Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 12288–12292, December 1994
Evolution

African origin of human-specific polymorphic *Alu* insertions

MARK A. BATZER*†‡, MARK STONEKING†§, MICHELLE ALEGRIA-HARTMAN*, HERNAN BAZAN¶, DAVID H. KASS¶, TAMIM H. SHAIKH¶, GABRIEL E. NOVICK‖, PANAYIOTIS A. IOANNOU**, W. DOUGLAS SCHEER††, RENE J. HERRERA‖, AND PRESCOTT L. DEININGER¶‡‡

*Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551; §Department of Anthropology, Pennsylvania State University, University Park, PA 16802; Departments of ¶Biochemistry and Molecular Biology and ††Pathology, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, LA 70112; ‖Department of Biological Sciences, Florida International University, University Park Campus, Miami, FL 33199; **The Cyprus Institute of Neurology and Genetics, P.O. Box 3462, Nicosia, Cyprus; and ‡‡Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121

Communicated by Bruce Wallace, August 10, 1994

ABSTRACT ***Alu* elements are a family of interspersed repeats that have mobilized throughout primate genomes by retroposition from a few "master" genes. Among the 500,000 *Alu* elements in the human genome are members of the human-specific subfamily that are not fixed in the human species; that is, not all chromosomes carry an *Alu* element at a particular locus. Four such polymorphic human-specific *Alu* insertions were analyzed by a rapid, PCR-based assay that uses primers that flank the insertion point to determine genotypes based on the presence or absence of the *Alu* element. These four polymorphic *Alu* insertions were shown to be absent from the genomes of a number of nonhuman primates, consistent with their arising as human genetic polymorphisms sometime after the human/African ape divergence. Analysis of 664 unrelated individuals from 16 population groups from around the world revealed substantial levels of variation within population groups and significant genetic differentiation among groups. No significant associations were found among the four loci, consistent with their location on different chromosomes. A maximum-likelihood tree of population relationships showed four major groupings consisting of Africa, Europe, Asia/Americas, and Australia/New Guinea, which is concordant with similar trees based on other loci. A particularly useful feature of the polymorphic *Alu* insertions is that the ancestral state is known to be the absence of the *Alu* element, and the presence of the *Alu* element at a particular chromosomal site reflects a single, unique event in human evolution. A hypothetical ancestral group can then be included in the tree analysis, with the frequency of each insertion set to zero. The ancestral group connected to the maximum-likelihood tree within the African branch, which suggests an African origin of these polymorphic *Alu* insertions. These data are concordant with other diverse data sets, which lends further support to the recent African origin hypothesis for modern humans. Polymorphic *Alu* insertions represent a source of genetic variation for studying human population structure and evolution.**

The *Alu* family of short interspersed repetitive DNA elements is distributed throughout primate genomes (recently reviewed in refs. 1 and 2). *Alu* repeats represent a highly successful class of mobile genetic elements; they have amplified in the last 65 million years to a copy number in excess of 500,000 within the human genome. *Alu* sequences were ancestrally derived from the 7SL RNA gene and are thought to mobilize in a process termed retroposition. The vast majority of the *Alu* elements located within the human genome are transcriptionally and presumably transpositionally silent. Once inserted at specific chromosomal locations, most *Alu* elements do not appear to be subject to loss or rearrangement, making them stable genetic markers.

The *Alu* sequences located within primate genomes may be subdivided into groups of related subfamily members that share common diagnostic nucleotide substitutions (3, 4). One of the most recently formed groups of *Alu* sequences within the human genome has been termed human-specific (HS) (5, 6), or predicted variant (7, 8). There are an estimated 500–2000 HS *Alu* elements (5–8), which are mostly (6, 9) but not exclusively (10) restricted to the human genome.

Some HS *Alu* elements have retroposed so recently that they have not fixed in the human species; that is, not all chromosomes carry an *Alu* element at a specific locus (6, 9). There are two reasons why these polymorphic *Alu* insertions should be particularly useful for population genetic studies. First, since the probability of independent retroposition at the same exact chromosomal site is virtually nil (6), all loci carrying a particular polymorphic *Alu* insertion are derived from a unique event and hence are identical by descent. Polymorphic *Alu* insertions should thus more accurately reflect population relationships than markers [such as restriction fragment length polymorphism (RFLP), variable numbers of tandem repeats and microsatellite loci] in which the sharing of the same allele by two individuals may reflect chance identity by state (i.e., independent mutations). Second, the ancestral state for polymorphic *Alu* insertions can be reasonably inferred to be the absence of the insertion, and the direction of mutational change is therefore the gain of the *Alu* element at a particular locus. Knowing the ancestral state and the direction of mutational change greatly facilitates the analysis of population relationships but is generally not possible for other types of loci.

