Alu Insertion Polymorphism: A New Type of Marker for Human Population Studies

NICOLE T. PERNA,[1,4] MARK A. BATZER,[2,5] PRESCOTT L. DEININGER,[2,3] AND MARK STONEKING[1]

Abstract A PCR-based method was used to screen 462 individuals from Japan, Papua New Guinea, Indonesia, and Australia for an *Alu* family insertion polymorphism. The frequency of this *Alu* insertion shows significant heterogeneity among island subgroups of the Indonesian sample and between the Japanese-Indonesian populations and the Australian–New Guinean populations. The simple, rapid PCR-based screening technique and the significant frequency differences among populations demonstrate that *Alu* insertion polymorphisms are potentially valuable markers for studies of the evolutionary history and migration patterns of modern humans.

DNA polymorphisms provide a rich source of information on the genetic structure and evolutionary history of human populations (Bowcock et al. 1987, 1991; Cann et al. 1987; Long et al. 1990; Chen et al. 1990; Stoneking et al. 1990; Vigilant et al. 1991). The introduction of the polymerase chain reaction (PCR) for in vitro amplification of specific DNA sequences (Saiki et al. 1985, 1988) has enabled new, rapid approaches for screening DNA polymorphisms. A significant advantage of PCR-based methods is that extremely small amounts of DNA are sufficient for analysis, such as the amount found in a single plucked hair bulb [e.g., Vigilant et al. (1989)]. By contrast, traditional methods of screening for DNA polymorphisms require much larger amounts of DNA, so that in many cases cell lines must be established [e.g., Bowcock et al. (1987, 1991)].

[1]Department of Anthropology, Pennsylvania State University, University Park, PA 16802.
[2]Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, LA 70112.
[3]Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121.
[4]Current address: Program in Genetics, University of New Hampshire, Spaulding Life Sciences, Durham, NH 03824.
[5]Current address: Human Genome Center, Biomedical Sciences Division, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550.
Address correspondence to M. Stoneking.

Human Biology, October 1992, Vol. 64, No. 5, pp. 641–648.

KEY WORDS: ALU INTERSPERSED REPEATS, PCR, DNA POLYMORPHISM, GENETIC VARIATION

However, once a PCR product has been obtained, it must be further manipulated to determine the genotype of an individual at a marker locus. These manipulations can involve several steps of varying complexity, such as digestion with restriction enzymes (Schurr et al. 1990), immobilization and hybridization with oligonucleotide probes (Helmuth et al. 1990; Schurr et al. 1990; Stoneking et al. 1991), or direct sequencing (Vigilant et al. 1989, 1991). Here, we describe an application of a new type of DNA polymorphism (*Alu* insertion polymorphism) to studies of population variation in which genotypes can be readily determined simply by means of agarose gel electrophoresis of the PCR product.

The *Alu* family of short, interspersed repeats is found in all primate genomes, with about 500,000 members distributed within the human genome (Deininger 1989). *Alu* elements have been inserted into new locations in primate genomes throughout evolutionary history (Slagel et al. 1987). Once an *Alu* element is inserted at a specific location, it appears not to be subject to loss or rearrangement (Deininger 1989); *Alu* elements are thus stable genetic markers.

Recently, Batzer et al. (1991) reported that a human-specific *Alu* element within the tissue plasminogen activator gene is dimorphic; that is, the *Alu* element (designated TPA-25) is not present in all individuals examined. Batzer et al. devised a simple PCR-based screen for the TPA-25 element that uses oligonucleotide primers, which flank the insertion location, to amplify a 400-bp fragment if TPA-25 is present and a 100-bp fragment if TPA-25 is absent. Agarose gel electrophoresis of the PCR products is thus sufficient to distinguish among homozygotes for the presence of TPA-25 (400-bp product only), homozygotes for the absence of TPA-25 (100-bp product only), and heterozygotes (both the 400-bp and the 100-bp products).

