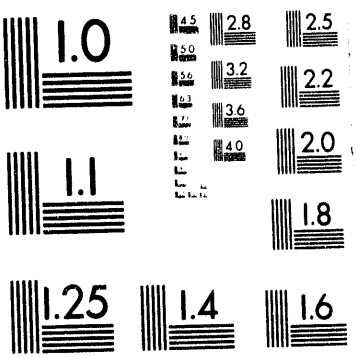

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3

1 of 1

2

Conf-940166--2

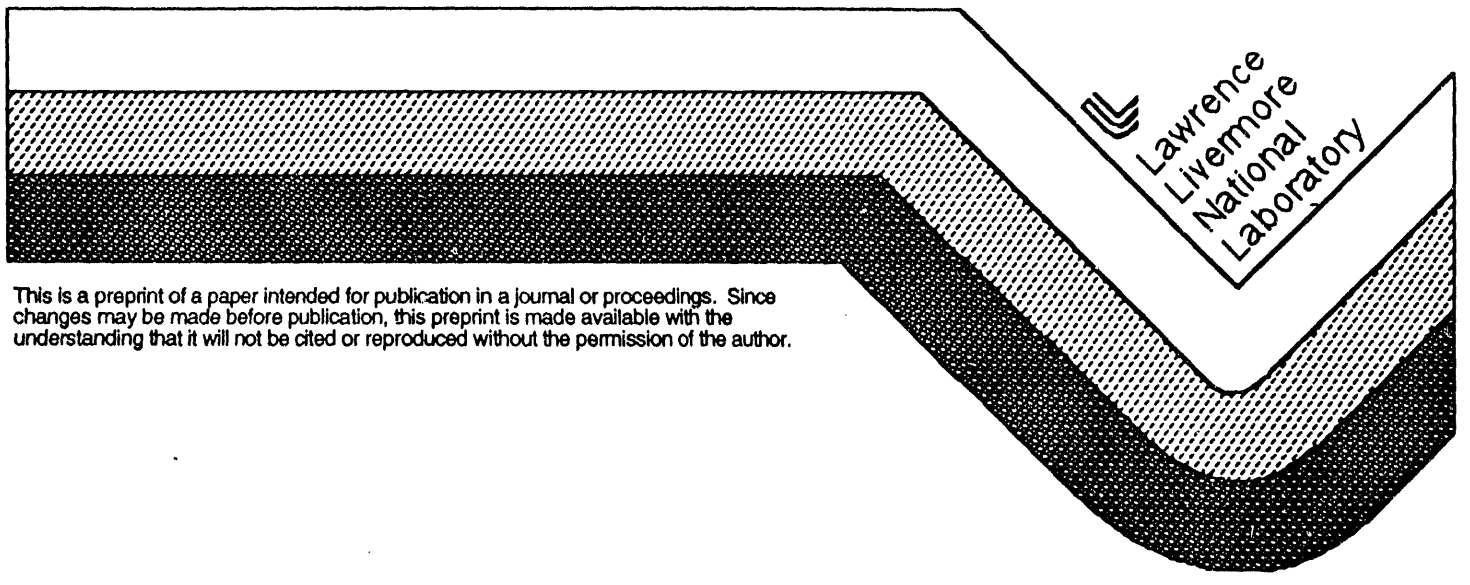
UCRL-JC-115967
PREPRINT

Alu repeats as markers for forensic DNA analyses

M. A. Batzer, M. Alegria-Hartman, D. H. Kass, G. E. Novick, T. H. Shaikh, H. A. Bazan, P. A. Ioannou, D. A. Boudreau, W. D. Scheer, R. J. Herrera, M. Stoneking, and P. L. Deininger

This paper was prepared for submittal to the First Forensic Experts Conference,
Dubai,United Araf Emirates, January 8-10, 1994

January 1994



This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Alu Repeats as Markers for Forensic DNA Analyses

Mark A. Batzer[1,*], Michelle Alegria-Hartman[1], David H. Kass[2], Gabriel E. Novick[3], Tamim H. Shaikh[2], Hernan A. Bazan[3], Panyiotis A. Ioannou[4], Donald A. Boudreau[5], W. Douglas Scheer[5], Rene J. Herrera[3], Mark Stoneking[6] and Prescott L. Deininger[2, 7].

1 Human Genome Center, L-452
Biology and Biotechnology Research Program
Lawrence Livermore National Laboratory
P.O. Box 808
Livermore, CA 94551

2 Department of Biochemistry and Molecular Biology
Louisiana State University Medical Center
1901 Perdido St.
New Orleans, LA 70112

3 Department of Biological Sciences
Florida International University
University Park Campus
Miami, FL 33199

4 The Cyprus Institute of Neurology and Genetics
P.O. Box 3462
Nicosia, Cyprus

5 Department of Pathology
Louisiana State University Medical Center
1901 Perdido St.
New Orleans, LA 70112

6 Department of Anthropology
Pennsylvania State University
University Park, PA 16802

7 Laboratory of Molecular Genetics
Alton Ochsner Medical Foundation
New Orleans, LA 70121

* To whom correspondence should be addressed.