We have previously described a rapid, PCR-based assay for determining genotypes (homozygous for the absence of the insertion, homozygous for the presence of the insertion, or heterozygous) for polymorphic *Alu* insertions (6, 9, 11). Here, we report on the distribution of four polymorphic *Alu* insertions in a worldwide survey of 664 individuals from 16 population groups. Our results indicate that these polymorphic *Alu* insertions probably have an African origin and that they are indeed useful loci for human population genetic studies.

MATERIALS AND METHODS

DNA Samples and Cell Lines. Individual DNA samples were isolated from peripheral blood lymphocytes as described (6). The geographic origin of each population group

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ML, maximum likelihood; NJ, neighbor joining; RFLP, restriction fragment length polymorphism; HS, human-specific; CAR, Central African Republic; PNG, Papua New Guinea.

†M.A.B. and M.S. contributed equally to this work.

‡To whom reprint requests should be addressed.

is shown in Fig. 1. The Alaska Natives were composed of Eskimos, Native Amerindians, and Aleuts (12). The Quechua group inhabits the South Andean regions of South America (13). The Arhuaco group resides in the Sierra Nevada region of Northern Colombia (14). The Caucasian group consisted of United States individuals with Northern European ancestry. The Asian group consisted of a mix of Chinese and Vietnamese samples. The African-American group was collected in New Orleans, Louisiana. The Cypriots were collected on the island of Cyprus and were composed of individuals from Greek-Cypriot and Turkish-Cypriot communities. African DNA samples were from two groups of Pygmies (Zaire and the CAR) and Nigerians. Samples from two Indonesian groups (Moluccas and Nusa Tengarras), two PNG groups (highland and coastal), and Australians were typed previously for the TPA 25 *Alu* element and are described in more detail elsewhere (11). Primate DNA samples consisting of five individual chimpanzees (*Pan troglodytes*), one gorilla (*Gorilla gorilla*), three orangutans (*Pongo pygmaeus*), one macaque (*Macaca fascicularis*), and one marmoset (*Leontopithecus saguinus*) were obtained from Bios (New Haven, CT). Rodent/human hybrid cell line DNA panels were obtained from the Coriell Institute for Medical Research (NIGMS panels 1 and 2). The other cell lines used in this study were the same as those described (6).

PCR Amplification. Amplification and analysis of DNA samples was carried out as described (6). The oligonucleotide primers for the TPA 25 and ACE loci were previously reported (6, 15). The primers and annealing temperatures for the PV 92 *Alu* repeat were 5′-AACTGGGAAAATTTGAAGAGAAAGT-3′ (5′ primer) and 5′-TGAGTTCTCAACTCCTGTGTGTTAG-3′ (3′ primer) (54°C); those for APO were 5′-AAGTGCTGTAGGCCATTTAGATTAG-3′ (5′ primer) and 5′-AGTCTTCGATGACAGCGTATACAGA-3′ (3′ primer) (50°C). The chromosomal location for PV 92 was determined by PCR amplification of Coriell Institute rodent/human hybrid cell line DNA panels (1 or 2). The distribution of each *Alu* element across primate species was determined using PCR-based analysis of orthologous positions within nonhuman primate genomes as described (6).

Data Analysis. Unbiased estimates of average heterozygosity, the associated standard error due to sampling, and G_{st} values (a measure of the relative magnitude of genetic differentiation among populations) were calculated according to equations in Nei (16). The GENDIST program in PHYLIP 3.4 (J. Felsenstein, University of Washington, Seattle) was used to compute genetic distances according to the methods of Cavalli-Sforza and Edwards (17), Nei (18), and Reynolds *et al.* (19), and neighbor-joining (NJ) trees (20) were constructed from these genetic distances using the program NEIGHBOR. The CONTML program was used to estimate a maximum-likelihood (ML) tree (21) directly from the allele frequencies. Trees were rooted as described in the text; the likelihood-ratio test of Kishino and Hasegawa (22), as implemented in PHYLIP 3.4, was used to evaluate the significance of alternative placements of the root on the tree.