Batzer et al. (1991) screened 79 individuals with this method and found that the frequency of the TPA-25 insertion varied from approximately 65% among Asians and Caucasians to approximately 40% among African and American blacks. These results suggested that the TPA-25 insertion polymorphism might be useful for studies of genetic variation in populations of anthropological interest. In this report we use the PCR-based screen to determine genotypes for the TPA-25 polymorphism for an additional 462 individuals from Japan, Indonesia, Australia, and Papua New Guinea. The rapidity and ease of the technique and the significant heterogeneity we find in the frequency of the TPA-25 element among these Pacific populations demonstrate the utility of this *Alu* insertion polymorphism for population genetic studies.

Materials and Methods

DNA Samples. A total of 462 individuals were analyzed. There were 142 samples from 6 Indonesian islands: Ternate and Hiri (part of the

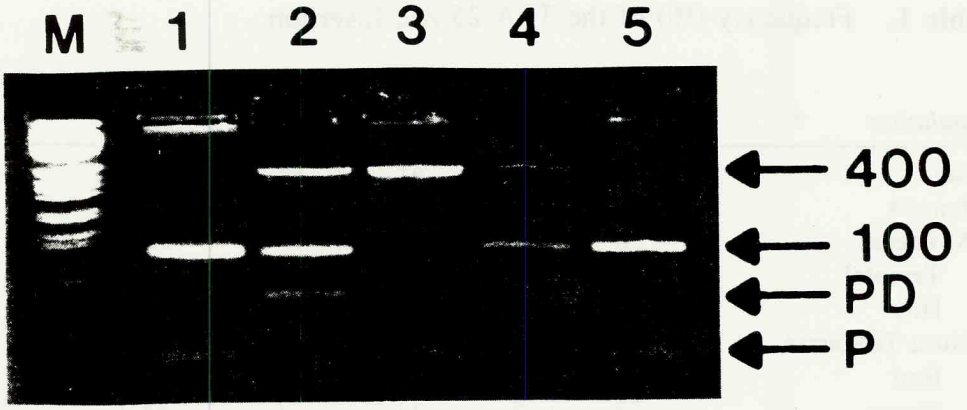

Figure 1. Agarose gel electrophoresis, illustrating the TPA *Alu* insertion banding patterns. Lanes: M, molecular size marker; 1 and 5, −/− homozygotes; 2 and 4, +/− heterozygotes; 3, +/+ homozygote. Arrows indicate bands corresponding to the 400-bp product, the 100-bp product, the primer-dimer (PD), and the residual primers (P).

Moluccas) and Alor, Flores, Roti, and Timor [part of the Nusa Tengarra (Lesser Sunda Islands)]. For further information see Stoneking et al. (1992). There were 116 samples from Papua New Guinea, including 68 samples from two highland localities (the Eastern and the Southern Highlands) and 48 samples from various coastal localities. For further information see Stoneking et al. (1990). There were 96 samples from four Australian villages (Balgo Mission, Turkey Creek, Christmas Creek, and Looma) along the northern edge of the Great Sandy Desert, collected by N. Kretchmer (University of California, Berkeley) as part of a nutritional study. An additional 22 samples came from various hospitals in Broome, Derby, Perth, and Alice Springs. There were 86 samples from Japan; they have been described previously by Helmuth et al. (1990).

PCR Amplification. Amplifications were performed in 50 μl reactions with 100–500 ng of target DNA, 400 ng of each primer, 100 μ*M* of each deoxynucleotide triphosphate, and 1 unit of *Taq* DNA polymerase (Promega). Each sample underwent 30 amplification cycles consisting of denaturation for 1 min at 94°C, annealing for 1 min at 58°C, and extension for 1 min at 74°C. The samples were visualized by UV fluorescence of 20 μl of PCR product fractionated on a 2% agarose gel containing 0.6 μg/ml ethidium bromide.