ABSTRACT

The Human-Specific (HS) subfamily of *Alu* sequences is comprised of a group of 500 nearly identical members which are almost exclusively restricted to the human genome. Individual subfamily members share an average of 98.9% nucleotide identity with the HS subfamily consensus sequence, and have an average age of 2.8 million years. We have developed a Polymerase Chain Reaction (PCR) based assay using primers complementary to the 5' and 3' unique flanking DNA sequences from each HS *Alu* that allow the locus to be assayed for the presence or absence of the *Alu* repeat. The dimorphic HS *Alu* sequences probably inserted in the human genome after the radiation of modern humans (within the last 200,000-one million years) and represent a unique source of information for human population genetics and forensic DNA analyses. These sites can be developed into Dimorphic *Alu* Sequence Tagged Sites (DASTS) for the Human Genome Project. HS *Alu* family member insertions differ from other types of polymorphism (e.g. Variable Number of Tandem Repeat [VNTR] or Restriction Fragment Length Polymorphism [RFLP]) in that polymorphisms due to *Alu* insertions arise as a result of a unique event which has occurred only one time in the human population and spread through the population from that point. Therefore, individuals that share HS *Alu* repeats inherited these elements from a common ancestor. Most VNTR and RFLP polymorphisms may arise multiple times in parallel within a population.

INTRODUCTION

Human genetic variability is the result of the accumulation of a number of different types of mutations. These mutations may be divided into three groups which are commonly used to characterize genetic variability and occur at different rates. The most frequent group of mutations is comprised of point mutations (e.g. G:C > A:T), and are commonly assayed (presence or absence of restriction endonuclease cleavage sites) using Restriction Fragment Length Polymorphisms (RFLPs) (Botstein *et al.*,1980). The average human (*Homo sapiens*) differs from our nearest non-human primate relative, the chimpanzee (*Pan troglodytes*), by approximately 40 million single base pair substitutions which have occurred over the last 4-6 million years (Miyamoto *et al.*,1987). Individual humans differ by approximately ten million point mutations. The information contained within this relatively small number of mutations (compared to the genome which is composed of 4×10^9 nucleotides) encodes all of the signals that define each and every human as unique. Other types of mutations involve changes in the length of simple sequence repeats or Variable Number of Tandem Repeats [VNTR] (Nakamura *et al.*,1987), or the insertion and deletion of DNA sequences, including mobile genetic elements such as Long INterspersed Repeated DNA Elements (LINEs) and Short INterspersed Repeated DNA Elements (SINEs). The characterization of point mutations as well as other types of variation facilitate the study of human population genetics as well as the unique DNA based forensic identification of individuals.

The *Alu* family of SINEs is distributed throughout primate genomes (for reviews see Weiner *et al.*,1986; Schmid and Shen, 1986; Deininger, 1989; and Deininger and Batzer, 1993). The *Alu* family represents a highly successful class of mobile genetic elements, having arisen as a repeated DNA sequence family in the last sixty-five million years (Deininger and Daniels, 1986) and amplified to a copy number in excess of 500,000 (Jelinek *et al.*,1980; Deininger *et al.*,1981). Each *Alu* repeat is approximately 300 bp in length and dimeric in structure with the right half containing an additional 31 bp (Figure 1). *Alu* repeats comprise 10% of the mass of the human genome (Deininger *et al.*,1981) and are distributed, on average, one *Alu* sequence every 5000 bp of DNA. The distribution of *Alu* repeats appears to be somewhat non-random (Sainz *et al.*,1992), with a global preference for the R bands of chromosomes (Korenberg and Rykowski, 1988). *Alu* sequences are thought to be ancestrally derived from the 7SL RNA gene (Ullu and Tschudi, 1984), and mobilize through the generation of an RNA polymerase III derived transcript in a process termed retroposition (Jagadeeswaran *et al.*,1981; Rogers, 1983).

The *Alu* sequences located within primate genomes may be subdivided into groups of related sequences that share common "diagnostic" nucleotide substitutions as first reported by Slagel *et al.*, (1987). Several overlapping subfamilies of various genetic ages have been identified (Willard *et al.*,1987; Jurka and Smith, 1988; Britten *et al.*,1988; Quentin, 1988; Shen *et al.*,1991; Jurka and Milosavljevic, 1991). Here, we will use the nomenclature originally proposed by Deininger and Slagel (1988) and subsequently modified by Shen *et al.* (1991) and reviewed in Deininger and Batzer (1993). The most recently formed group of *Alu* sequences within the human genome has been termed Human-Specific (HS) (Batzer *et al.*,1990; Batzer and Deininger, 1991). This subfamily is identical to the Predicted Variant (PV) subfamily reported by Matera *et al.*, (1990 a and b).