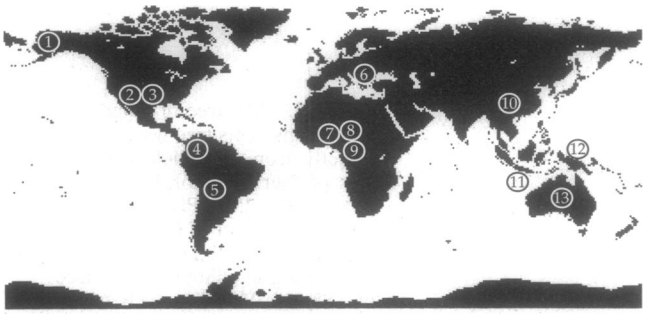

FIG. 1. Geographical map of surveyed human population groups. The figure is a map of the world with the geographic location of all the population groups used in this study denoted by circles with numbers. The populations surveyed were 1, Alaska Natives; 2, United States Caucasians; 3, African-Americans; 4, Arhuaco; 5, Quechua; 6, Greek and Turkish Cypriots; 7, Nigerians; 8, Central African Republic (CAR) Pygmies; 9, Zaire Pygmies; 10, Asians (Vietnamese and Chinese); 11, Indonesians (Nusa Tengarras and Moluccas); 12, coastal and highland Papua New Guineans (PNGs); and 13, Australian Aborigines.

RESULTS

Chromosomal Location and Distribution of Polymorphic *Alu* Insertions. The chromosomal locations of the TPA 25 (23), APO (24), and ACE (15) repeats were previously reported as 8, 11, and 17, respectively. Amplification of the hybrid cell line DNA panel showed that the PV 92 *Alu* insertion site is located on chromosome 16, and hence each of the four loci reported here reside on different chromosomes. To verify that the polymorphic *Alu* insertions are indeed HS, DNA samples from a number of representative nonhuman primate genomes were analyzed. We sampled a total of six chimpanzees, two gorillas, three orangutans, one macaque, one green monkey, one owl monkey, and one marmoset, each of which did not contain any of the polymorphic *Alu* insertions reported here (data not shown). Therefore, we conclude that the four *Alu* insertions analyzed here are indeed HS and that the polymorphism at each locus is due to the recent insertion of each *Alu* element sometime after the divergence of human and African ape lineages.

Human Genetic Variability. A total of 664 individuals from 16 populations were screened for the four polymorphic *Alu* insertion loci. For each locus, the frequency of each allele and the average heterozygosity in each population are reported in Table 1. All loci were polymorphic in all populations with the exception of APO, which was fixed for the presence of the *Alu* insertion in the Quechua and Arhuaco. Only three departures from Hardy–Weinberg equilibrium were noted using a χ^2 test for goodness of fit (Nusa Tengarra for ACE, Greek-Cypriot for APO, and African-American for PV92). The most likely explanation for these deviations is that they represent normal statistical fluctuations, since there were 62 tests for goodness-of-fit to Hardy–Weinberg proportions, and 1 out of 20 tests are expected to be significant at the 5% level. In addition, χ^2 tests for associations between each pair of loci were carried out for each population. No significant associations were detected, which is not surprising since each *Alu* insertion is located on a different chromosome.

The heterozygosity for each population, averaged across all four loci, was substantial, ranging from 0.134 in the Arhuaco to 0.447 in the Nusa Tengarras (Table 1). The Arhuaco, Australians, and Quechua have the lowest heterozygosities, which is not surprising since they probably represent smaller population sizes and/or fewer founders than the other populations. The heterozygosity for each locus, averaged across all 16 populations, ranged from 0.290 for APO to 0.479 for PV 92 (Table 1). These heterozygosities are quite high, especially since the *Alu* insertion loci are biallelic and hence have a maximum possible heterozygosity of 0.5 and attest to the value of these polymorphic *Alu* insertion loci for population genetic studies.

To further investigate the utility of the polymorphic *Alu* insertions for human population studies, the G_{st} value (16) was calculated for each locus (Table 1). The G_{st} value estimates the proportion of the total variance that is due to differences among populations: the higher the G_{st} value, the greater the magnitude of genetic differentiation among populations. The G_{st} values ranged from 0.097 for TPA 25 to 0.283 for PV 92, and all were statistically significant by contingency χ^2 analysis. We therefore conclude that there are