Results

The three possible phenotypes resulting from the TPA-25 insertion polymorphism are shown in Figure 1. The presence of the TPA-25 ele-

Table 1. Frequency (%) of the TPA-25 *Alu* Insertion

Population	*n*	*No. +/+*	*No. +/−*	*No. −/−*	*Frequency of +*	*Frequency of −*
Japan	86	29	41	16	57.6	42.4
Indonesia	142	34	58	50	44.4	55.6
Moluccas	50	18	19	13	55.0	45.0
Ternate[a]	25	11	7	7	58.0	42.0
Hiri	25	7	12	6	52.0	48.0
Nusa Tengarras	92	16	39	37	38.6	61.4
Roti	24	6	11	7	47.9	52.1
Timor	23	5	11	7	45.7	54.3
Alor	23	3	8	12	30.4	69.6
Flores	22	2	9	11	29.5	70.5
Papua New Guinea	116	5	27	84	15.9	84.1
Highlands	68	3	16	49	16.2	83.8
Eastern Highlands	29	1	10	18	20.7	79.3
Southern Highlands	39	2	6	31	12.8	87.2
Coast	48	2	11	35	15.6	84.4
Australia	118	4	24	90	13.6	86.4
Balgo Mission	25	0	8	17	16.0	84.0
Looma	17	1	3	13	14.7	85.3
Turkey Creek	24	1	4	19	12.5	87.5
Christmas Creek	30	1	4	25	10.0	90.0
Broome-Derby	22	1	5	16	15.9	84.1

a. Deviates from Hardy-Weinberg equilibrium proportions ($\chi^2 = 4.52$, d.f. = 1, $p < 0.05$).

ment results in amplification of a 400-bp fragment, whereas the absence of the insertion results in a 100-bp fragment. Genotypes can thus be readily and simply determined by means of gel electrophoresis of the PCR products; single 400-bp and 100-bp fragments indicate homozygotes for the presence of the insertion (+/+) and the absence of the insertion (−/−), respectively, and heterozygotes (+/−) possess both fragments (Figure 1).

A total of 462 individuals from 4 populations were screened for the TPA-25 element. The observed number of each genotype and the resulting allele frequencies are given in Table 1. The frequency of the TPA-25 element varied from 58% in Japan to 14% in Australia. Tests for the goodness of fit of observed genotype proportions to Hardy-Weinberg equilibrium proportions revealed that all subpopulations except one (Ternate) fit Hardy-Weinberg expectations.

The method of unplanned tests (Sokal and Rohlf 1981, p. 728) was used to assess heterogeneity in the frequency of the TPA-25 element among various subsets of the populations. Analysis of the six Indonesian subpopulations revealed significant heterogeneity, with four islands (Ternate, Hiri, Roti, and Timor) forming one homogeneous group that dif-

fered significantly from a homogeneous group consisting of the remaining two islands (Alor and Flores). The frequency of the TPA-25 element in the first group varied from 46% to 58%, whereas in the second group it was about 30%. This significant break in the frequency of the *Alu* insertion does not parallel the geographic distribution of these islands, as two of the subpopulations in the first group (Roti and Timor) are part of the same island chain as the second group.

Similar comparisons of the Papua New Guinea subpopulations revealed no significant heterogeneity between the two highland subpopulations nor between the highland and coastal subpopulations. There was also no significant heterogeneity among the five Australian subpopulations. Finally, analysis of the four main populations revealed that Japan and Indonesia formed one homogeneous group (frequency of the TPA-25 element = 44–58%) that differed significantly from a second homogeneous group consisting of Australia and Papua New Guinea (frequency of the TPA-25 element = 14–16%).

Discussion

Agarose gel electrophoresis of PCR products provides a rapid and efficient means of determining genotypes for the TPA-25 *Alu* insertion polymorphism. This is because of the large size difference between the two alleles. PCR-based techniques are also available for other loci in which alleles differ in size, such as variable number of tandem repeat (VNTR) and microsatellite loci (Jeffreys et al. 1988; Edwards et al. 1992). However, allelic size differences for VNTR and microsatellite loci are much smaller, usually requiring polyacrylamide gel electrophoresis and radiolabeling of the PCR product to accurately size alleles. Furthermore, these loci tend to be highly mutable and thus may not be stable markers for assessing intercontinental population relationships. By contrast, the TPA-25 *Alu* insertion probably is a stable marker, because it is likely that insertion occurred once during human evolution with no subsequent deletions of the element (Batzer et al. 1991).