The HS subfamily of *Alu* sequences has been found to be transcriptionally active *in vivo* suggesting the transpositional competence of this group of *Alu* repeats (Matera *et al.*,1990a). There are an estimated 500-2000 HS *Alu* sequences (Batzer *et al.*,1990; Matera *et al.*,1990a; Batzer and Deininger, 1991), the majority (Batzer and Deininger, 1991; Batzer *et al.*,1991), but not all (Leeflang *et al.*,1992) of which are restricted to the human genome. Individual HS *Alu* sequences share a high degree of nucleotide identity with each other as well as the subfamily consensus sequence (>98%) (Batzer *et al.*,1990) suggesting that they were derived from a single, or at most, a small highly related subset of source genes (Batzer *et al.*,1990; Batzer and Deininger 1991). The low level of random mutations within HS *Alu* family members has been used, along with a neutral rate of evolution, to calculate an average age of 2.8 million years (Batzer *et al.*,1990). The evolution of the putative *Alu* "master" or source *Alu* gene(s) has occurred at a neutral rate, suggesting a lack of function for these elements (Shen *et al.*,1991; Deininger *et al.*,1992). Thus, the majority of *Alu* repeats appear to be fossil relics that integrate in the genome and accumulate mutations in a random manner after integration (Labuda and Striker, 1989; Batzer *et al.*,1990). Here we report on the distribution of an *Alu* repeat within a number of human groups as determined using a PCR based assay for the presence and absence of the *Alu* insertion.

MATERIALS AND METHODS

Cell Lines and DNA Samples.

The cell lines used in this study were as follows *Homo sapiens*, HeLa (ATCC CCL2); *Pan troglodytes*, Wes (ATCC CRL1609) and *Gorilla gorilla*, Ggo-1 (primary gorilla fibroblasts) provided by Dr. Stephen J. O'Brien. Cell lines were maintained as directed by the source, DNA isolations were performed as previously described (Ausabel *et al.*, 1987). Additional human DNA samples were isolated from peripheral lymphocytes (Ausabel *et al.*, 1987) that were made available from previous studies.

Polymerase Chain Reaction (PCR) Amplification.

Amplification of DNA samples was carried out in 100 µl reactions using 100 ng of target DNA, 750 ng of each oligonucleotide, 200 µM dNTPs in 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl pH 8.4 and *Taq* polymerase (3 u) according to the supplier's (Roche Molecular Diagnostics) instructions. Each sample was subjected to the following amplification cycle: 1 min at 94°C (denature), 2 min at the appropriate annealing temperature, and 2 min at 72°C (extension) for 32 cycles in an Perkin Elmer 9600 PCR machine. The TPA 25 *Alu* family member was originally reported in one of two independent clones of the human tissue plasminogen activator locus (Friezner Degen *et al.*, 1986). Oligonucleotide primers complementary to the 5' and 3' flanking unique DNA sequence region around this HS *Alu* repeat were utilized to facilitate characterization of the locus with PCR (Batzer and Deininger, 1991; Batzer *et al.*, 1991; Perna *et al.*, 1992). Twenty microliters of each PCR reaction was subjected to fractionation on a 2% agarose gel with 0.5 µg/ml ethidium bromide. The reaction products were directly visualized using UV fluorescence.

RESULTS

A schematic representation of the *Alu* insertion PCR assay is shown in Figure 2. Primers for PCR analysis are chosen in the 5' and 3' unique flanking DNA sequences outside each *Alu* insertion. Amplification of the locus results in one of three outcomes as shown in Figure 2B. Genomes that are homozygous for the *Alu* insertion amplify a 400 bp DNA fragment (B-1). Genomes that lack the *Alu* family member amplify a 100 bp DNA fragment (B-3), while heterozygous samples will amplify both bands (B-2). This makes the *Alu* insertion assay quite easy to interpret since the amplified DNA fragments are separated by over 300 bp in length. This type of assay can be applied to address evolutionary questions about the time of origin for a single *Alu* sequence (Batzer and Deininger, 1991; Shen *et al.*, 1991: Batzer *et al.*, 1993) or to answer questions about human population genetics (Batzer and Deininger, 1991; Batzer *et al.*, 1991; Perna *et al.*, 1992) and forensics (Novick *et al.*, 1993).

