA sequences from humans and rodents.

Primer design and PCR amplification

PCR primers were designed from flanking unique DNA sequences adjacent to individual Ya8 Alu elements using the Primer3 software (Whitehead Institute for Biomedical Research, Cambridge, MA, USA) (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). The resultant PCR primers were screened against the GenBank non-redundant 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. PCR amplification was carried out in 25 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 Taq® DNA polymerase (1.25 U) as recommended by the supplier (Life Technologies). Each sample was subjected to the following amplification cycle: an initial denaturation of 2:30 min at 94°C, 1 min of denaturation at 94°C, 1 min at the annealing temperature, 1 min of extension at 72°C, repeated for 32 cycles, followed by a final extension at 72°C for 10 min. 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, annealing temperatures, PCR product sizes and chromosomal locations are shown in Table 1. Phylogenetic analysis of all the Alu elements listed in Table 1 was determined by PCR amplification of human and non-human primate DNA samples. The human genomic diversity associated with each element was determined by the amplification of 20 individuals from each of four populations (African–American, Greenland Native, European and Egyptian) (160 total chromosomes). The chromosomal location of Alu repeats identified from clones that had not been previously mapped was determined by PCR amplification of National Institute of General Medical Sciences (NIGMS) human/rodent somatic cell hybrid mapping panel 2 (Coriell Institute for Medical Research, Camden, NJ).

Allele-Specific Alu PCR (ASAP)

We used a modification of the IRE-Bubble PCR method (Munroe et al., 1994), utilizing the same amplification (anchor) primer, but altering the annealed anchor/linker primers. The annealed linkers formed a Y instead of a bubble to avoid end-to-end ligation. Also, instead of blunt-end digestion, genomic DNA was digested with *Mse*I; that cuts 5′-T′TAA-3′ and does not cut in the Alu consensus. Otherwise the genomic-anchor ligations were prepared according to (Munroe et al., 1994). The annealed linker primers are: MSET: 5′-TAGAAGGAGAGG-ACGCTGTCTGTCGAAGG-3′ and MSEB: 5′-GAG-CGAATTCGTCAACATAGCATTTCTGTCCTCTCC TTC-3′. The amplification (linker) primer is: LNP:

Table 1. Ya8 accession numbers, primers, location, and product sizes

Name	Accession #	5′ Primer sequence (5′-3′)	3′ Primer sequence (5′-3′)	A.T.[1]	Chromosomal location[2]	Product size[3]	
						Filled	Empty
Ya8NBC1	AC006959	CCTGCTGACATTTAGAAATGACTCT	ATATACAAGTCATCAGATGGGGACAC	60°C	5	504	293
Ya8NBC2	AC006556	GCCTGTGTACCTCCTTTAAATATCTTG	CTCAAAAACTGGAGCAGGAGTAA	50°C	21	503	242
Ya8NBC3	AC006989	GGTGGTCATCCATATACTATCTCATAGG	AGAGTTCTGGAAAAGTTGACAGGAT	55°C	Y/X[4]	498	178
Ya8NBC4	AL049871	CATTCCACCCTGTCAGCATT	GCTTTGGAAGTAGGCAGGTTAC	60°C	14	536	204
Ya8NBC6	AC004066	ACTTAGCTTTGAGTATTTTTCTGAACTATC	CTAAAATGGAGGTACCGATATACTTTATTA	60°C	4	470	132
Ya8NBC8	AL034422	GGATCACAAACCTAAATGAAAGAGGTAA	CCGTCTCAAAACAAACAGACAAATA	60°C	20	501	155
Ya8NBC10	AC004893	GGATTACTTTGATGAAAATATCTTAGTAGG	AACTGGATTGTACTTTGAAGACCAC	60°C	7	757	371
Ya8NBC11	AC007688	GAGTGCCTATTATGTGTTAGGTACTTTGCT	ACTCTCACTAGATTATAAGCCCCATAAGGA	60°C	12	419	105
Ya8NBC12	AL022302	CATCTTAAAAGACATTAGAAAAGTACACAG	CTGGCCACTTAGTATATTTTCAATCAG	60°C	22	530	211
Ya8NBC13	AL008722	CCATTTTCTATAAGAAGGCTTCACC	AAAGTAATGTGAAAGTATTGGAGAAGAGAT	60°C	22	402	77
Ya8NBC14	AF094481	GAATCTCTATCTCTGACACTAGCCACT	GGCAACAAGTCTGATGAATACTTAAAGGAG	60°C	3	500	189
Ya8NBC15	AF179296	CTCTACAGTACAGATGAGAAAGTACAGACA	CGCCTTGCTAGATTTCTTTCTAATG	60°C	8	620	299
Ya8NBC17	AC005205	CTAGTTCCCACATACCGAAAACAC	CCTGTCTCGTTCAGTCTTCTTTG	58°C	19	501	155
Ya5NBC60	AC006553	CAGTCCATAGCAGTCATGGTAAATAAG	AAGTCTATACCGGTTACCTCTTTCTT	58°C	4	456	149

[1] Amplification of each locus required 2:30 min @ 94°C initial denaturing, and 32 cycles for 1 min 94°C, 1 min Annealing Temperature (A.T.) and 1 min elongation at 72°C, with a final extension time of 10 min at 72°C.
[2] Chromosomal location determined from Accession information or by PCR analysis of monochromosomal hybrid cell lines.
[3] Empty product sizes calculated by removing the Alu element and one direct repeat from the filled sites that were identified.
[4] Ya8NBC3 is located in the pseudoautosomal region of the X and Y chromosome.

