

## ORIGINAL INVESTIGATION

Sarah A. Tishkoff · Gualberto Ruano · Judith R. Kidd  
Kenneth K. Kidd

## Distribution and frequency of a polymorphic *Alu* insertion at the plasminogen activator locus in humans

Received: 10 July 1995 / Revised: 17 November 1995

**Abstract** We have investigated the frequency distribution, across a broad range of geographically dispersed populations, of alleles of the polymorphic *Alu* insertion that occurs within the 8th intron of the tissue plasminogen activator gene (PLAT). This *Alu* is a member of a recently derived subfamily of *Alu* elements that has been expanding during human evolution and continues to be transpositionally active. We used a “population tube” approach to screen 10 chromosomes from each of 19 human populations for presence or absence of this *Alu* in the PLAT locus and found that all tested populations are dimorphic for presence/absence of this insertion. We show that the previously published *EcoRI*, *HincII*, *PstI*, *TaqI*, and *XmnI* polymorphisms at the PLAT locus all result from insertion of this *Alu* and we use both restriction fragment length polymorphism and polymerase chain reaction analysis to examine the frequency of *Alu*(+) and *Alu*(-) alleles in a sample of 1003 individuals from 27 human populations and in 38 nonhuman primates. Nonhuman primates are monomorphic for the *Alu*(-) allele. Human populations differ substantially in allele frequency, and in several populations both alleles are common. Our results date the insertion event prior to the spread and diversification of modern humans.

### Introduction

Nearly one million members of the *Alu* family of short interspersed repeated DNA elements (SINEs) are distributed throughout the primate genome (for reviews see Schmid and Shen 1985; Weiner et al. 1986; Deininger 1989). These *Alu* elements share a 282-base pair (bp) consensus sequence, which is believed to be derived from the 7SL RNA gene (Ullu 1982). The DNA sequence of *Alu* elements shows a dimerized structure. The 5' half contains an RNA poly-

merase III promoter, which in some *Alu* elements can be active (Weiner et al. 1986), and the 3' half has a sequence similar to that of the 5' half but contains an additional 31 bp and is lacking an active promoter (Deininger et al. 1981; Deininger 1989). There is a short A-rich region in the center of the *Alu* element and a longer A-rich region at the 3' end which is more variable in length and sequence (Deininger et al. 1981; Orita et al. 1990). *Alu* elements are flanked by short direct repeats and are believed to transpose via an RNA intermediate (Rogers 1983).

The *Alu* family arose in the primate genome within the last 65 million years (Deininger and Daniels 1986). Analyses of many *Alu* sequences have led to the conclusion that *Alu* elements can be separated into at least three distinct subfamilies, each of which is the progeny of one or a few founder sequences (Willard et al. 1987; Quentin 1988; Slagel et al. 1987; Jurka and Smith 1988; Britten et al. 1988). Willard et al. (1987) named these three subfamilies diverged, major, and conserved. The different degrees of divergence of members of each subfamily from the *Alu* consensus indicate that the diverged subfamily is the most ancient and the conserved subfamily most recent. Two *Alus*, at the MLVI2 locus (Economou-Pachnis and Tsiichlis 1985) and the plasminogen activator (PLAT) locus (Friezner-Degen et al. 1986), share five sequence differences compared to the conserved subfamily consensus and have been proposed to be derived from one or more closely related source genes that diverged from the conserved subfamily and which comprise a fourth subfamily (Deininger and Slagel 1988; Schmid et al. 1989). This new subfamily was further characterized by Matera et al. (1990a,b) who named it predicted variant (PV) and by Batzer et al. (1990,1991) who introduced the designation human specific (HS). We will refer to this *Alu* subfamily as predicted variant because there is evidence that this *Alu* subfamily is not specific to humans and is present in the African apes as well (Matera et al. 1990a; LeeFlang et al. 1992).

Matera et al. (1990a) have shown that there has been an expansion of the PV *Alu* subfamily within humans following their divergence from African apes. Further evidence for a recent origin for some PV subfamily members

S. A. Tishkoff · G. Ruano · J. R. Kidd · K. K. Kidd (✉)  
Department of Genetics, Yale University, 333 Cedar Street,  
New Haven, CT 06510, USA  
Tel. +1-203-785-2654; Fax +1-203-785-6568

is that several of the known PV *Alus* are polymorphic for presence or absence in humans (Economou-Pachnis and Tsihchlis 1985; Antonarakis et al. 1988; Lim et al. 1986; Matera et al. 1990a; Batzer et al. 1991, 1994). The de novo germ-line insertions of a PV *Alu* into the NF1 gene of a patient with neurofibromatosis type 1 (Wallace et al. 1991) and into the factor IX gene of a patient with hemophilia B (Vidaud et al. 1993) document the ongoing transposition of PV *Alus* and their ability to disrupt normal gene function and cause disease (Wallace et al. 1991). One of the first identified polymorphic PV *Alu* elements was the TPA *Alu*, which is located within the 8th intron of the tissue plasminogen activator gene at the PLAT locus (Friezner-Degen et al. 1986). In this report we show that five different published restriction fragment length polymorphisms (RFLPs; based on five different restriction enzymes) recognize the *Alu* insertion polymorphism at the PLAT locus, and we examine the frequency distribution of the TPA *Alu* among various geographically dispersed human populations.