Table 1. Distribution of polymorphic *Alu* insertions

Population	*n*	TPA 25			PV 92			APO			ACE			Pop. Het ± SE
		f_{Alu}	Het	SE	f_{Alu}	Het	SE	f_{Alu}	Het	SE	f_{Alu}	Het	SE	
PNG														
Coastal	47	0.160	0.271	0.052	0.362	0.467	0.028	0.660	0.454	0.032	0.660	0.454	0.032	0.411 ± 0.047
Highland	69	0.159	0.270	0.043	0.239	0.367	0.038	0.681	0.438	0.029	0.739	0.388	0.036	0.366 ± 0.035
Australian														
Aborigine	99	0.126	0.222	0.035	0.152	0.258	0.036	0.869	0.229	0.035	0.909	0.166	0.033	0.219 ± 0.019
Indonesian														
Nusa														
Tengarras	91	0.385	0.476	0.017	0.500	0.503	0.004	0.780	0.345	0.035	0.637	0.465	0.020	0.447 ± 0.035
Moluccas	49	0.561	0.498	0.014	0.694	0.429	0.037	0.755	0.374	0.045	0.673	0.444	0.033	0.436 ± 0.025
Asian														
Chinese and														
Vietnamese	16	0.531	0.514	0.025	0.813	0.315	0.087	0.906	0.175	0.084	0.688	0.444	0.064	0.362 ± 0.075
Amerindian														
Quechua	20	0.675	0.450	0.054	0.875	0.224	0.079	1.000	0.000	0.000	0.700	0.431	0.060	0.276 ± 0.105
Arhuaco	20	0.125	0.224	0.079	0.975	0.050	0.047	1.000	0.000	0.000	0.850	0.262	0.080	0.134 ± 0.064
Alaska Native	62	0.363	0.466	0.024	0.645	0.462	0.025	0.992	0.016	0.016	0.637	0.466	0.024	0.353 ± 0.112
Cypriot														
Greek	50	0.530	0.503	0.009	0.250	0.379	0.044	0.950	0.096	0.039	0.390	0.481	0.023	0.365 ± 0.094
Turkish	33	0.576	0.496	0.021	0.333	0.451	0.040	0.985	0.030	0.029	0.333	0.451	0.040	0.357 ± 0.109
Caucasian	32	0.641	0.468	0.035	0.141	0.246	0.063	0.922	0.146	0.057	0.469	0.506	0.014	0.341 ± 0.087
Pygmies														
Zaire	17	0.235	0.371	0.079	0.353	0.471	0.052	0.853	0.258	0.086	0.324	0.451	0.060	0.388 ± 0.048
CAR	17	0.206	0.337	0.083	0.265	0.401	0.073	0.735	0.401	0.073	0.118	0.214	0.085	0.338 ± 0.044
Nigerian	11	0.409	0.506	0.050	0.091	0.173	0.101	0.500	0.524	0.033	0.273	0.416	0.090	0.405 ± 0.081
African-Amer.	31	0.419	0.495	0.023	0.177	0.297	0.063	0.565	0.500	0.020	0.355	0.465	0.037	0.439 ± 0.048
Locus Het		0.455 ± 0.008			0.479 ± 0.006			0.290 ± 0.014			0.472 ± 0.006			
G_{st}		0.097			0.283			0.140			0.138			

f_{Alu}, frequency of the *Alu* insertion; Het, heterozygosity; Pop. Het, heterozygosity for each population, averaged across all four loci; Locus Het, heterozygosity for each locus, averaged across all 16 populations; Amer., American.

significant differences among human populations in the frequency of each polymorphic *Alu* insertion, further testifying to their usefulness as markers in studying genetic variation in humans.

Population Relationships. To investigate the genetic relationships of the 16 populations based on the four polymorphic *Alu* insertion loci, three different measures of genetic distance were calculated from the allele frequencies in Table 1, and a NJ tree was constructed for each distance measure. We also constructed a ML tree directly from the allele frequencies in Table 1; the ML tree is depicted in Fig. 2. The topology of the ML tree consists of four major branches composed of African (top left), Caucasian (bottom left), Pacific (top right), and Asian as well as New World groups (bottom right). The three NJ trees shared these same four major branches and only differed in the order of branching among the Asian and New World groups (data not shown); the order of branching of populations was the same everywhere else for all four trees.