5′GAATTCGTCAACATAGCATTTCT-3′. We placed an *Eco*RI site at the 5′ end of the primer for the option of cloning PCR products into cloning sites of common vectors. No bands are observed on a gel when this primer is used alone with the anchored template at an annealing temperature of 55°C.

Unless otherwise noted, PCR conditions (for all ASAP reactions) were performed in 20 μl using a Perkin-Elmer 9600 thermal cycler with the following conditions: 1 × Promega buffer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 0.25 μM primers, 1.5 U Taq polymerase (Promega) at 94°C – 2 min, 94°C – 20 s, 62°C – 20 s, 72°C – 1 min, 10 s, for 5 cycles; 94°C – 20 s, 55°C – 20 s, 72°C – 1 min, 10 s, for 25 cycles; 72°C – 3 min. Nested Alu primers were used that move along the Alu in an upstream direction as follows: ASII (Ya5-specific): 5′-CTGGAGTGCAGTGGCGG-3′; HS18R (Ya8-specific): 5′-CTCAGCCTCCCAAGTAGCTA-3′; HS16R (Ya8-specific): 5′-CGCCCGGCTATTTTT-GTAG-3′.

The ASII primer has Ya5 diagnostic nucleotides (present in both Ya5 and Ya8 subfamilies). In the first round of PCR, stock genomic DNA (2.4 ng anchored DNA) was used as the template. For subsequent rounds of amplification, PCR products were purified through microcon-30 (Amicon) columns using two centrifuge spins following the addition of 400 μl of water. For the second round of amplification, 1 μl of microcon-purified first round PCR reaction was used as the template, and for the third round 1 μl of microcon-purified second round PCR products was used. For display analysis (see below) the PCR products were ‘equalized’ in volume following microcon purification.

Display of anchor-Alu PCR products

Third round PCR was performed utilizing a 5′ end-labeled primer incorporating [γ-^{32}P] ATP (Amersham) with T4 polynucleotide kinase (New England BioLabs). PCR conditions were as above with the exception of using 0.188 μM of each Ya8 and LNP cold primers and 0.075 μM of end-labeled Ya8 primer. Anchor-PCR and end-labeled molecular weight markers (ϕX174 DNA digested with *Hinf*I; Promega) were separated by electrophoresis on denaturing 5% long ranger (AT Biochem) gels, and examined by autoradiography following exposure to Amersham Hyperfilm at room temperature. DNA samples from different ethnic groups were utilized in the display to identify variants that resulted from recent Alu insertion events (polymorphism).

Verification of PCR generated DNA fragments as Ya8 products

Gels were aligned to autoradiographs by either small cuts in various parts of the gel, or placement of low-level radioactive dye on the gel prior to re-exposure. Bands were then sliced out of the gels, placed in 200 μl of water and eluted by heating at 65°C for 15 min. Samples were re-amplified with third round PCR primers, cloned and sequenced as described above. Following verification these bands were amplified by the third round primer pair, new nested oligonucleotides based on the flanking unique sequences were designed to move, by PCR, downstream through the Alu element to the opposite flank. Annealing temperatures were adjusted to reflect the Tm of the oligonucleotide primers. Generally two or three rounds of PCR were utilized to obtain the 3′ flanking sequences of the Alu. These PCR products were also cloned and sequenced in the same manner.

Results

We present two complementary approaches that facilitate rapid detection of newly inserted Alu elements from the human genome. First, computational analyses of human genomic DNA sequences from the GenBank database are used in the identification of recently integrated Alu elements. Second, allele-specific PCR amplification is used for the selective enrichment of young Alu elements. To compare and contrast these two approaches, we present the data obtained when these methods are applied to the identification of members of the Ya8 Alu subfamily, the smallest previously reported subfamily of Alu repeats in the human genome.

Copy number and sequence diversity

In order to estimate the copy number of Ya8 Alu family members, we determined the number of exact matches to our subfamily specific oligonucleotide query sequence as a proportion of the human genome that had been sequenced in the non-redundant database. We obtained 27 matches to the subfamily specific query sequence from the non-redundant database. Upon further sequence annotation using the RepeatMasker2 web site, five matched the Ya8 Alus

previously sequenced in our laboratories (Batzer et al., 1990; Batzer & Deininger, 1991; Batzer et al., 1995). Eight of the elements identified in the search were classified as Alu Sx subfamily members, and two matched the TPA 25 Ya8 Alu family member. A total of 13 independent Ya8 Alu elements were identified from the search of the non-redundant database that were not sequenced as part of a project to specifically identify recently integrated Alu elements. The non-redundant database contained 45.3% human DNA sequences for a total of 590,140,703 bases of human sequence on the date of the search. The estimated size of the Ya8 subfamily is (3×10^9 bp/590, 140, 703 bp) $\times 13$ unique Ya8 matches $= 66$ Ya8 subfamily members. This estimate compares favorably with that of 50 previously reported based upon library screening, restriction digestion or Southern blotting (Batzer et al., 1995). An additional six matches to the Ya8 subfamily query sequence were identified in the HTGS. One of these elements was an Alu Sq subfamily member, while a second element was a duplicate copy of Ya8NBC60. PCR analyses of two elements identified in the high throughput database, Ya8NBC7 and Ya8NBC16 (GenBank accession numbers AL109937 and AC008944), were inconclusive and these elements were eliminated from further analysis. These two elements were identified from low pass first sequence runs in the HTGS database. It is not surprising that the PCR analyses failed, since the DNA sequences are of presumably lower quality than finished DNA sequences contained in the non-redundant database. However, two additional Ya8 Alu repeats (Ya8NBC8 and Ya8NBC15) were identified in the HTGS database and subjected to further analysis.