## Material and methods

### Amplification of the TPA *Alu*

Primers designed to flank the TPA *Alu* were generated from the published sequence (Friezner-Degen et al. 1986). With the position from Genbank accession number K03021 of the most 3' nucleotide presented in parentheses, the sequence of primer PLAT.A 5' to 3' is GTGAAAAGCAAGGTCTACCAG (28,628) and the sequence of primer PLAT.B 5' to 3' is GACACCGAGTTCATCTTGAC (29,193). The polymerase chain reaction (PCR) was performed in a 25- $\mu$ l volume with 10 pmol each of primers PLAT.A and PLAT.B, 200  $\mu$ M each dNTP, 50 mM KCl, 10 mM Tris.HCl (pH 8.4), 3 mM MgCl<sub>2</sub>, 0.625 units *Taq* DNA polymerase (Perkin-Elmer/Cetus), and 100 ng genomic DNA. The samples were subjected to 30 cycles consisting of 1 min at 94°C (denaturation), 1 min at 60°C (annealing), and 1 min at 72°C (extension) in a Perkin-Elmer model 480 or 9600 thermal cycler. A 10  $\mu$ l sample of this reaction was analyzed on a 2.5% agarose gel. Following electrophoresis the DNA was visualized with EtBr and UV light. Amplification produces a 570-bp fragment from chromosomes with the *Alu* insertion and a 260-bp fragment from those without.

### Southern blotting of TPA *Alu* amplification products

The genomic DNAs of several *Alu*(+) homozygous, *Alu*(-) homozygous, and *Alu*(+)/*Alu*(-) heterozygous individuals were amplified using primers PLAT.A and PLAT.B (flanking the TPA *Alu* insertion) and the products were analyzed on a 2.5% agarose gel. Following electrophoresis the DNA was transferred onto a nylon membrane using standard Southern blotting methods as described by Sambrook et al. (1989), prehybridized, and then hybridized to the human *Alu* clone CBE (kindly provided by Dr. I. Gonzalez); this clone is derived from human ribosomal spacer sequence and contains four *Alu* elements. The CBE clone was labeled with [5'- $\alpha$ -<sup>32</sup>P]dCTP by random oligonucleotide labeling (using Stratagene's Prime-It II kit) to a specific activity of at least  $1 \times 10^9$ . Hybridized filters were washed to a stringency of  $0.1 \times$  SSC at 65°C for 5 min and exposed to X-ray film for 1-5 days, depending on the specific activity of the probe.

### Population samples

Most of the populations used in this study have been previously described in other studies from or done in collaboration with our

laboratory (Bowcock et al. 1987, 1991; Kidd et al. 1991, 1993; Barr and Kidd 1993; Lichter et al. 1993; Lu et al. 1995). Those populations and samples not previously described by us include the Jemez Pueblo samples from New Mexico and the Southern Cheyenne samples from Western Oklahoma (both collected by Dr. D. Goldman of the National Institute of Alcohol Abuse and Alcoholism and described in Goldman et al. 1993); the Pima samples from Arizona (collected by Dr. W. Knowler from the National Institute of Diabetes and Digestive and Kidney Disease; see Knowler et al. 1978); the Papua New Guinea samples from the Goroka area (collected by Dr. F. Black and Dr. J. Yates from Yale University in collaboration with Dr. K. Bhatia of the Institute of Research in Papua New Guinea); the Micronesian samples of individuals from several different islands of Micronesia (collected by Dr. D. Upson from the University of Hawaii and Dr. S. Williams of Smith College); the Danish samples (random blood bank samples collected by Dr. Josef Parnas from the Institute of Preventive Medicine, Denmark); the Finnish samples (collected by Dr. Leena Peltonen from the National Public Health Institute, Finland); the Yakut sample of individuals who speak the Yakutian language from the Yakutian Autonomous Republic of Russia and the Adygei sample of individuals who speak the Adygeian language from north of the Caucasus mountains in the Krasnodar region of Russia (both collected by Drs. V. Galkina, E. Grigorenko, and M. Kadoshnikova with support from the Bristol Company in Russia). Cell lines of unrelated individuals in these populations were established using a modification of the protocol of Anderson and Gusella (1984). Cell lines of selected individuals in many of these populations are available through the Coriell Institute for Medical Research, Camden, N.J. DNA was purified from these cell lines using standard procedures (Sambrook et al. 1989). All blood samples were obtained with informed consent and genetic marker studies of these samples are carried out under a protocol approved by the Human Investigations Committee at Yale.