A representation of an agarose gel electrophoretic separation of PCR products derived from an *Alu* insertion PCR assay for TPA 25 are shown in Figure 3. In this figure the human DNA sample (Lane 1) amplified both 400 bp and 100 bp DNA fragments indicating that this individual (HeLa cell line) was in fact heterozygous for the insertion of the TPA 25 *Alu* repeat. Amplification of DNA samples from two closely related primates (Chimpanzee and Gorilla) shown in Lanes 2 and 3, respectively, resulted in the production of a 100 bp fragment indicating that the *Alu* insertion was absent at the orthologous positions within the genomes of these non-human primates. The products derived from the amplification of several human DNA samples are shown in Lanes 4-8. Individuals were either homozygous for the presence of the *Alu* repeat (Lanes 4 and 5), heterozygous (Lane 6), or completely lacked the *Alu* insertion (Lanes 7 and 8). This type of data allows us to establish the approximate time that this polymorphism originated within the human genome, and enables us to address questions about the genetic affinity of different human population groups.

The data derived from the amplification of the TPA 25 *Alu* repeat in different human population groups is shown in Table 1. The frequency of the TPA 25 *Alu* repeat within each group varied from 0.12 - 0.64. The frequency of the *Alu* insertion was lower within African groups as compared to the non-African groups with the exception of Arhuaco (South American) Indians and Australian Aborigines. The Arhuaco Indians of northern Colombia exhibit the lowest frequency (0.12) for the *Alu* insertion. The low frequency of the TPA 25 Alu insertion may be the result of limited outbreeding within a small population group, or may be partially reflective of small sample size. Within the major ethnic groups, the variation in *Alu* insertion frequency ranged by 11% in Caucasoid groups with 15% difference in African groups demonstrating the potential use of these markers for studies of relatively closely related populations. Further studies are underway characterizing additional *Alu* insertions and different population groups.

DISCUSSION

The insertion of new *Alu* elements within the human genome appears to be occurring at a rate of 100-200/million years (Batzer *et al.*,1990; Batzer and Deininger, 1991). After insertion into the genome these elements appear to mutate at a neutral rate (Labuda and Striker, 1989; Batzer *et al.*,1991; Shen *et al.*,1991). Previous studies of orthologous members of the globin gene family have shown that *Alu* insertions represent relatively stable integrations into the genome (Sawada *et al.*,1985; Sawada and Willard, 1986; Bailey and Shen, 1993), although the partial deletion of a single *Alu* repeat has been characterized (Edwards and Gibbs, 1992). Given that the rate of *Alu* family amplification is currently very low, the independent insertion of two different *Alu* repeats at the same location in the genome is very improbable. Therefore, individuals who share *Alu* insertions inherited them from a common ancestor. This differentiates *Alu* insertions from other types of polymorphisms including RFLP (Botstein *et al.*,1980), *Alu* VpA (Economou *et al.*,1990), and VNTR (Nakamura *et al.*,1987), which may arise independently multiple times within a population. After integration in the genome, a new *Alu* repeat will suffer one of two fates, drift toward fixation for the absence of the *Alu* sequence, or drift toward an increased frequency representing a new polymorphic genome marker.

The absence of the TPA 25 *Alu* element from the genomes of non-human primates indicates that the element inserted relatively recently within the human genome. This *Alu* insertion has arisen as a polymorphism after the human/great ape divergence which is thought to have occurred within the last 4-6 million years (Miyamoto *et al.*,1987). The fact that this *Alu* element is polymorphic within different humans serves as another measure of the recent time of origin of this insertion. In this case, the *Alu* insertion has arisen as a polymorphism at or about the time that modern humans began to radiate from Africa (within the last 100,000 -1 million years, Cann *et al.*,1987). Individual *Alu* insertions represent a unique fossil record of evolutionary history. Depending on how recent a particular *Alu* element inserted within the genome it will have a different distribution. The oldest *Alu* elements inserted within primate genomes about 65 million years ago at the time of the primate-rodent divergence and are dispersed throughout the genomes of all primates. Other elements are unique to Prosimian and Anthropoid primates, and began to amplify 40-65 million years ago. Many *Alu* repeats reside within a single genus, such as the HS subfamily of *Alu* repeats which has arisen within the last 4-6 million years of evolutionary history (Batzer and Deininger, 1991). A few HS *Alu* repeats display geographic origin bias, such as the TPA 25 *Alu* family member (Batzer and Deininger, 1991; Batzer *et al.*,1991; Perna *et al.*,1992). Recent *Alu* insertions may also be unique to a family, such as the cholinesterase *Alu* family member (Muratani *et al.*,1991). Two *Alu* insertions have been reported that represent unique (*de novo*) insertions into the Factor IX (Vidaud *et al.*,1993) and NF-1 (Wallace *et al.*,1991) loci that result in hemophilia and neurofibromatosis, respectively.