A comparison of the nucleotide sequences of all of the Ya8 Alu family members is shown in Figure 2. In order to determine the time of origin for the Ya8 subfamily we divided the nucleotide substitutions within the elements into those that have occurred in CpG dinucleotides and those that have occurred in non-CpG positions. 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 & Striker, 1989; Batzer et al., 1990) as a result of the deamination of 5-methylcytosine (Bird, 1980). A total of 14 non-CpG mutations and 8 CpG mutations occurred within the 14 Alu Ya8 subfamily members reported. Using a neutral rate of evolution for primate intervening DNA sequences of 0.15% per million years (Miyamoto, Slightom & Goodman, 1987) and the non-CpG mutation rate of 0.413%

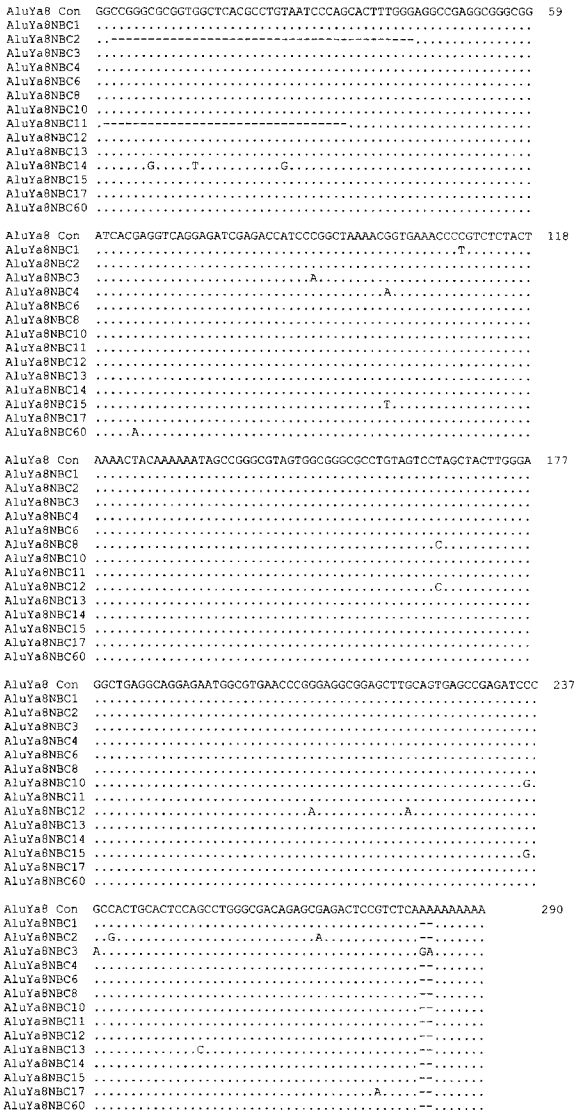

Figure 2. Multiple alignment of Ya8 subfamily members. The Ya8 subfamily consensus (con) is derived from the most common nucleotide found at each position within the subfamily members. Nucleotide substitutions at each position are indicated with the appropriate nucleotide. Deletions are marked by '–'.

(14/3388 using only non-CpG bases) within the 14 Ya8 Alu elements yields an estimated age of 2.75 million years old for the Ya8 subfamily members. This estimate of age is somewhat higher than the 660,000 years previously reported (Batzer et al., 1995). However, the previous study of Ya8 Alu family members involved only four elements making the calculated age more subject to random statistical fluctuation. This estimate is also consistent with the expansion of a family of mobile elements that began around the time humans

Ya8NBC1	AAGAGGGGGAGAG	[Alu]	A_{18}	AAGAGGGGGAGAG
Ya8NBC2	GGA	[Alu]	$A_{16}CA_4$	TGGA
Ya8NBC3	GAAGAAGTTTTGC	[Alu]	$ACA_{21}CA_2$	GAAGAAGTTTTGC
Ya8NBC4	CGACAATTT	[Alu]	$A_{17}CA_{13}CA_{10}$	CCGACAATTT
Ya8NBC6	AAATTTAAAATATT	[Alu]	A_{44}	AAATTTAAAATATT
Ya8NBC8	AAGAAAATATAGGCATA	[Alu]	$A_{11}CA_{14}CA_{23}$	AAGAAAATATAGGCATA
Ya8NBC10	AAAAATAAAATA	[Alu]	A_4	AAAAATAAAATA
Ya8NBC11	AAGGAATGAGACTG	[Alu]	A_{20}	AAGGAATGAGACTG
Ya8NBC12	AAAGTTCTTTGCA	[Alu]	A_{27}	AAAGTTCTTTGCA
Ya8NBC13	AAGAAGGCTTCACCAG	[Alu]	A_{30}	AAGAAGGCTTCACCAG
Ya8NBC14	ATCCC	[Alu]	A_{26}	ATCCC
Ya8NBC15	AGAACCACCAGGAA	[Alu]	A_{27}	AGAACCACCAGGAA
Ya8NBC17	AAGGAATCTC	[Alu]	A_{17}	AAGGAATCTC
Ya8NBC60	GGTAAATAAGCTTTCTT	[Alu]	A_{25}	GGTAAATAAGCTTTCTT

Figure 3. Nucleotide sequences flanking Ya8 subfamily members. Nucleotide sequences flanking the Ya8 Alu family members are shown. Nucleotides encompassed in the direct repeats are underlined. The length of the oligo-dA rich tail is denoted by an (A) and a subscript indicating the number of adenine residues.

and African apes diverged, which is thought to have occurred 4–6 million years ago (Miyamoto, Slightom & Goodman, 1987).