### Population tubes

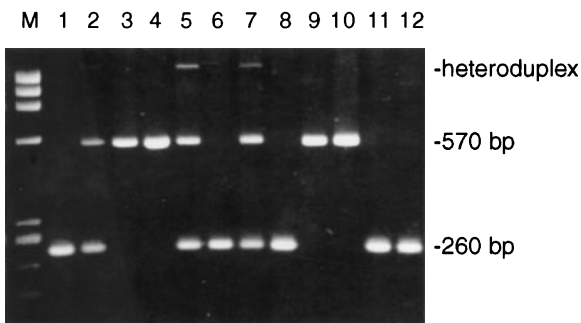
Equal concentrations of genomic DNA from five unrelated individuals from each population were combined into "poptubes" as described by Ruano et al. (1994). The populations included in our poptube survey were (1) Biaka Pygmies, (2) Mbuti Pygmies, (3) Yemenite Jews, (4) Druze, (5) Ethiopian Jews, (6) Roman Jews, (7) Sardinians, (8) Thoti Indians, (9) Kachari (Assam), (10) Cambodian, (11) Chinese, (12) Japanese, (13) Nasioi Melanesians, (14) mixed Europeans, (15) Moskoke, (16) Mayans, (17) Guahiba (Colombia), (18) Karitiana (Brazil), and (19) Rondonian Surui (Brazil). Since each population tube contains equal concentrations of the genomic DNA from five unrelated individuals from a particular population, this mixed DNA has 10 potentially different copies of each autosomal locus. Amplification with the PLAT.A and PLAT.B primers allows us to quickly screen for presence or absence of the *Alu* in multiple populations.

### RFLP analysis

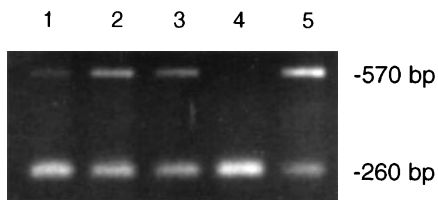
RFLPs were analyzed using standard Southern blotting procedures (Sambrook et al. 1989). 1- to 10- $\mu$ g of DNA from each individual was digested with *Eco*RI, *Hinc*II, *Pst*I, *Taq*I, or *Xmn*I restriction enzyme following the suppliers' protocols, size-fractionated by electrophoresis in an agarose gel, transferred by blotting onto a nylon membrane, prehybridized, and then hybridized with a PLAT *Alu*(-) probe. This probe was produced by amplification of DNA from an *Alu*(-) homozygous individual with the PLAT.A and PLAT.B primers; the 260-bp PCR product was run on a 2% agarose gel and isolated using DEAE-cellulose membrane as described by Sambrook et al. (1989). This product was oligolabeled with [5'- $\alpha$ -<sup>32</sup>P]dCTP as described above for the CBE *Alu* clone. Hybridized filters were washed to a stringency of  $0.1 \times$  SSC at 65°C and exposed to X-ray film for 1-5 days, depending on the specific activity of the probe.

## Results

Figure 1 shows the results of PCR amplification with primers PLAT.A and PLAT.B of 12 Cambodian samples. Heterozygotes and both types of homozygotes are clearly distinguishable. The products from several homozygous and heterozygous individuals were Southern blotted and hybridized to the *Alu*-specific CBE probe. This probe hybridized only to the 570-bp product, thereby confirming



**Fig. 1** PCR analysis of individuals for the TPA *Alu* insertion at the plasminogen activator (PLAT) locus. Ethidium bromide stained 2.5% agarose gel of PCR products from amplification of total genomic DNA using the PLAT.A and PLAT.B primers. Lane M *Hae*III-digested  $\phi$ X174 RF DNA marker, lanes 1–12 PCR product from 12 Cambodian individuals. Homozygotes for the *Alu*(+) allele produce a 570-bp fragment, homozygotes for the *Alu*(-) allele produce a 260-bp fragment, and heterozygotes produce fragments of both sizes in addition to a slowly migrating heteroduplex band formed by annealing of the *Alu*(+) and *Alu*(-) strands

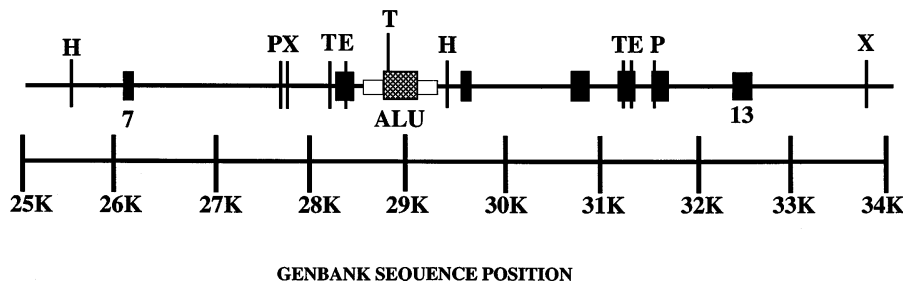


**Fig. 2** PCR products from amplification of “poptube” DNA samples with primers flanking the *Alu* insertion. Each tube contains the pooled genomic DNA from five unrelated individuals from a single population. Band intensities reflect the respective allele frequencies. Lane 1 Mbuti Pygmy; lane 2 Thoti; lane 3 Kachari (Assam); lane 4 Nasioi; lane 5 Guahiba

that this longer PCR product results from the insertion of an *Alu* element (data not shown).