The allele frequencies alone do not provide any information as to the location of the root of the tree (i.e., the position of the ancestral population on the tree). Placement of the root on the tree requires additional information (e.g., knowledge of ancestral frequencies) or assumptions (e.g., a constant rate

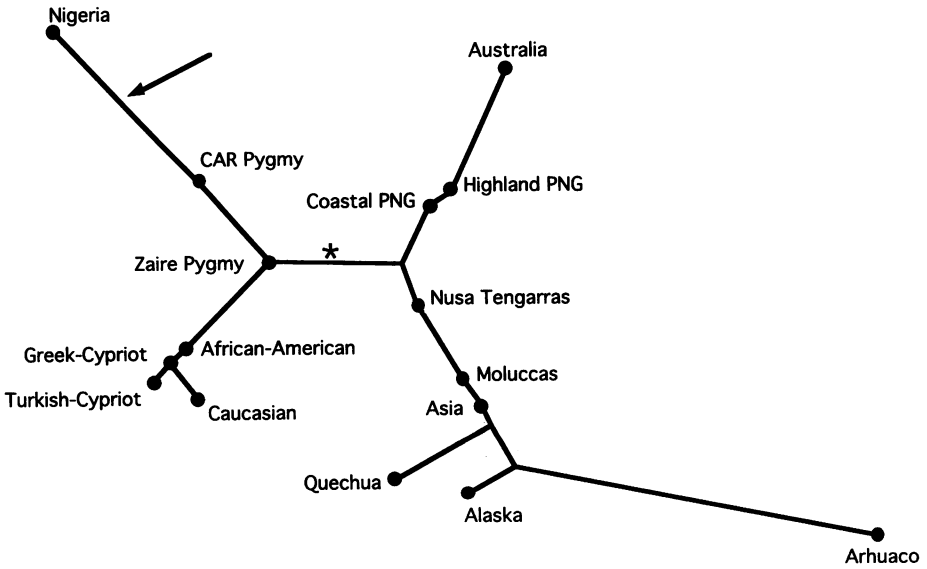

FIG. 2. ML tree of human relationships. This tree was derived directly from the allele frequencies of four polymorphic *Alu* repeats (TPA 25, APO, ACE, and PV 92) in a total of 664 unrelated individuals (Table 1) using PHYLIP 3.4; the log-likelihood of this tree was 88.936. The distance between population groups is proportional to the branch lengths on the tree. Addition of a hypothetical ancestor that does not contain the *Alu* repeats results in a branch that connects with the tree at the position denoted by the arrow in the African branch; the asterisk indicates an alternative placement of the root that is not significantly worse than the optimal placement.

of evolution). For most genetic polymorphisms, there is no way to know what the allele frequencies might have been in the ancestral population. However, since the ancestral state for the HS polymorphic *Alu* insertions is the absence of the insertion, it seems reasonable to suppose that the ancestral population would have been fixed for the absence of the insertion. We therefore included in the phylogenetic analyses a hypothetical ancestral population with allele frequencies of zero for the *Alu* insertion at each locus. The point at which this hypothetical ancestral population attached to the ML population tree is indicated by the arrow in Fig. 2 and is within the African branch, suggesting an African origin of these polymorphic *Alu* insertions. A similar placement of the root was found with the NJ trees (data not shown).

The robustness of the placement of the root in the African branch of the ML tree was evaluated in two ways. First, even though the ancestral state of the HS polymorphic *Alu* insertions is the absence of the insertion, all this really means is that the *Alu* insertion occurred sometime after human and African ape lineages diverged. The frequency of the insertion may have been greater than zero (i.e., polymorphic) in the actual ancestral human population from which all contemporary human populations are derived. We therefore sequentially incremented the frequency of the *Alu* insertion in steps of 0.05 at each locus in the hypothetical ancestral population and repeated the ML analysis. The placement of the root did not change until the frequency of each *Alu* insertion reached 0.45 or more in the ancestral population, indicating that the placement of the root is actually relatively insensitive to the hypothetical ancestral frequency of each *Alu* insertion. Second, the likelihood-ratio test (22) was used to evaluate if alternative placements of the root on the ML tree were significantly inferior to the actual placement in Fig. 2. Moving the root outside of the African, European, Pacific, and Asian/Americas groupings to the branch indicated by an asterisk in Fig. 2 does not significantly decrease the likelihood, but moving the root within either the European or Asian/Americas branches does result in trees with significantly lower likelihoods (data not shown). Since the likelihood-ratio test is based on the number of loci, with only four loci it is not surprising that small alterations in the placement of the root do not significantly decrease the likelihood.

DISCUSSION

Alu insertion polymorphisms have several desirable properties for studying genetic variation in human populations. First, the nonradioactive, PCR-based detection method for these polymorphisms makes it feasible to rapidly screen large numbers of DNA samples isolated from a wide variety of sources. By contrast, traditional methods for detecting DNA polymorphisms are more time-consuming, often require radioactive isotopes, and need so much DNA that cell lines often must be established (e.g., refs. 25 and 26).