Inspection of the nucleotide sequences flanking each Ya8 Alu family member shows that all of the elements were flanked by short perfect direct repeats (Figure 3). The direct repeats ranged in size from 3–17 nucleotides. These direct repeats are fairly typical of recently integrated Alu family members. Two of the Alu Ya8 Alu family members contained 5′ truncations (Ya8NBC2 and Ya8NBC11). Since Ya8NBC2 and Ya8NBC11 are both flanked by perfect direct repeats the truncations in these elements probably occurred as a result of incomplete reverse transcription or improper integration into the genome rather than by post-integration instability. All of the Ya8 Alu family members had oligo-dA rich tails that ranged in length from a minimum of four nucleotides to over 40 bases in length. It is also interesting to note that the 3′ oligo-dA rich tails of several of the elements (Ya8NBC2, Ya8NBC3, Ya8NBC4, and Ya8NBC8) have accumulated random mutations beginning the process of the formation of simple sequence repeats of varied sequence 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., 1995b).

Phylogenetic distribution, and chromosomal location

The phylogenetic distribution of each Ya8 Alu element was determined by amplifying genomic DNA from two non-human primates (common chimpanzee and gorilla). All of the Ya8 Alu family members were absent from the genomes of non-human primates. This suggests that the majority of these elements dispersed within the human genome sometime after the human and African ape divergence. The chromosomal location of each Ya8 Alu element was taken directly from the GenBank database entry or determined by PCR amplification of human/rodent monochromosomal hybrid cell line DNA samples (Table 1).

Human genomic diversity

In order to determine the human genomic variation associated with each of the Ya8 Alu family members we subjected a panel of human DNA samples to PCR amplification (Table 2). The panel was composed of 20 individuals of European origin, African Americans, Greenland Natives and Egyptians for a total of 80 individuals (160 chromosomes). Using this approach four of the 14 (Ya8NBC8, Ya8NBC10, Ya8NBC14 and Ya8NBC15) Alu Ya8 subfamily members were monomorphic for the presence of the Alu element suggesting that these elements integrated in the genome prior to the radiation of modern humans from Africa. Three of the elements (Ya8NBC2, Ya8NBC13 and Ya8NBC17) appeared heterozygous in all of the individuals that were analyzed, suggesting that they had integrated into previously undefined repetitive elements within the human genome as previously described (Batzer et al., 1991). However, the remaining seven elements were polymorphic for the presence of an Alu repeat within the genomes of the test panel individuals (Table 2). The unbiased heterozygosity values (corrected for small sample sizes) for these polymorphic Alu insertions were variable, and approached the theoretical maximum in several cases. This is quite interesting since the maximum uncorrected heterozygosity for these biallelic elements is 50% and suggests that these Alu insertion polymorphisms will make excellent markers for the study of human population genetics. In addition, 50% of the randomly identified Ya8 Alu family members are polymorphic. These results suggest that the Ya8 subfamily is younger than either the Ya5 (from which Ya8 was derived) or Yb8 Alu subfamilies, since only 25% of the members of these Alu subfamilies are polymorphic in the human genome (Batzer et al., 1995).

Allele-Specific Alu PCR (ASAP)

Although database screening is extremely efficient for identifying recent Alu elements, it will not allow identification of new elements from genomes not included in the sequencing efforts. Our primary objective with the ASAP technique is to rapidly identify newly inserted Alu elements from a background of 500,000 older Alus. To accomplish this feat, we utilized a

Table 2. Alu Ya8 associated human genomic diversity

Elements	African American					Greenland natives					European					Egyptian					
	Genotypes			*f* Alu	Het	Genotypes			*f* Alu	Het	Genotypes			*f* Alu	Het	Genotypes			*f* Alu	Het[1]	Avg Het[2]
	+/+	+/−	−/−			+/+	+/−	−/−			+/+	+/−	−/−			+/+	+/−	−/−			
Ya8NBC1	10	2	7	0.58	0.50	5	0	9	0.36	0.35	10	5	1	0.78	0.48	8	0	10	0.44	0.51	0.46
Ya8NBC3	0	12	7	0.32	0.44	0	6	14	0.15	0.44	0	12	7	0.32	0.26	0	9	10	0.24	0.51	0.41
Ya8NBC4	1	4	13	0.17	0.29	6	0	7	0.46	0.51	8	5	6	0.55	0.52	18	0	1	0.95	0.10	0.35
Ya8NBC6	8	2	6	0.56	0.51	11	0	3	0.85	0.00	16	0	0	1.00	0.35	12	2	3	0.76	0.37	0.31
Ya8NBC11	13	2	0	0.93	0.13	12	0	0	1.00	0.09	10	1	0	0.95	0.00	13	3	0	0.91	0.18	0.10
Ya8NBC12	17	0	0	1.00	0.00	19	0	0	1.00	0.05	18	1	0	0.97	0.00	17	0	0	1.00	0.00	0.01
Ya8NBC60	6	9	3	0.58	0.50	6	7	5	0.53	0.51	5	9	3	0.56	0.51	10	5	4	0.66	0.46	0.49

[1]This is the unbiased heterozygosity.
[2]Average heterozygosity is the average of the population heterozygosity.