Amplification with these primers was used to screen 19 human populations for presence of the TPA *Alu* at the PLAT locus using the population tube assay for detection of DNA sequence variation (Ruano and Kidd 1991; Ruano et al. 1994). Figure 2 shows the PCR amplification products from these poptubes for five of the human populations. All but 1 of the 19 pop tubes gave both 570-bp and 260-bp PCR products showing presence of both alleles. The pop tube for the Nasioi Melanesians yielded only the 260-bp PCR product indicating that all 10 alleles were *Alu*(-). When an additional 18 members of this population were screened individually, one heterozygous individual was detected; this was the only *Alu*(+) allele of a total of 46 Nasioi chromosomes examined. In addition to the human populations, we screened 24 chimpanzees, five gorillas, five orangutans, and four gibbons for the *Alu* insertion. We did not detect the presence of the TPA *Alu* in any of the nonhuman primates; all had amplification products of approximately 260 bp (data not shown).

Five RFLPs, based on five different enzymes, have been published for the PLAT locus: *Eco*RI (Benham et al. 1984), *Hinc*II (Donis-Keller et al. 1987), *Pst*I (Donis-Keller et al. 1987), *Taq*I (Ludwig et al. 1991), and *Xmn*I (Wohn et al. 1990). Because the reported allele sizes for four of these RFLPs (all but *Taq*I) differed by approximately 300 bp, we suspected they might all result from insertion of the TPA *Alu* into the PLAT gene. Restriction maps for those five enzymes of *Alu*(+) and *Alu*(-) PLAT chromosomes were generated from the published sequence (Genbank accession #K03021) for the region flanking the position of the *Alu* insertion (Fig. 3); the predicted fragment sizes are shown in Table 1. All agree well with the published descriptions of the RFLPs. Because of the presence of a *Taq*I restriction site within the TPA *Alu* element, digestion with *Taq*I of chromosomes containing the *Alu* insertion produces two smaller sized fragments rather than a single larger sized fragment (see Fig. 3 and Table 1). We used Southern blot analysis to confirm the identity of these RFLPs with the *Alu* insertion polymorphism. The 260-bp PCR product, which contains the DNA flanking the *Alu* insertion but not the *Alu* itself (see Fig. 3), was used as a probe of genomic DNA digested with each of the five en-



**Fig. 3** Restriction map of *Alu*(+) and *Alu*(-) chromosomes in the region of the *Alu* insertion polymorphism based on published sequence (Genbank accession #K03021). The polymorphic *Alu* element is located within the 8th intron of the PLAT gene, which spans sequence position 28,463–29,469. Exons 7–13 are repre-

sented by solid boxes. The TPA *Alu* is represented by a hatched box and the position of the 260-bp *Alu*(-) PCR probe is represented by the shaded bar. E *Eco*RI, H *Hinc*II, P *Pst*I, T *Taq*I, X *Xmn*I restriction sites

zymes of several *Alu*(+) and *Alu*(-) homozygous and several *Alu*(+)/*Alu*(-) heterozygous individuals (based on PCR typing as described); the autoradiographs showed fragments of sizes expected from the published sequence and phenotypes produced by the two methods were completely concordant. A comparison of phenotypes produced by PCR amplification with the primers flanking the *Alu* and by RFLP analysis of an *Eco*RI digest hybridized with the *Alu*(-) probe is shown in Fig. 4; similar results were obtained with *Hinc*II, *Pst*I, *Taq*I, and *Xmn*I digests (data not shown).

The frequency of the *Alu*(+) and *Alu*(-) alleles within several geographically distinct populations was determined by PCR amplification and by RFLP analysis (see Table 2). We PCR-amplified genomic DNA of 46 Ethiopian Jews, 8 Yemenite Jews, 34 Finns 19 Danes, 52 Adygei, 26 Micronesians, 23 Nasioi Melanesians, 20 Papua New Guineans, 38 Chinese, 38 Ami Taiwanese, 42 Atayal Taiwanese, 34 Yakuts, 31 Pima, 47 Karitiana, 44 Rondonian Surui, 48 Maya, and 15 Quechua. RFLP analysis using the 260-bp *Alu*(-) PCR product as a probe was used to screen Southern blots of genomic *Eco*RI digests of 65 Biaka Pygmies,

39 Mbuti Pygmies, 15 Ethiopian Jews, 23 Roman Jews, 33 Yemenite Jews, 51 Druze, 14 Assamese, 24 Cambodians, 42 Japanese, 44 Jemez Pueblo, 42 Cheyenne, and 67 Ticuna. All populations are polymorphic for the TPA *Alu* insertion and usually both alleles occur with frequencies greater than 25%. Two exceptions are the Nasioi and the Ticuna populations, which have allele frequencies for the *Alu* insertion chromosome of 0.02 and 0.98, respectively. In general, we see a pattern of increasing allele frequencies of the *Alu*(+) allele from Africa (average allele frequency of 0.25) east into Europe and Asia (average allele frequency of 0.52) and into the New World (average allele frequency of 0.63).