The isolation and characterization of additional polymorphic *Alu* insertions is presently underway. Recent *Alu* insertions have previously been isolated in a random manner through the sequence analysis of human genes (Ryan and Dugaiczyk, 1989; Stoppa-Lyonet *et al.*, 1990; Muratani *et al.*, 1991; Jurka, 1993; Hutchinson *et al.*, 1993) or through a subfamily specific directed approach (Batzer *et al.*, 1990; Matera *et al.*, 1990 a and b; Leeflang *et al.*, 1992). The directed type of approaches take advantage of specific DNA sequence characteristics and highly specific short oligonucleotide primers (Batzer *et al.*, 1993). The specificity of directed approaches have both positive and negative aspects. On the positive side, most of the members of a single subfamily of repeated DNA sequences can be isolated permitting exhaustive analysis. Recently however, a great deal of nucleotide sequence (subfamily) heterogeneity has been reported among polymorphic *Alu* insertions. As a result directed hybridization approaches would not facilitate the isolation of every polymorphic *Alu* repeat located within the genome. Global genomic comparisons such as Representational Difference Analysis [RDA cloning] (Lisitsyn *et al.*, 1993) may represent the most efficient method for the isolation of these types of markers. The isolation and characterization of additional Dimorphic *Alu* Sequence Tagged Sites (DASTS) should eventually provide a novel system for the forensic identification of humans, as well as mapping the human genome.

REFERENCES

Ausabel, F. M., Brent, R., Kingston, R. E., More, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., 1987, "Current Protocols in Molecular Biology," John Wiley and Sons Publishing, New York, N.Y.

Bailey, A. D., and Shen, C.-K. J., 1993, Sequential insertion of *Alu* family repeats into specific genomic sites of higher primates, *Proc. Natl. Acad. Sci. USA* **90:** 7205.

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

Batzer, M. A., Schmid, C. W., and Deininger, P. L., 1993, Evolutionary Analyses of Repetitive DNA sequences, *Methods Enzymol.* **224:** 213.

Batzer, M. A., Gudi, V. A., Mena, J. C., Foltz, D. W., Herrera R. J., and Deininger, P. L., 1991, Amplification dynamics of Human-Specific (HS) *Alu* family members, *Nucleic Acids Res.* **19:** 3619.

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

Botstein, D., White, R. L., Skolnick, M. H. and Davis, R. W., 1980, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *Amer. J. Hum. Genet.* **32:** 314.

Britten, R. J., Baron, W. F., Stout, D. B., and Davidson, E. H., 1988, Sources and evolution of human *Alu* repeated sequences, *Proc. Natl. Acad. Sci. USA* **85:** 4770.

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

Deininger, P. L., 1989, SINEs short interspersed repeated DNA elements in higher eucaryotes. In "Mobile DNA" (M. Howe and D. Berg, Eds.), p. 619. American Society for Microbiology Press, Washington D.C.

Deininger, P. L., and Batzer, M. A., 1993, Evolution of Retroposons. In "Evolutionary Biology, Volume 27" (M. K. Hect *et al.*,Eds), p. 157. Plenum Press, New York N.Y.

Deininger, P. L., and Daniels, G. R., 1986, The recent evolution of mammalian repetitive DNA elements, *Trends Genet.* **2:** 76.

Deininger, P. L., and Slagel, V. K., 1988, Recently amplified *Alu* family members share a common parental *Alu*sequence, *Mol. Cell. Biol.* **8:** 4566.

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

Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. and Schmid, C. W., 1981, Base sequence studies of 300 nucleotide renatured repeated human DNA clones, *J. Mol. Biol.* **151:** 17.

Economou, E. P., Bergen, A. W., Warren, A. C., and Antonarakis, S. E., 1990, The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome, *Proc. Natl. Acad. Sci. USA* **87:** 2951.

Edwards, M. C., and Gibbs, R. A., 1992, A human dimorphism resulting from loss of an *Alu*, *Genomics* **14:** 590.

Friezner Degen, S. J., Rajput, B., and Reich, E., 1986, The human tissue plasminogen activator gene, *J. Biol. Chem.* **261:** 6972.

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

Jagadeeswaran, P., Forget, B. G., and Weissman, S. M., 1981, Short interspersed repetitive DNA elements in eucaryotes: Transposable DNA elements generated by reverse transcription of RNA pol III transcripts?, *Cell* **26:** 141.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L., and Schmid, C. W., 1980, Ubiquitous interspersed repeated sequences in mammalian genomes, *Proc. Natl. Acad. Sci. USA* **77:** 1398.

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

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

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

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

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

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

Lisitsyn, N., Lisitsyn, N., and Wigler, M., 1993, Cloning the differences between two complex genomes, *Science* **259:** 946.