Second, *Alu* insertions appear to be relatively stable integrations into the genome that are rarely deleted (27, 28). Even when deletion of an *Alu* element occurs, the deletion is not a precise excision of the *Alu* element, but rather it leaves behind a signature of the original insertion event (29). Also, the rate of insertion and fixation of new *Alu* elements is about 100–200 per million years (5, 6), so the independent insertion of two different *Alu* elements at the same location in the genome has essentially no chance of occurring. Therefore, individuals who share polymorphic *Alu* insertions inherited them from a common ancestor, making *Alu* insertion polymorphisms identical by descent. This distinguishes *Alu* insertions from other types of polymorphisms including RFLP (30) and variable numbers of tandem repeats (31), which may arise multiple times within a population and are merely identical by state.

Third, the four *Alu* insertion loci studied here were highly variable both within and among populations. Although these four loci were first detected as polymorphisms in Caucasian populations, and hence might be subject to ascertainment bias, with just two exceptions all four loci were polymorphic in all 16 populations. Heterozygosity values were substantial, exceeding 0.45 for three of the four loci, which is even more remarkable when one considers that these are biallelic loci and hence have a maximum heterozygosity of 0.5. The insertion frequency at each locus varied significantly among the 16 populations, with G_{st} values ranging from 0.097 to 0.283. By comparison, of 42 biallelic DNA markers studied by Bowcock *et al.* (25), 23 had F_{st} values (comparable to G_{st} values) exceeding 0.097, and only four had F_{st} values exceeding 0.283. Since Bowcock *et al.* (25) studied different populations, the comparison with the present study is not strictly accurate, but it does illustrate that there is an appreciable amount of interpopulation differentiation for these *Alu* insertion polymorphisms. Thus, *Alu* insertion polymorphisms provide a useful set of DNA markers for studying human population relationships.

This is also supported by the tree analysis of population relationships. The genetic affinities among the population groups reported here, based on the four polymorphic *Alu* insertion loci, appear to be quite reasonable. Four main groups were revealed, corresponding to Africa, Europe, Asia/Americas, and Australia/New Guinea. The Australia/New Guinea grouping is consistent with previous studies on the genetic structure of these populations (32), as is the placement of the New World populations (Alaska, Quechua, and Arhuaco) with the other Asian groups (33). It is interesting to note that the African-American group is placed between Caucasians and Africans. This is not surprising since previous studies have shown that there is a 10–30% contribution of Caucasian genes to the African-American gene pool (34). Finally, the Caucasian branch of the tree places Greek and Turkish-Cypriots closer together than to a generic group of U.S. Caucasians, which presumably reflects greater ethnic heterogeneity in the group of U.S. Caucasians sampled.

The tree of population relationships based on *Alu* insertion polymorphisms is very similar to population trees based on classical blood protein markers (35, 36), nuclear DNA RFLP loci (26, 37), microsatellite loci (38), and mitochondrial DNA (39). While these previous studies also agree with the present study in placing the ancestral population in Africa, these studies had no reliable information concerning the probable ancestral allele frequencies. Instead, the placement of the root of the tree in all of the above studies was obtained by assuming a constant rate of evolution. This was done either by using the unweighted pair-group method with arithmetic mean type of tree construction (e.g., ref. 35), which assumes a constant rate of evolution, or by first constructing a tree by a different method (such as NJ) and then using midpoint rooting (e.g., refs. 36 and 38), which then invokes a constant rate of evolution by placing the root at the midpoint of the longest path connecting two populations. In either case, it should be noted that the assumption of a constant rate of evolution for such trees of population relationships means that the rate of allele frequency change (not just the rate of mutation) within and between populations would have been constant in time. This assumption can be examined by constructing trees with a method that does not assume a constant rate of evolution (such as ML) and then comparing the lengths of the terminal branches. For example, in Fig. 2 the terminal branches leading to the Alaska and Arhuaco groups should have similar lengths if the rate of evolution was approximately constant, and yet the lengths are clearly very different. The assumption of a constant rate of evolution, when applied to allele frequency data, is therefore at best dubious.