## Discussion

We have used "poptubes" containing mixed DNA from five individuals per population as a method for quickly screening populations for DNA sequence variation. We used this method to show that 18 of 19 geographically dispersed human populations are dimorphic for presence/absence of the TPA *Alu* insertion (the 19th population, Nasioi, was subsequently shown to have a low frequency of the TPA *Alu* insertion allele).

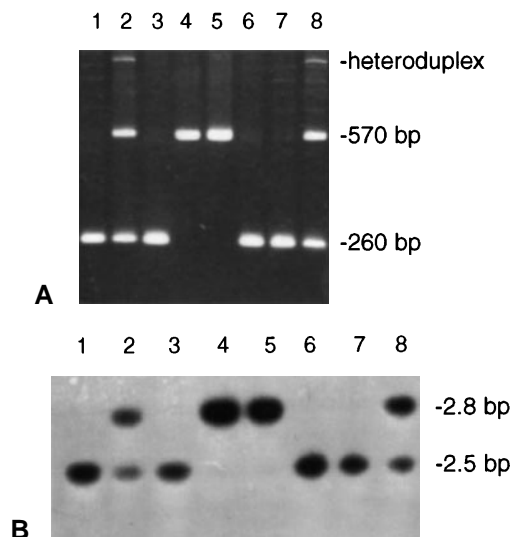
Our demonstration that the PLAT *Eco*RI, *Hinc*II, *Pst*I, *Taq*I, and *Xmn*I RFLPs are the same as the TPA *Alu* insertion polymorphism essentially eliminates the possibility of constructing haplotypes incorporating these RFLPs, which are, in fact, only different detection systems for the same polymorphism. This finding has important implications for the use of these RFLPs as markers for genetic mapping studies; clearly, typing these multiple RFLPs at the PLAT locus would merely provide duplicative data and would not be additively informative.

However, the fact that these five RFLPs all result from insertion of the TPA PV *Alu* allows the various datasets to be combined; it is clear that the TPA PV *Alu* at the PLAT locus is highly polymorphic in many human populations (Batzer and Deininger 1991; Batzer et al. 1991, 1994; Benham et al. 1984, 1985; Donis-Keller et al. 1987; Ludwig et al. 1991; Yang-Fang et al. 1986; Wohn et al. 1990; Perna et al. 1992). We have extended these observations to include 27 geographically dispersed populations. PCR and RFLP analysis of these populations shows that the PLAT locus is highly polymorphic in nearly all human populations regardless of location. The two most extreme frequencies (2% for the *Alu*(+) allele in Nasioi and 98% for the *Alu*(+) allele in Ticuna) occur in populations that are or have been small and isolated. Thus, random genetic drift may have a particularly strong effect in these populations. The pattern of increasing frequencies for the *Alu*(+) allele from Africa eastward across Asia and into the New World may represent a cumulative effect of genetic drift that occurred during the historical migration of humans out of Africa and around the world (Tishkoff et al. 1995).

The fact that the TPA PV *Alu* at PLAT is not present in any of the African apes examined implies that the *Alu* insertion event occurred after the human/great ape divergence 4–6 million years ago (Wilson and Sarich 1969;

**Table 1** Fragment sizes produced by digestion of *Alu*(+) and *Alu*(-) chromosomes with *Eco*RI, *Hinc*II, *Pst*I, *Taq*I, and *Xmn*I restriction enzymes. Sizes were determined by restriction mapping of the published sequence (Genbank accession #K03021)

| Restriction enzyme | <i>Alu</i> (+)    | <i>Alu</i> (-) |
|--------------------|-------------------|----------------|
| <i>Eco</i> RI      | 2840 bp           | 2530 bp        |
| <i>Hinc</i> II     | 3759 bp           | 3449 bp        |
| <i>Pst</i> I       | 3764 bp           | 3454 bp        |
| <i>Taq</i> I       | 2314 bp<br>677 bp | 2681 bp        |
| <i>Xmn</i> I       | 5954 bp           | 5644 bp        |



**Fig. 4A, B** Comparison of PCR amplification of the segment encompassing the polymorphic *Alu* and an *Eco*RI RFLP caused by presence/absence of the *Alu*. **A** 2.5% Agarose gel of PCR products of eight Cambodian individuals. **B** Autoradiograph of Southern blot of an *Eco*RI digest of genomic DNA from the same individuals as above using the 260-bp PCR fragment that flanks the site of the *Alu* insertion as a probe. Hybridization produces 2.5-kb and 2.8-kb fragments. All results are concordant with those of PCR

**Table 2** Allele frequencies for the *Alu* insertion polymorphism at the PLAT locus in humans. (*n* total number of individuals screened.)