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

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

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

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

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R., 1987, Variable number of tandem repeat (VNTR) markers for human gene mapping, *Science* **235:** 1616.

Novick, G. E., Gonzalez, T., Garrison, J. Novick, C. C., Batzer, M. A., Deininger, P. L., and Herrera, R. J., 1993, The use of polymorphic *Alu* insertions in human DNA fingerprinting. In "DNA Fingerprinting: State of the Science (S. D. J. Pena *et al.*,Eds.), p. 283. Birkhauser Verlag Press, Basal, Switzerland.

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

Quentin, Y., 1988, The *Alu* family developed through successive waves of fixation closely connected with primate lineage history, *J. Mol. Evol.* **27:** 194.

Rogers, J., 1983, Retroposons defined, *Nature* **301:** 460.

Ryan, S. C., and Dugaiczyk, A., 1989, Newly arisen repeats in primate phylogeny, *Proc. Natl. Acad. Sci. USA* **86:** 9360.

Sainz, J., Pevny, L., Wu, Y., Cantor, C. R., and Smith, C. L., 1992, Distribution of interspersed repeats (*Alu*and *Kpn*) on *Not* I restriction fragments of human chromosome 21, *Proc. Natl. Acad. Sci. USA* **89:** 1080.

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

Sawada, I., Willard, C., Shen, C.-K. J., Chapman, B., Wilson, A. C., and Schmid, C. W., 1985, Evolution of *Alu* family repeats since the divergence of human and chimpanzee, *J. Mol. Evol.* **22:** 316.

Schmid, C. W., and Shen, C.-K. J., 1986, The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. In "Molecular Evolutionary Genetics" (R.J. MacIntyre Ed.), p. 323. Plenum Publishing, New York.

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

Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw Jr., H., and Deininger, P. L., 1987, Clustering and sub-family relationships of the *Alu* family in the human genome, *Mol. Biol. Evol.* **4:** 19.

Stoppa-Lyonnet, D., Carter, P. E., Meo, T., and Tosi, M., 1990, Clusters of intragenic *Alu* repeats predispose the human C1 inhibitor locus to deleterious rearrangements, *Proc. Natl. Acad. Sci. USA* **87:** 1551.

Ullu, E., and Tschudi, C., 1984, *Alu* sequences are processed 7SL RNA genes, *Nature* **312:** 171.

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

Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T.W., and Collins, F. S., 1991, A *de novo Alu* insertion results in neurofibromatosis type 1, *Nature* **353:** 864.

Weiner, A. M., Deininger, P. L., and Efstratiadis, A., 1986, The reverse flow of genetic information: pseudogenes and transposable elements derived from nonviral cellular RNA, *Annu. Rev. Biochem.* **55:** 631.

Willard, C., Nguyen, H. T., and Schmid, C. W., 1987, Existence of at least three distinct *Alu* subfamilies, *J. Mol. Evol.* **26:** 180.

ACKNOWLEDGMENTS

This work was supported by U. S. Department of Energy LDRD 94-LW-103 (MAB), Univ. of California Center for Molecular Cytometry CMC 07 (MAB), NIH grants RO1 HG00340 and HG 00770 (PLD), USPHS grant RR08205 (RJH) and NSF grant BNS 90-20567 (MS). PAI was partially funded by the ministry of health in Cyprus. Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy contract No.W-7405-ENG-48.

Figure 1. Structure of a Typical Human *Alu* Repeat.

The diagram shows a schematic depiction of a generic *Alu* repeat. Each *Alu* element is dimeric in structure with the right half containing an additional 31 bp insertion relative to the left half. The total length of each *Alu* sequence is approximately 300 bp. Each repeat also contains a middle A rich region and is flanked by short intact direct repeats (arrows). The left half of each sequence contains an RNA polymerase III promoter represented by the A and B.

Figure 2. Schematic of the *Alu* Repeat PCR Assay.

A schematic of the *Alu* insertion assay is shown at the top in part A. The *Alu* sequence is represented by the cross-hatched line. A chromosome is represented by the thick dark line, and the flanking unique PCR primers are denoted by the arrows. Amplification in this assay results in the production of a 100 bp DNA fragment, 400 bp DNA fragment, or both depending on the genotype of the sample being assayed. Samples that are homozygous for the *Alu* repeat will amplify a 400 bp DNA fragment (B-1). Samples that are heterozygous for the *Alu* insertion will amplify both 400 and 100 bp DNA fragments (B-2). Lastly, samples that are homozygous for the lack of the *Alu* sequence only amplify a 100 bp DNA fragment (B-3).