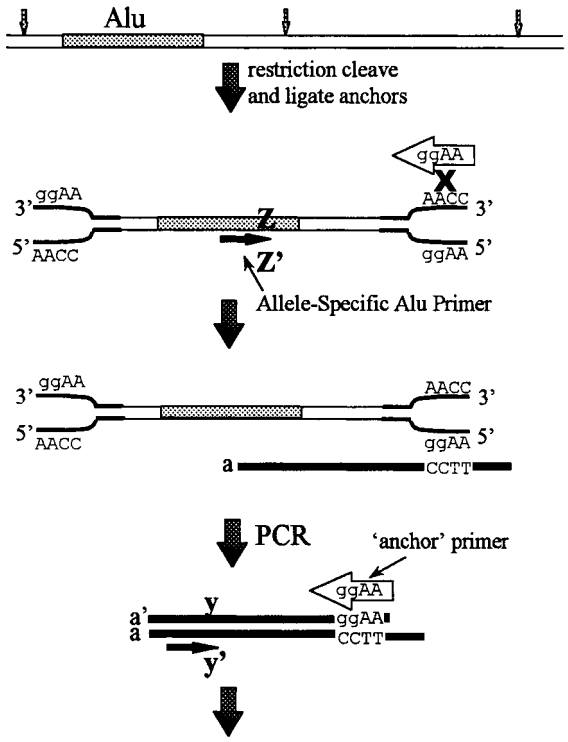

Figure 4. The Allele-Specific Alu PCR (ASAP) anchor strategy. Schematic diagram of the technique for the isolation of a designated subset of Alu repeats based on a modification of the IRE-bubble PCR technique (Munroe et al., 1994). The shaded rectangle represents an Alu sequence in genomic DNA. The *Mse*I (or an alternative restriction enzyme) cleaves in unique sequences flanking the Alu repeat (small arrows). The anchors with the complementary *Mse*I site are ligated. The anchors are designed so that the two oligonucleotide strands base-pair only at the *Mse*I site end, but not at the other end (represented here schematically with four arbitrary bases). PCR is initiated using an allele-specific Alu primer (Z′). The anchor primer will not be able to base pair preventing anchor-to-anchor amplification. Only those fragments (a) generated by the Alu primer are available for amplification by the anchor primer. The amplified product (a and a′) provides a template for nested PCR (primer y′) to further decrease the background.

modification of the IRE-bubble PCR technique (Munroe et al., 1994). The procedure utilizes an anchored PCR strategy (Figure 4) in which genomic DNA is cleaved with an enzyme that does not cleave within the Alu repeat. The modified anchor is then ligated to the fragment ends. This anchor will only allow PCR amplification if a primer first primes within the fragment and replicates across the linker eliminating any problems with amplification from anchor to anchor. We take advantage of the base changes that identify the younger Alu subfamily members (Batzer et al., 1996b; Batzer & Deininger, 1991). In addition, this allows the selective enrichment for a smaller fraction of the Alu elements from the genome, as there are only 1000 Ya5 and 1000 Yb8 Alu repeats and approximately 50 Ya8 Alu family members in the human genome (Batzer et al., 1995). We gain the specificity for the recent inserts by using a PCR primer that matches the particular Alu subfamily with the diagnostic positions at its 3′ end. Each amplification will extend from a specific Alu subfamily member through its upstream flanking sequences to the randomly located flanking restriction site. The numerous older Alu repeats have accumulated many mutations and may compete for the PCR primers with the Ya5/8 elements. Therefore, although the first amplification provides a great deal of subfamily specificity, we then carry out a 'nested' reaction using a second allele-specific primer to improve the specificity, followed by a third round with another allele-specific primer. In theory, we can utilize primers for each of the 5–8 diagnostic mutations in a subfamily.

In the example presented in this paper, we focused our attention on the identification and display of the lower copy number Alu Ya8 subfamily. Also, to better display the results, we used nested primers in the upstream direction of Ya8 to avoid amplification problems through the A-rich tail. Using the primers described in the Materials and methods section, by the third round of PCR, we were able to visualize discrete DNA fragments on an agarose gel (data not shown). The size range of these fragments appeared to be between 150 bp and 800 bp. To enhance this display, we chose an alternative method of electrophoretic separation and end-labeled the nested primer to further minimize background (see below). To verify these were Ya8 repeats, we directly cloned the third round PCR products and sequenced them. Partial or complete sequences of these products, using vector primers in both directions, demonstrated all 12 clones to be amplified by the Alu-anchor primer pair, although in one case the unique linker sequence was imprecise. All these elements contained the Ya5/8 diagnostic nucleotides (There were no further upstream diagnostics to declare these as Ya8 elements.).

For eight of the 12 isolated clones, there were between 12 and 18 unique nucleotides between the linker and the Alu (or truncated Alu) sequences. Since Alu elements preferentially insert into A-T rich regions (Daniels & Deininger, 1985) and *Mse*I cuts at the sequence TTAA, then this result is not surprising. The advantage of using *Mse*I for the restriction digestion is that most of the Alu-linker products are

small enough to be amplified. Although it would be difficult to perform nested PCR in the opposite direction with those few A-T rich nucleotides, searching GenBank using the BLAST program with the obtained flanking unique DNA sequences as the query may in some cases identify the rest of the genomic sequence for each Alu element. This will provide the Alu location with both its flanking sequences. Flanking unique sequence primers can then be designed and the Alu polymorphism can then be confirmed using other human DNA sources. Once the polymorphism is confirmed subsequent population studies can be performed.