A significant advantage of using the *Alu* insertion polymorphisms to study human population relationships is that one can make inferences regarding ancestral allele frequencies, thereby avoiding the assumption of a constant rate of evolution. Since the ancestral state of each *Alu* insertion is the absence of the insertion (which is supported by the absence of these HS *Alu* insertions at orthologous positions in nonhuman primate genomes), the root of the tree can be obtained by including a hypothetical ancestral population in the analysis in which the frequency of the insertion is zero for each locus. Furthermore, one can also test if alternative placements of the root are significantly inferior, an issue that has not been addressed in previous genetic studies of human population relationships.

The results of such analyses indicate that the most probable placement of the ancestral population for these four polymorphic *Alu* insertions is in Africa. The placement of the root does not shift unless one greatly increases the presumed ancestral frequencies of the insertions; however, placing the root outside the African branch in the tree is not significantly inferior to the optimal rooting within the African branch. This latter result is not surprising, since only four loci were analyzed, which are too few to obtain a reliable indication of population relationships (16). Analysis of more loci is required to obtain strong statistical support for a particular tree topology. Nevertheless, the placement of the root within the African branch of the tree does suggest an African origin for these polymorphic *Alu* insertions. An African origin has similarly been inferred for classical markers (35, 36), DNA RFLP markers (26, 37), microsatellite loci (38), and mitochondrial DNA (39–41). Such concordant results over such diverse data sets provides strong support for a recent African origin of modern humans (42).

In conclusion, the application of *Alu* insertion polymorphisms to the study of human population genetics provides a new set of rapidly and easily screened nuclear DNA markers for investigating relationships among populations. It has been estimated that there may be as many as 400 polymorphic *Alu* insertions in the human genome (9). The isolation of additional polymorphic *Alu* insertions should facilitate incisive investigation of both the evolutionary history and genetic structure of modern population groups, as well as the analysis of admixture between groups. In addition, *Alu* insertions provide the opportunity to study evolution over even more recent time scales. Some *Alu* insertions appear to be restricted to single families, such as the cholinesterase *Alu* family member (43). Other HS *Alu* insertions have been reported that represent unique (*de novo*) insertions into the NF-1 (44) and factor IX (45) loci and result in neurofibromatosis and hemophilia, respectively. These *Alu* insertions represent unique genetic variants located in the genomes of single individuals from the human population. Polymorphic *Alu* insertions clearly represent an ongoing evolutionary process in the human genome, and the *Alu* family of repeats represent a unique source of genetic variation for human population genetics and forensic identity testing.

We thank K. Bhatia, L. L. Cavalli-Sforza, N. Kretchmer, A. S. M. Sofro, J. Kidd, J. Yunis, E. Yunis, S. J. O'Brien, and J. Wainscoat for samples. P.A.I. was partially funded by the Ministry of Health in Cyprus. This research was supported by grants from the U.S. Department of Energy (LDRD 94-LW-103) to M.A.B.; National Institutes of Health (RO1 HG 00770) to P.L.D., (HL-42082) to W.D.S., and (RR08205) to R.J.H.; and National Science Foundation (BNS 90-20567) to M.S. Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy contract no. W-7405-ENG-48.