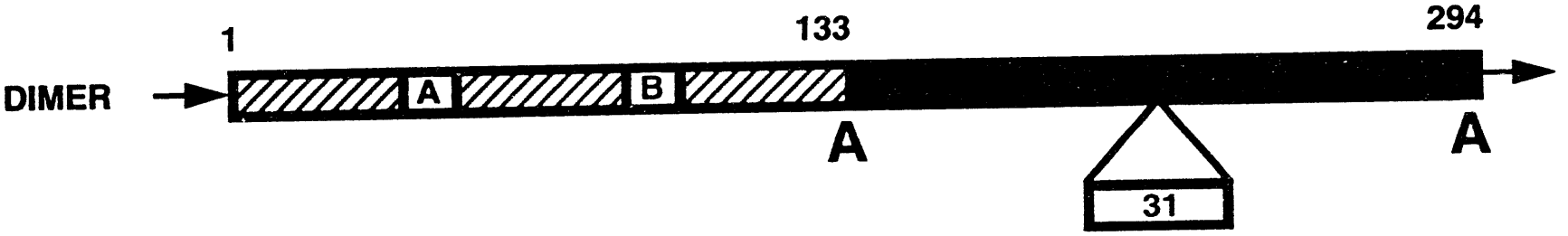
Figure 3. Representation of an *Alu* Insertion Assay.

The diagram depicts the results of an agarose gel chromatograph of PCR products generated through the amplification of DNA from the genomes of a number of human and non-human primates for TPA 25 *Alu* family member. Lane M is the marker ϕX174 *Hae* III digest, Lanes 1-3 depict the amplification products from the HeLa human (*Homo sapiens*) cell line, chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla*), respectively. Lanes 4-8 show the products derived from the amplification of an Asian (Lane 4), two Caucasians (Lanes 5 and 6), and two American Blacks (Lanes 7 and 8). The genotype of each individual is presented below the schematic of the gel with the location of the 400 and 100 bp DNA fragments indicated by arrows.

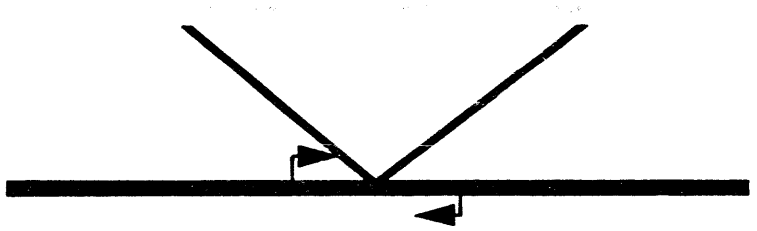
Table 1. Distribution of a Highly Dimorphic HS *Alu* Family Member [1]

	Subfamily member TPA 25	
Asians	+ +	12 (9.1) {0.57}
	+ -	8 (13.72)
	- -	8 (5.18) {0.43}
Indonesians	+ +	35 (28.85) {0.44}
	+ -	61 (73.42)
	- -	53 (46.73) {0.56}
Caucasians	+ +	16 (13.56) {0.64}
	+ -	10 (15.2)
	- -	7 (4.28) {0.36}
Greek Cypriots	+ +	15 (14.05) {0.53}
	+ -	23 (24.91)
	- -	12 (11.04) {0.47}
American Blacks	+ +	6 (5.6) {0.4}
	+ -	16 (16.8)
	- -	13 (12.6) {0.6}
African Blacks (Nigerians)	+ +	2 (3.15) {0.37}
	+ -	13 (10.72)
	- -	8 (9.13) {0.63}
African Blacks (Pygmies)	+ +	0 (1.66) {0.22}
	+ -	15 (11.71)
	- -	19 (20.63) {0.78}
Arhuaco	+ +	0 (0.29) {0.12}
	+ -	5 (4.22)
	- -	15 (15.49) {0.88}
Australian (Aborigines)	+ +	4 (2.09) {0.14}
	+ -	22 (25.85)
	- -	82 (80.06) {0.86}

[1] Genotypes, followed by expected numbers in () and allele frequency in { }.