Display and rapid identification of Ya8 associated variants

To alleviate the need for testing every Ya8 element obtained by this assay, we chose to end-label the third round nested PCR primer to enable a display of individual Ya8 repeats following electrophoretic separation and autoradiography. Observed variations may be due to primer mismatch, genomic rearrangements, small insertion/deletions or Alu based insertion/deletions (I/D).

We carried out the procedure with four different individuals to discern which bands represent variants (Figure 5), and to effectively display variants as DNA fingerprints. We obtained about 40 bands per individual from a single reaction. Among the four individuals analyzed, about one half of the bands appeared variant (Figure 5). We have developed a potent method for the generation of Ya8 associated DNA fingerprints that is in reasonable agreement with the database mining approach and seems to display the majority of Alu subfamily members. This necessitated addressing what proportion of the fragments generated were the result of the presence of a Ya8 Alu element and whether the lack of the same band in another individual represented an Alu insertion polymorphism. We chose 12 bands to re-amplify and verify as Ya5/8 elements. Those bands that appeared variant were analyzed for Alu insertion polymorphisms. Other bands were selected for future testing of dimorphisms as these individual Ya8 elements may display variation among other people/populations. Occasionally, upon re-amplification from the isolated band, we obtained background products and therefore, generally more than one clone was sequenced. Of the 12 isolated bands (Figure 5) nine were verified as precisely amplified HS16R-LNP products. Two others each contained

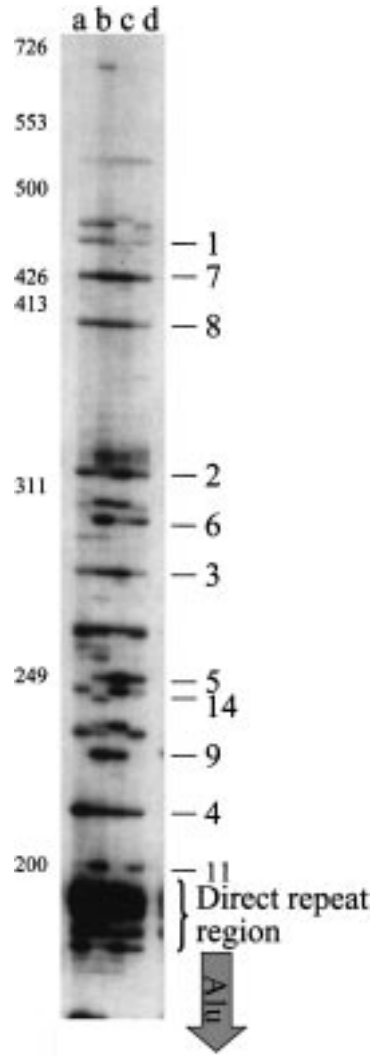

Figure 5. DNA fingerprints of unrelated individuals based on anchored-Alu PCR. Individual bands are numbered for identification purposes. Fragment lengths are shown in nucleotides to the left. DNA samples used are of Caucasian (lane a), Hispanic (lane b), Hindu–Indian (lane c) and Chinese (lane d) descent.

a Ya5/8 Alu, one randomly amplified by HS16R (anc-8) in lieu of the linker primer, while anc-3 contained sequences downstream of HS16R. Anc14 apparently was an amplified J (PS) Alu element (data not shown). Therefore, this demonstrates the majority of the bands visualized on the autoradiograph are AluYa5/8 repeats and most probably Ya8. The numerous bands at about 178 nt coincide with our previous finding that many of the products will have between 12 and 18 unique sequences. Of the nine bands where we attempted to obtain the opposite flank by nested anchored PCR, we reached the opposite (downstream) flank of the Alu for

three of them (anc-5, anc-6, anc-4). In some cases the amount of unique sequence was too small to employ nested primers, and in some cases there was a high level of A-T richness. In one case we merely got a non-specific product. All three sequences obtained were authentic Ya8 Alu elements based on the diagnostic nucleotide positions and the high level of conservation of the sequence in relation to the consensus. This demonstrates the successful nature of our protocol to select for this subfamily of repeats amongst a large background of Alu repeats.

When 'crossing' the anc-5 Alu by nested PCR using four individuals (not all identical to Figure 5), we found a correspondence between the generation of a distinct band among the individuals that also had the anc-5 band on an autoradiograph. However, we obtained a short 3′ flank of 12 nucleotides that proved difficult in amplifying DNA from various individuals with unique flanks. It is still possible that this variant represents an I/D event. Besides anc-5, anc-6 also appeared polymorphic on the autoradiograph, although anc-4 did not. However, since we had both flanks, for these Alu elements, we developed primers to rapidly assess various individuals for an insertion variant. For anc-6, one of a few different primer sets worked well, yielding the band of expected size, although also generating a few non-specific bands. However, a band was present for 11 unrelated individuals analyzed (data not shown), including those observed on the autoradiograph, suggesting that the anc6 polymorphism was not the result of an I/D variant. In addition, this band was absent in the chimpanzee, possibly indicating the absence of the Alu or perhaps primer mismatch due to nucleotide divergence. Although anc-4 was not variant on the autoradiograph, we tested 13 individuals of various ethnic backgrounds for an I/D event and observed it to be monomorphic. Although we have not verified any of the displayed variants to be the result of an Alu insertion, this potential remains, as we observed Ya8 elements to be highly polymorphic, and all the bands, but one, analyzed were Ya8 repeats.