The TPA-25 *Alu* marker was polymorphic in all populations studied, with the frequency of the insertion ranging from 13.6% to 58.0%. A significant deviation from Hardy-Weinberg equilibrium expectations was observed in one of the 15 subpopulations. Several explanations can be postulated for this deviation in the Ternate sample, which is due to a deficit of heterozygotes. One possibility is that the samples were not typed correctly. We have occasionally observed preferential amplification of the shorter fragment, which could lead to misclassification. However, increasing the amount of primer compensates for this problem, and all samples in which this might have occurred were reanalyzed with higher

primer concentrations. Another explanation is that the Ternate sample reflects a true heterozygote deficiency on the island, possibly because of inbreeding. Although all the Ternate samples came from one village on the island, the island is relatively accessible and there is mate exchange with other villages and islands. A third possibility is that the observed deviation is a result of sampling and reflects that 1 out of 20 comparisons are expected to be statistically significant at the $p = 0.05$ level. Because 15 independent comparisons to Hardy-Weinberg expectations were done, we consider this last explanation to be the most likely one.

Batzer et al. (1991) surveyed 79 individuals for the TPA-25 insertion and found that the frequency of the insertion was 66% in Asians, 63% in Caucasians, 38% in American blacks, and 42% in African blacks. By surveying additional populations of anthropological interest, we have confirmed the utility of the *Alu* insertion marker for population genetic studies. In particular, the frequency of the TPA-25 element in the present Japanese sample is 58%, which is similar to the frequency of 66% observed by Batzer et al. (1991) for a general Asian population. Also, the frequency of the TPA-25 element in the Papua New Guinea and Australian populations is nearly identical (14–16%) and much lower than in any other population surveyed. This similarity of the frequency of the insertion in Papua New Guinea and Australia (and their difference from the other populations sampled) would seem to reflect the known shared ancestry of these populations and the probable founder and drift effects associated with the colonization of New Guinea and Australia (Hill and Serjeantson 1989).

To further measure the utility of the *Alu* marker for population genetic studies, we calculated the F_{ST} value (Weir 1990) for the four major populations (Japan, Indonesia, Australia, and Papua New Guinea) and obtained a value of 0.14. By comparison, Bowcock et al. (1987) found in their survey of worldwide human populations that 27 out of 42 biallelic DNA polymorphisms had F_{ST} values less than 0.14. Although the comparison between their study and the present study is not strictly accurate, because different populations were examined, this does illustrate that for a biallelic marker there is appreciable interpopulation differentiation.

The results presented here are part of an ongoing study of autosomal and mitochondrial DNA variation in Pacific populations (Helmuth et al. 1990; Stoneking et al. 1990, 1991, 1992). The TPA-25 *Alu* insertion is only one locus, but there are interesting differences between the pattern of variation for this locus and the pattern of mtDNA variation in the Papua New Guinea and Indonesian populations. There are significant differences between highland and coastal Papua New Guinea populations with respect to mtDNA (Stoneking et al. 1990) but not with respect to the *Alu* insertion. Conversely, there are significant differences among the Indonesian populations with respect to the *Alu* insertion but

not with respect to the frequency of an mtDNA marker (Stoneking et al. 1992). These patterns point to the need for screening both mitochondrial and autosomal DNA variation to fully comprehend the genetic structure of human populations.

In conclusion, the rapidity and ease of the PCR-based assay for the TPA-25 *Alu* insertion polymorphism establishes the *Alu* insertion as a useful marker for other studies of molecular genetic variation. It is estimated that there may be as many as 400 *Alu* insertion polymorphisms in the human genome (Batzer et al. 1991). Several additional *Alu* insertion polymorphisms are currently being characterized, and it is hoped that eventually PCR-based screens will be available for a set of unlinked *Alu* insertion polymorphisms as a means of rapidly surveying genetic variation in any human population.