1
133
294
DIMER
A
B
A
A
31

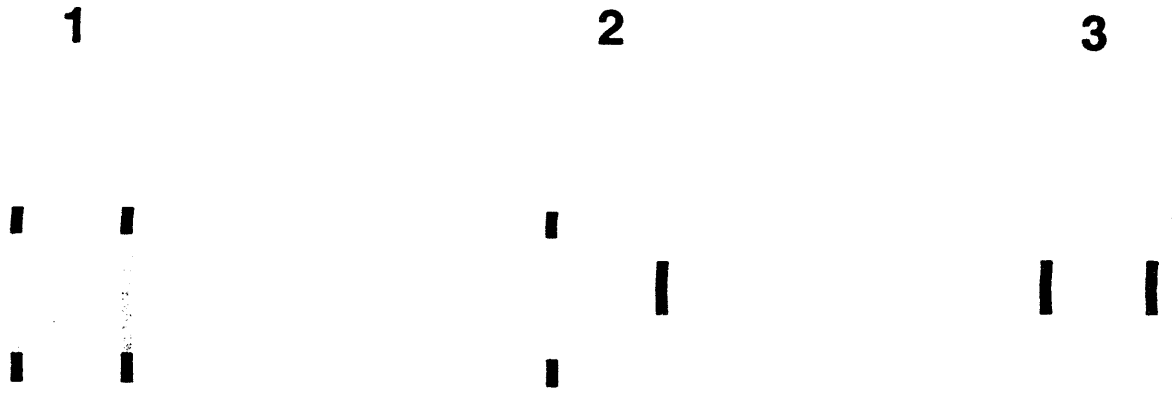
A



400 bp

100 bp

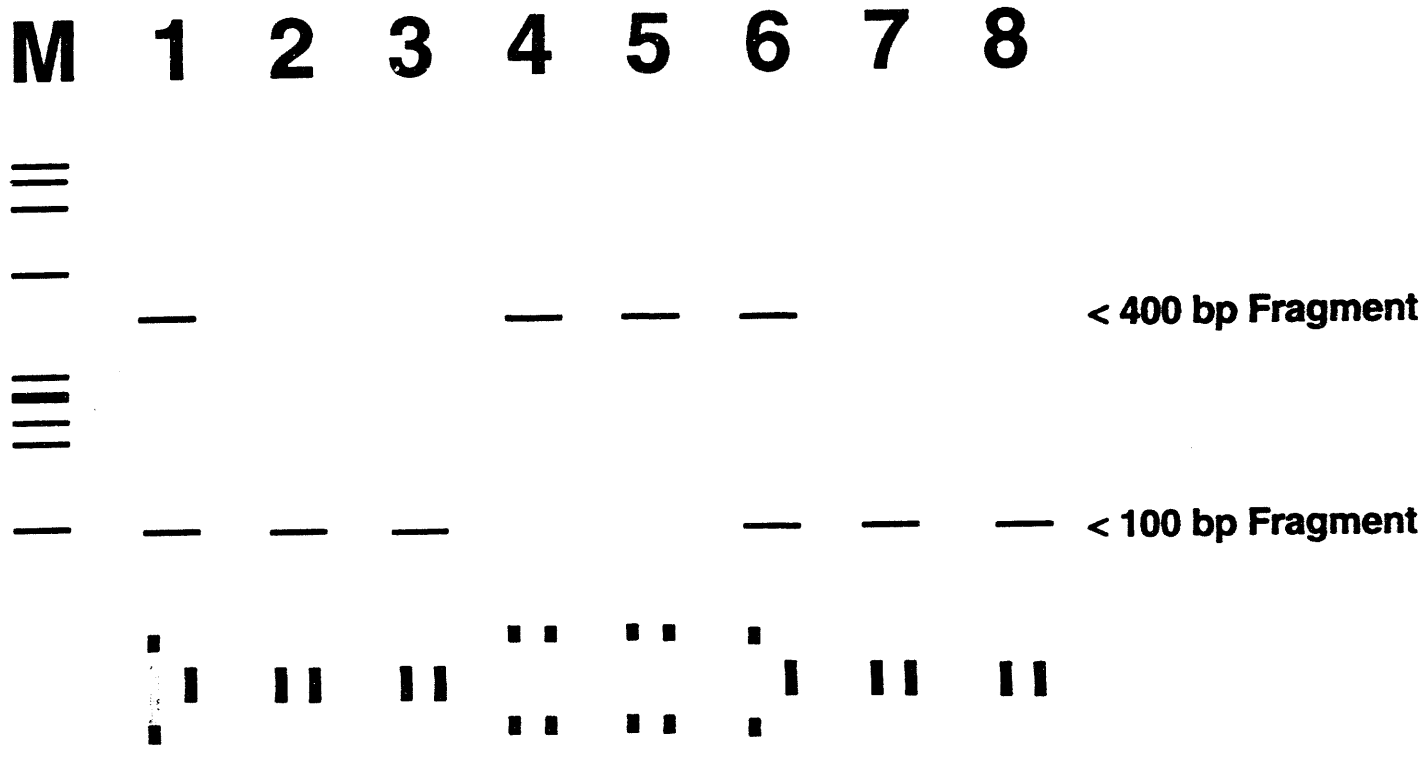
B



400 bp Fragment

400 & 100 bp Fragments

100 bp Fragment


M 1 2 3 4 5 6 7 8
< 400 bp Fragment
< 100 bp Fragment

DATE
FILMED
5/4/94
END