Discussion

In this manuscript we present an analysis of the smallest defined subfamily of Alu elements located within the human genome termed Ya8. This subfamily of Alu elements was derived from the Ya5 subfamily of Alu elements. The Ya5 subfamily is composed of approximately 1000 members and has largely integrated into the human genome sometime after the human-African ape divergence. The main reasons that supported the more recent origin of the Ya8 subfamily are the accumulation of three additional diagnostic mutations as compared to the Ya5 subfamily and the lower copy number for the Ya8 subfamily. It is also important to note that a higher percentage of the Ya8 Alu family members (50%) are polymorphic for insertion presence/absence as compared to only 25% polymorphism in the Yb8 and Ya5 Alu subfamilies. These data also suggest a recent origin for the Alu Ya8 subfamily within the human genome. However, it is still possible that the Ya8 Alu subfamily may have amplified from an allelic variant of the Ya5 subfamily that was not as efficient at mobilization as the Ya5 source gene.

The ability to detect a handful of Alu repeats from the background of several hundred thousand Alu elements in the human genome is impressive. The application of computational biology to the analysis of large multigene families such as Alu repeats offers the potential to address a number of new questions in comparative genomics as an increasing proportion of the human genome is sequenced. Studies of the present, as well as ancient, integration patterns of mobile elements in the human genome may begin to be addressed. In addition, the patterns of diversity generated by the integration of mobile elements into the human genome may be analyzed at a scale that was previously unimaginable. These types of studies will shed new insight into the relationships between different types of mobile elements in the human genome, integration site preferences, impact, and the biological properties of these elements.

The development of the ASAP technique facilitated the display of a subset of Ya8 Alu elements from a large and complex background. The preferential isolation of the young Alu elements, as demonstrated here, enhances the identification of recent Alu insertion events in the genome. We focused our efforts on the smallest known defined subfamily of Alu repeats to best address issues of sensitivity of the display of individual elements. One of the advantages of this technique is its flexibility. Altering the restriction enzyme used for digestion of genomic DNA selects for distinct subsets of Alu elements within a particular subfamily, since this technique preferentially amplifies products that range from 200 and 800 bp in size. In addition, modifications to the ASAP technique, such as the use of a less frequent restriction endonuclease, may allow for a display of subsets of the larger groups of Alu repeats such as Ya5 elements. Alternatively, the

use of primers that select for subfamily 'subgroups' may also be used to reduce the complexity of the resultant display by decreasing the number of PCR products. Although we focused on Ya8 Alu elements due to their low copy number, the young Yb8 Alu subfamily is another alternative for ASAP with an estimated copy number of only 1000 elements (Batzer et al., 1995; Zietkiewicz et al., 1994) and some polymorphic members (Hutchinson et al., 1993; Hammer 1994; Arcot et al., 1998). We have previously demonstrated the isolation of young Alu elements (based on sequence identity to a consensus) using a Yb8 diagnostic primer, and a generic Alu as an anchor in the amplification reaction, that can be profiled with minimal background (Kass, Batzer & Deininger, 1996). It is conceivable that variations on the anchored-Alu PCR technique can be employed to rapidly localize individual elements from all three subfamilies of young Alu elements.

Once the flanking sequences of the young Alu elements are obtained, the PCR strategy can be employed to trace polymorphisms that have resulted from recent Alu insertions and are not yet fixed in human populations. The anchored-Alu PCR approach not only facilitates rapid identification of young elements by displaying the amplification products, but will also increase the potential for selecting only those mobile element fossils that exhibit presence/absence variation. Selection in this manner also shifts the spectrum for new elements toward the elements that are lower frequency and less likely to be held in common between individuals or populations. Therefore, this approach should prove to be quite useful for the ascertainment of mobile element fossils to address questions about more recent human diversifications. In contrast, the identification of mobile element fossils using computational biology affords the opportunity to identify multiple frequency classes of Alu elements that are shared at different geographic levels within the human population.

The ASAP method's strength comes from its ability to isolate a subset of interspersed repeat sequences from different DNA sources and compare them at the same time. In other words, this approach is not limited to Alu elements, but may be used with other SINEs (from other organisms) or even long interspersed elements (LINEs) or for that matter any repeated DNA sequence family that has a defined subfamily structure. A second potential application would be the use of ASAP to monitor genomic instability associated with different forms of cancer by providing a multi-locus monitoring system. Due to its high flexibility the ASAP technique has an enormous range of potential applications.

Mobile element fossils have proven to be simple powerful tools for tracing the origin of human populations (Perna et al., 1992; Batzer et al., 1994a,b, 1996a; Stoneking et al., 1997). These elements should also prove quite useful to the forensic community as paternity identity testing reagents (Batzer & Deininger, 1991; Novick et al., 1993). Some Alu insertion polymorphisms have been identified by chance (Deininger & Batzer, 1995) while others have been identified by library screening in a directed approach (Batzer & Deininger, 1991; Batzer et al., 1995; Arcot et al., 1995a, b, c; Batzer et al., 1996a; Arcot et al., 1998). Here, we have presented two complementary methods involving computational biology and PCR based displays that will enhance our ability to identify the genomic fossils of recently integrated mobile elements from complex genomes. These approaches will contribute to a new era in biological sciences that will increasingly rely upon informatics/computational biology as well as hard-core bench molecular biology to answer global questions in comparative genomics.

Acknowledgements

AMR, MLC, DHK and SUN contributed equally to this paper. DHK was supported by an Eastern Michigan University Spring/Summer Research Award and a University Research in Excellence Fund. AS was supported by a fellowship from the Academy of Scientific Research, Ministry of Scientific Research and Informatics, Cairo, Egypt. This research was supported by National Institutes of Health RO1 GM45668 to PLD, Department of the Army DAMD17-98-1-8119 to PLD and MAB, and award number 1999-IJ-CX-K009 from the Office of Justice Programs, National Institute of Justice, Department of Justice to MAB. Points of view in this document are those of the authors and do not necessarily represent the official position of the US Department of Justice.