| Population                     | N  | Observed genotypes |     |     | Allele frequencies <sup>b</sup> |                | $\chi^2$ |
|--------------------------------|----|--------------------|-----|-----|---------------------------------|----------------|----------|
|                                |    | +/+                | +/- | -/- | <i>Alu</i> (+)                  | <i>Alu</i> (-) |          |
| Ticuna <sup>+</sup>            | 67 | 57                 | 10  | 0   | 0.93 ± 0.02                     | 0.07 ± 0.02    | 0.13     |
| Karitiana*                     | 47 | 20                 | 26  | 1   | 0.70 ± 0.05                     | 0.30 ± 0.05    | 4.88     |
| Rondonian Surui*               | 44 | 12                 | 20  | 12  | 0.50 ± 0.05                     | 0.50 ± 0.05    | 0.36     |
| Quechua*                       | 15 | 6                  | 8   | 1   | 0.67 ± 0.09                     | 0.33 ± 0.09    | 0.60     |
| Maya*                          | 48 | 19                 | 24  | 5   | 0.65 ± 0.05                     | 0.35 ± 0.05    | 0.41     |
| Jemez Pueblo <sup>+</sup>      | 44 | 10                 | 25  | 9   | 0.51 ± 0.05                     | 0.49 ± 0.05    | 0.81     |
| Pima*                          | 31 | 11                 | 7   | 13  | 0.47 ± 0.06                     | 0.53 ± 0.06    | 8.52     |
| Cheyenne <sup>+</sup>          | 42 | 14                 | 22  | 6   | 0.59 ± 0.05                     | 0.41 ± 0.05    | 0.32     |
| Yakut*                         | 34 | 16                 | 9   | 9   | 0.60 ± 0.06                     | 0.40 ± 0.06    | 6.80     |
| Japanese <sup>+</sup>          | 42 | 13                 | 21  | 8   | 0.56 ± 0.05                     | 0.44 ± 0.05    | 0.06     |
| Ami*                           | 38 | 9                  | 22  | 7   | 0.53 ± 0.06                     | 0.47 ± 0.06    | 0.99     |
| Atayal*                        | 42 | 0                  | 23  | 19  | 0.27 ± 0.05                     | 0.73 ± 0.05    | 2.82     |
| Chinese*                       | 38 | 9                  | 16  | 13  | 0.45 ± 0.06                     | 0.55 ± 0.06    | 0.84     |
| Cambodian <sup>+</sup>         | 24 | 6                  | 8   | 10  | 0.42 ± 0.07                     | 0.58 ± 0.07    | 2.27     |
| Assamese <sup>+</sup>          | 14 | 7                  | 6   | 1   | 0.71 ± 0.09                     | 0.29 ± 0.09    | 0.04     |
| Micronesian*                   | 26 | 8                  | 13  | 5   | 0.56 ± 0.07                     | 0.44 ± 0.07    | 0.01     |
| New Guinean*                   | 20 | 1                  | 9   | 10  | 0.28 ± 0.07                     | 0.72 ± 0.07    | 0.33     |
| Nasioi Melanesian*             | 23 | 0                  | 1   | 22  | 0.02 ± 0.02                     | 0.98 ± 0.02    | 0.10     |
| Adygei*                        | 52 | 18                 | 24  | 10  | 0.58 ± 0.05                     | 0.42 ± 0.05    | 0.15     |
| Finnish*                       | 34 | 17                 | 10  | 7   | 0.65 ± 0.06                     | 0.35 ± 0.06    | 4.31     |
| Danish*                        | 19 | 12                 | 4   | 3   | 0.53 ± 0.08                     | 0.47 ± 0.08    | 0.35     |
| Roman Jewish <sup>+</sup>      | 23 | 5                  | 11  | 7   | 0.46 ± 0.07                     | 0.54 ± 0.07    | 0.03     |
| Yemenite Jewish <sup>++</sup>  | 41 | 5                  | 19  | 17  | 0.35 ± 0.05                     | 0.65 ± 0.05    | 0.01     |
| Druze <sup>+</sup>             | 51 | 15                 | 23  | 3   | 0.52 ± 0.05                     | 0.48 ± 0.05    | 0.48     |
| Ethiopian Jewish <sup>++</sup> | 61 | 6                  | 21  | 34  | 0.27 ± 0.04                     | 0.73 ± 0.04    | 0.99     |
| Biaka Pygmy <sup>+</sup>       | 65 | 4                  | 22  | 39  | 0.23 ± 0.04                     | 0.77 ± 0.04    | 0.14     |
| Mbuti Pygmy <sup>+</sup>       | 39 | 2                  | 16  | 21  | 0.26 ± 0.05                     | 0.74 ± 0.05    | 0.28     |

<sup>a</sup>Populations screened by PCR analysis are marked by \*.

<sup>b</sup>Populations screened by RFLP analysis after digestion with EcoRI are marked by +.

<sup>c</sup>Allele frequencies for a subset of the same Mbuti and Biaka Pygmy samples as reported here were previously reported in Bowcock et al. (1987,1991) and Batzer et al. (1994) and are concordant with the allele frequencies we obtained with our larger sample size. The genotype frequencies for all populations are close to predicted Hardy Weinberg frequencies; even the most deviant, the Pima, is not significant when corrected for the multiple tests performed

Hill 1994). However, the fact that all human populations studied to date have chromosomes containing the *Alu* insertion at the PLAT locus implies that it inserted before modern humans spread from Africa around the globe approximately 100,000 years ago (Tishkoff et al. 1996; Stringer and Andrews 1988). The high level of polymorphism of this *Alu* insertion in many geographically dispersed populations suggests that it is a relatively recent insertion event because there was not enough time for one allele to go to fixation before humans dispersed around the world. Thus, our results support the findings of Batzer et al. (1994), suggesting that the TPA *Alu* (and several other PV *Alus*) originated in Africa, but relatively recently in human evolution.