1. Deininger, P. L. & Batzer, M. A. (1993) in *Evolutionary Biology*, eds. Hect, M. K., MacIntyre, R. J. & Clegg, M. T. (Plenum, New York), pp. 157–196.
2. Schmid, C. W. & Maraia, R. (1992) *Curr. Opin. Genet. Dev.* **2**, 874–882.
3. Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H., Jr. & Deininger, P. L. (1987) *Mol. Biol. Evol.* **4**, 19–29.
4. Willard, C., Nguyen, H. T. & Schmid, C. W. (1987) *J. Mol. Evol.* **26**, 180–186.
5. Batzer, M. A., Kilroy, G. E., Richard, P. E., Shaikh, T. H., Desselle, T. D., Hoppens, C. L. & Deininger, P. L. (1990) *Nucleic Acids Res.* **18**, 6793–6798.
6. Batzer, M. A. & Deininger, P. L. (1991) *Genomics* **9**, 481–487.
7. Matera, A. G., Hellmann, U. & Schmid, C. W. (1990) *Mol. Cell. Biol.* **10**, 5424–5432.
8. Matera, A. G., Hellmann, U., Hintz, M. F. & Schmid, C. W. (1990) *Nucleic Acids Res.* **18**, 6019–6023.
9. Batzer, M. A., Gudi, V. A., Mena, J. C., Foltz, D. W., Herrera, R. J. & Deininger, P. L. (1991) *Nucleic Acids Res.* **19**, 3619–3623.
10. Leeflang, E. P., Liu, W.-M., Hashimoto, C., Choudary, P. V. & Schmid, C. W. (1992) *J. Mol. Evol.* **35**, 7–16.
11. Perna, N. T., Batzer, M. A., Deininger, P. L. & Stoneking, M. (1992) *Human Biol.* **64**, 641–648.
12. Newman, W. P., Middaugh, J. P., Propst, M. T. & Rogers, D. R. (1993) *Lancet* **341**, 1056–1057.
13. Cooper, J. M. (1963) in *Handbook of South American Indians*, ed. Steward, J. H. (Cooper Square, New York), Vol. 2, p. 160.
14. Park, W. Z. (1963) in *Handbook of South American Indians*, ed. Steward, J. H. (Cooper Square, New York), Vol. 1, p. 868.
15. Tiret, L., Rigat, B., Visvikis, S., Breda, C., Corval, P., Cambien, F. & Soubrier, F. (1992) *Am. J. Hum. Genet.* **51**, 197–205.
16. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
17. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) *Evolution* **32**, 550–570.
18. Nei, M. (1972) *Am. Nat.* **106**, 283–292.
19. Reynolds, J. B., Weir, B. S. & Cockerham, C. C. (1983) *Genetics* **105**, 767–779.
20. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
21. Felsenstein, J. (1981) *Evolution* **35**, 1229–1242.
22. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29**, 170–179.
23. Yang-Feng, T. L., Opdenakker, G., Volckaert, G. & Franke, U. (1986) *Am. J. Hum. Genet.* **39**, 79–87.
24. Karathanasis, S. K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6374–6378.
25. Bowcock, A. M., Bucci, C., Hebert, J. M., Kidd, J. R., Kidd, K. K., Friedlaender, J. S. & Cavalli-Sforza, L. L. (1987) *Gene Geogr.* **1**, 47–64.
26. Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K. & Cavalli-Sforza, L. L. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 839–843.
27. Sawada, I. & Schmid, C. W. (1986) *J. Mol. Biol.* **192**, 693–703.
28. Bailey, A. D. & Shen, C.-K. J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7205–7209.
29. Edwards, M. C. & Gibbs, R. A. (1992) *Genomics* **14**, 590–597.
30. Botstein, D., White, R. L., Skolnick, M. H. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331.
31. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. & White, R. (1987) *Science* **235**, 1616–1622.
32. Serjeantson, S. W. & Hill, A. V. S. (1989) in *The Colonization of the Pacific: A Genetic Trail*, eds. Hill, A. V. S. & Serjeantson, S. W. (Oxford Univ. Press, Oxford), pp. 286–294.
33. Meltzer, D. J. (1993) *Evol. Anth.* **1**, 157–169.
34. Chakraborty, R., Kamboh, M. I., Nwankwo, M. & Ferrell, R. E. (1992) *Am. J. Hum. Genet.* **50**, 145–155.
35. Cavalli-Sforza, L. L., Piazza, A., Menozzi, P. & Mountain, J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6002–6006.
36. Nei, M. & Roychoudhury, A. K. (1993) *Mol. Biol. Evol.* **10**, 927–943.
37. Wainscoat, J. S., Hill, A. V. S., Boyce, A. L., Flint, J., Hernandez, M., Thein, S. L., Old, J. M., Lynch, J. R., Falusi, A. G., Weatherall, D. J. & Clegg, J. B. (1986) *Nature (London)* **319**, 491–493.
38. Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. & Cavalli-Sforza, L. L. (1994) *Nature (London)* **368**, 455–457.
39. Merriweather, D. A., Clark, A. G., Ballinger, S. W., Schurr, T. G., Soodyall, H., Jenkins, T., Sherry, S. T. & Wallace, D. W. (1991) *J. Mol. Evol.* **33**, 543–555.
40. Cann, R. L., Stoneking, M. & Wilson, A. C. (1987) *Nature (London)* **325**, 31–36.
41. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) *Science* **253**, 1503–1507.
42. Stoneking, M. (1993) *Evol. Anth.* **2**, 60–73.
43. Muratani, K., Hada, T., Yamamoto, Y., Kaneko, T., Shigeto, Y., Ohue, T., Furuyama, J. & Higashino, K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11315–11319.
44. Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W. & Collins, F. S. (1991) *Nature (London)* **353**, 864–866.
45. Vidaud, D., Vidaud, M., Bahnak, B. R., Siguret, V., Sanchez, S. G., Laurin, Y., Meyer, D., Goossens, M. & Lavergne, J. M. (1993) *Eur. J. Hum. Genet.* **1**, 30–36.