References

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

Arcot, S.S., T.H. Shaikh, J. Kim, L. Bennett, M. Alegria-Hartman, D.O. Nelson, P.L. Deininger & M.A. Batzer, 1995a. Sequence

diversity and chromosomal distribution of 'young' Alu repeats. Gene 163: 273–278.

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

Arcot, S.S., A.W. Adamson, G.W. Risch, J. LaFleur, M.B. Robichaux, J.E. Lamerdin, A.V. Carrano & M.A. Batzer, 1998. High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. J. Mol. Biol. 281: 843–856.

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

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith & K. Struhl, 1996. Current Protocols in Molecular Biology, Wiley, Canada.

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 & 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 & E. Zuckerkandl, 1996b. Standardized nomenclature for *Alu* repeats. J. Mol. Evol. 42: 3–6.

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 & 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. Alegria-Hartman, H. Bazan, D.H. Kass, G. Novick, P.A. Ioannou, D. Boudreau, W.D. Scheer, R.J. Herrera, M. Stoneking & P. Deininger, 1994a. Alu repeats as markers for human population genetics. IVth International Symposium on Human Identification, 49–57.

Batzer, M.A., M. Stoneking, M. Alegria-Hartman, H. Bazan, D.H. Kass, T.H. Shaikh, G. Novick & P.A. Ioannou, 1994b. African origin of human-specific polymorphic Alu insertions. Proc. Natl. Acad. Sci., USA 91: 12288–12292.

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

Batzer, M.A., V. Gudi, J.C. Mena, D.W. Foltz, R.J. Herrera & P.L. Deininger, 1991b. Amplification dynamics of human-specific (HS) Alu family members. Nucleic Acids Res. 19: 3619–3623.

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

Batzer, M.A., C.W. Schmid & P.L. Deininger, 1993. Evolutionary analyses of repetitive DNA sequences. Methods Enzymol. 224: 213–232.

Bird, A.P., 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic. Acids. Res. 8: 1499–1504.

Daniels, G. & P.L. Deininger, 1985. Repeat sequence families derived from mammalian tRNA genes. Nature 317: 819–822.

Deininger, P.L. & M.A. Batzer, 1995. SINE master genes and population biology, pp. 43–60 in The Impact of Short, Interspersed Elements (SINEs) on the Host Genome, edited by, R. Maraia, R.G. Landes, Georgetown, TX.

Deininger, P.L. & M.A. Batzer, 1993. Evolution of Retroposons, pp. 157–196 in Evolutionary Biology edited by M.K. Heckht et al., Plenum Publishing, New York.

Deininger, P.L. & M.A. Batzer, 1999. Alu repeats and human disease. Mol. Genet. Metab. 67: 183–193.

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.

Hutchinson, G.B., S.E. Andrew, H. McDonald, Y.P. Goldberg, R. Graham, J.M. Rommens & M.R. Hayden, 1993. An Alu element retroposition in two families with Huntington disease defines a new active Alu subfamily. Nucleic. Acids. Res. 21: 3379–3383.

Jurka, J., P. Klonowski, V. Dagman & P. Pelton, 1996. CENSOR – a program for identification and elimination of repetitive elements from DNA sequences. Computers and Chemistry 20(1): 119–122.

Kass, D.H., C. Alemán, M.A. Batzer & P.L. Deininger, 1994. An HS Alu insertion caused a factor XIIIB gene RFLP. Genetica 94: 1–8.

Kass, D.H., M.A. Batzer & P.L. Deininger, 1996. Characterization and population diversity of interspersed repeat sequence variants (IRS-morphs). Genome 39: 688–696.

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

Miyamoto, M.M., J.L. Slightom & M. Goodman, 1987. Phylogenetic relations of human and African apes from DNA sequences in the psi eta-globin region. Science 238: 369–373.

Munroe, D.J., M. Haas, E. Bric, T. Whitton, H. Aburatani, K. Hunter, D. Ward & D.E. Housman, 1994. IRE-bubble PCR: a rapid method for efficient and representative amplification of human genomic DNA sequences from complex sources. Genomics 19: 506–514.

Novick, G., T. Gonzalez, J. Garrison, C. Novick, M. Batzer, P. Deininger & R. Herrera, 1993. The use of polymorphic Alu insertions in human DNA fingerprinting, in pp. 283–291 DNA Fingerprinting: State of the science, edited by S.D.J. Pena, R. Chakraborty, J.T. Epplen and A.J. Jeffreys, Birkhauser Verlag, Basel.

Perna, N.T., M.A. Batzer, P.L. Deininger & M. Stoneking, 1992. Alu insertion polymorphism: A new type of marker for human population studies. Human Biology 64: 641–648.

Shen, M.R., M.A. Batzer & P.L. Deininger, 1991. Evolution of the Master Alu Gene(s). J. Mol. Evol. 33: 311–320.

Stoneking, M., J.J. Fontius, S.L. Clifford, H. Soodyall, S.S. Arcot, N. Saha, T. Jenkins, M.A. Tahir, P.L. Deininger & M.A. Batzer, 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. 7: 1061–1071.

Zietkiewicz, E., C. Richer, W. Makalowski, J. Jurka & D. Labuda, 1994. A young Alu subfamily amplified independently in human and African great apes lineages. Nucleic. Acids. Res. 22: 5608–5612.