Acknowledgments We thank K. Bhatia, R. Cann, N. Kretchmer, T. Sasazuki, L. Schumacher, and A.S.M. Sofro for samples and S. Sherry and L. Vigilant for comments on this manuscript. This work was supported by the National Science Foundation under grant BNS 90-20567 to M. Stoneking and the US Public Health Service under grant HG00340 to P.L. Deininger.

Received 18 February 1992; revision received 23 March 1992.

Literature Cited

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

Bowcock, A.M., C. Bucci, J.M. Hebert, J.R. Kidd, K.K. Kidd, J.S. Friedlaender, and L.L. Cavalli-Sforza. 1987. Study of 47 DNA markers in five populations from four continents. *Gene Geog.* 1:47–64.

Bowcock, A.M., J.R. Kidd, J.L. Mountain, J.M. Hebert, L. Carotenuto, K.K. Kidd, and L.L. Cavalli-Sforza. 1991. Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* 88:839–843.

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

Chen, L.Z., S. Easteal, P.G. Board, and R.L. Kirk. 1990. Evolution of β-globin haplotypes in human populations. *Molec. Biol. Evol.* 7:423–437.

Deininger, P.L. 1989. SINEs: Short interspersed repeated DNA elements in higher eucaryotes. In *Mobile DNA*, M. Howe and D. Berg, eds. Washington, DC: ASM Press, 619–636.

Edwards, A., H.A. Hammond, L. Jin, C.T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241–253.

Helmuth, R., N. Fildes, E. Blake, M.C. Luce, J. Chimera, R. Madej, C. Gorodezky, M. Stoneking, N. Schmill, W. Klitz, R. Higuchi, and H.A. Erlich. 1990. HLA-DQα allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. *Am. J. Hum. Genet.* 47:515–523.

Hill, A.V.S., and S.W. Serjeantson, eds. 1989. *The Colonization of the Pacific: A Genetic Trail*. Oxford: Oxford University Press.

Jeffreys, A.J., V. Wilson, R. Neumann, and J. Keyte. 1988. Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. *Nucleic Acids Res.* 16:10953–10971.

Long, J.C., A. Chakravarti, C.D. Boehn, S. Antonarakis, and H.H. Kazazian. 1990. Phylogeny of human β-globin haplotypes and its implications for recent human evolution. *Am. J. Phys. Anthropol.* 81:113–130.

Saiki, R.K., S. Scharf, F. Faloona, K. Mullis, G. Horn, H.A. Erlich, and N.A. Arnheim. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–1354.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S. Scharf, R.H. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491.

Schurr, T.G., S.W. Ballinger, Y.Y. Gan, J.A. Hodge, D.A. Merriwether, D.N. Lawrence, W.C. Knowler, K.M. Weiss, and D.C. Wallace. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet.* 46:613–623.

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

Sokal, R.R., and F.J. Rohlf. 1981. *Biometry*. New York: W.H. Freeman.

Stoneking, M., A.S.M. Sofro, and A.C. Wilson. 1992. Implications of a mitochondrial marker for the colonization of the Pacific. In *Human Evolution in the Pacific Region*, C.K. Ho, G. Krantz, and M. Stoneking, eds. Pullman, WA: Washington State University Press (in press).

Stoneking, M., L.B. Jorde, K. Bhatia, and A.C. Wilson. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733.

Stoneking, M., D. Hedgecock, R.G. Higuchi, L. Vigilant, and H.A. Erlich. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.* 48:370–382.

Vigilant, L., R. Pennington, H. Harpending, T.D. Kocher, and A.C. Wilson. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci. USA* 86:9350–9354.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A.C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.

Weir, B. 1990. Intraspecific differentiation. In *Molecular Systematics*, D.M. Hillis and C. Moritz, eds. Sunderland, MA: Sinauer Associates, 373–410.