The *Alu* insertion polymorphism at the PLAT locus has the special characteristic that we know the direction of mutation: the *Alu*(+) (or insertion allele) is the derived allele. Moreover, the highly specific nature of this insertion essentially excludes the possibility of recurrent mutation. Thus, all *Alu*(+) alleles are descended from the single chromosome in which that mutation originally occurred. Because the insertion of a PV *Alu* into the PLAT locus is easily typed by PCR and Southern blotting techniques, because it is a unique and stable event that is highly polymorphic in humans, because there is significant allele frequency variation among human populations, and because the polarity of the mutation event is known, it is a particularly good genetic marker for population genetic studies

and for studies of isolation and genetic drift during recent human evolution. The large number of populations now typed for this polymorphism provide a global set of reference frequencies for comparison with future studies of individual populations.

**Acknowledgements** We would like to thank Svante Pääbo and Greg Matera for their helpful discussion and advice. This work was supported in part by NSF grant SBR 9408934 to J.R.K. The development of the cell line resource was also supported, in part, by NIH grants NIH39239 and AA09379 to K.K.K. and by a grant from the Sloan Foundation.

## References

- Anderson MA, Gusella J (1984) Use of cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. *In Vitro* 20:856-858
- Antonarakis SE, Chakravanti A, Halloran SL, Hudson RR, Feissee L, Karathanasis SK (1988) DNA polymorphism haplotypes of the human apolipoprotein APO A1-APOC3-APOA4 gene cluster. *Hum Genet* 80:265-273
- Barr CL, Kidd KK (1993) Population frequencies of the A1 allele at the dopamine D<sub>2</sub> receptor locus. *Biol Psychiatry* 34:204-209
- Batzer MA, Deininger PL (1991) A human-specific subfamily of *Alu* sequences. *Genomics* 9:481-487
- Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hopeens CL, Deininger PL (1990) Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Res* 18:6793-6798
- Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL (1991) Amplification dynamics of human-specific (HS) *Alu* family members. *Nucleic Acids Res* 19:3619-3623

- Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger P (1994) African origin of human-specific polymorphic *Alu* insertions. *Proc Natl Acad Sci USA* 91:12288–12292
- Benham FJ, Spurr N, Povey S, Brinton BT, Goodfellow PN, Solomon E, Harris TJR (1984) Assignment of tissue-type plasminogen activator to chromosome 8 in man and identification of a common restriction length polymorphism within the gene. *Mol Biol Med* 2:251–259
- Benham FJ, Spurr N, Povey S, Brinton BT, Solomon E, Goodfellow PN, Harris TJR (1985) Tissue-type plasminogen activator (PLAT) maps to chromosome 8 and there is a common restriction length polymorphism within the gene. *Cytogenet Cell Genet* 40:581
- Bowcock AM, Bucci C, Hebert JM, Kidd JR, Kidd KK, Friedlander J, Cavalli-Sforza LL (1987) Study of 47 DNA markers in five populations from four continents. *Gene Geogr* 1:47–64
- Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci USA* 88:839–843
- Britten RJ, Baron WF, Strout DB, Davidson EH (1988) Sources and evolution of human *Alu* repeated sequences. *Proc Natl Acad Sci USA* 85:4770–4774
- Deininger PL (1989) SINES, short interspersed repeated DNA elements in higher eukaryotes. In: Howe M, Berg D (eds) *Mobile DNA*. Washington DC, ASM Press, pp 619–636
- Deininger PL, Daniels GR (1986) The recent evolution of mammalian repetitive DNA elements. *Trends Genet* 2:76–80
- Deininger PL, Slagel VK (1988) Recently amplified *Alu* family members share a common parental *Alu* sequence. *Mol Cell Bio* 8:4566–4569
- Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J Mol Biol* 151:17–33
- Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES et al. (1987) A genetic linkage map of the human genome. *Cell* 51:319–337
- Economou-Pachnis A, Tschlis PN (1985) Insertion of an *Alu* SINE in the human homologue of the *MLVI-2* locus. *Nucleic Acids Res* 13:8379–8387
- Friezner Degen SJ, Rajput B, Reich E (1986) The human tissue plasminogen activator gene. *J Biol Chem* 261:6972–6985
- Goldman D, Brown GL, Albaugh B, Robin R, Goodson S, Trunzo M, Akhtar L, Lucas-Derse L, Long J, Linnoila M, Dean M (1993) *DRD2* dopamine receptor genotype, linkage disequilibrium, and alcoholism in American Indians and other populations. *Alcohol Clin Exp Res* 17:199–204
- Hill A (1994) Late Miocene and early Pliocene hominoids from Africa. In: Corruccine RS, Ciochon RL (eds) *Integrative paths to the past: paleoanthropological advances in honor of F. Clark Howell*. Prentice Hall Press, Englewood Cliffs, NJ, pp 123–145
- Jurka J, Smith T (1988) A fundamental division in the *Alu* family of repeated sequences. *Proc Natl Acad Sci USA* 85:4775–4778
- Kidd JR, Black FL, Weiss KM, Balazs I, Kidd KK (1991) Studies of three Amerindian populations using nuclear DNA polymorphisms. *Hum Biol* 63:775–794
- Knowler WC, Bennet PH, Hamman RF, Miller M (1978) Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. *Am J Epidemiol* 108:497–505
- Leefflang EP, Liu W, Hashimoto C, Choudary PV, Schmid CW (1992) Phylogenetic evidence for multiple *Alu* source genes. *J Mol Evol* 35:7–16
- Lichter JG, Barr CL, Kennedy JL, Van Tol HHM, Kidd KK, Livak KJ (1993) A hypervariable segment in the human dopamine receptor D4 (*DRD4*) gene. *Hum Mol Genet* 2:767–773
- Lim D, Coleman RT, Assmann G, Frossard PM (1986) Deletion of an *Alu* sequence in the 5' of the apolipoprotein AI gene associated with decreased serum HDL-cholesterol levels (abstract). *Am J Hum Genet* 39:621
- Lu RB, Ko HC, Chang FM, Castiglione CM, Schoolfield G, Pakstis AJ, Kidd JR, Kidd KK (1995) No association between alcoholism and multiple polymorphisms at the dopamine D2 receptor gene (*DRD2*) in three distinct Taiwanese populations. *Biol Psychiatry* (in press)
- Matera GA, Hellmann U, Schmid CW (1990a) A transpositionally and transcriptionally competent *Alu* subfamily. *Mol Cell Biol* 10:5424–5432
- Matera GA, Hellmann U, Hintz MF, Schmid CW (1990b) Recently transposed *Alu* repeats result from multiple source genes. *Nucleic Acids Res* 18:6019–6023
- Orita M, Sekiya T, Hayashi K (1990) DNA sequence polymorphisms in *Alu* repeats. *Genomics* 8:271–278
- Perna NT, Batzer MA, Deininger PL, Stoneking M (1992) *Alu* insertion polymorphism: a new type of marker for human population studies. *Hum Biol* 64:641–648
- Quentin Y (1988) The *Alu* family developed through successive waves of fixation closely connected with primate lineage history. *J Mol Evol* 27:194–202
- Rogers J (1983) Retroposons defined. *Nature* 301:460
- Ruano G, Kidd KK (1991) Coupled amplification and sequencing of genomic DNA. *Proc Natl Acad Sci USA* 88:2815–2819
- Ruano G, Deinard AS, Tishkoff S, Kidd KK (1994) Detection of DNA sequence variation via deliberate heteroduplex formation from genomic DNAs amplified en masse in “population tubes.” *PCR Methods Appl* 3:225–231
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Schmid CW, Shen CKJ (1985) The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. In: MacIntyre RJ (ed) *Molecular evolutionary genetics*. Plenum Press, New York, pp 323–358
- Schmid CW, Deka N, Matera AG (1989) Repetitive human DNA: the shape of things to come. In: Adolph KW (ed) *Chromosomes: eukaryotic, prokaryotic and viral*, vol 1. CRC Press, Boca Raton, Fla, pp 3–29
- Slagel V, Flemington E, Traina-Dorge V, Bradshaw H, Deininger PL (1987) Clustering and subfamily relationships of the *Alu* family in the human genome. *Mol Biol Evol* 4:19–29
- Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. *Science* 239:1263–1268
- Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Cheung K, Kidd JR, Bonn -Tamir B, Santachiara-Benerecetti AS, Moral P, Watson E, Krings M, P  bo S, Risch N, Jenkins T, Kidd KK (1996) Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. *Science* 271:1380–1387
- Ullu E, Murphy S, Melli M (1982) Human 7S RNA consists of a 140 nucleotide middle repetitive sequence inserted in an *Alu* sequence. *Cell* 29:195–202
- Vidaud D, Vidaud M, Bahnak BR, Siguret V, Gispert Sanchez S, Laurian Y, Meyer D, Goossens M, Lavergne JM (1993) Haemophilia 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* 1:30–36
- Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS (1991) A de novo *Alu* insertion results in neurofibromatosis type 1. *Nature* 353:864–866
- Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem* 55:631–661
- Willard C, Nguyen HT, Schmid CW (1987) Existence of at least three distinct *Alu* subfamilies. *J Mol Evol* 26:180–186
- Wilson AC, Sarich VM (1969) A molecular time scale for human evolution. *Proc Natl Acad Sci USA* 63:1088–1093
- Yang-Fang TL, Opendakker G, Volckaert G, Francke U (1986) Human tissue-type plasminogen activator gene located near chromosomal breakpoint in myeloproliferative disorder. *Am J Hum Genet* 39:79–87