J Mol Evol (1996) 42:22–29

JOURNAL OF MOLECULAR EVOLUTION

Genetic Variation of Recent Alu Insertions in Human Populations

Mark A. Batzer,[1] Santosh S. Arcot,[1] Joshua W. Phinney,[1] Michelle Alegria-Hartman,[1] David H. Kass,[2] Stephen M. Milligan,[3] Colin Kimpton,[4] Peter Gill,[4] Manfred Hochmeister,[5] Panayiotis A. Ioannou,[6] Rene J. Herrera,[7] Donald A. Boudreau,[8] W. Douglas Scheer,[8] Bronya J. B. Keats,[9] Prescott L. Deininger,[2,10] Mark Stoneking[11]

[1] Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551, USA
[2] Department of Biochemistry and Molecular Biology, Center for Human and Molecular Genetics, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112, USA
[3] DNA Unit, Michigan State Police, 714 South Harrison Road, East Lansing, MI 48823, USA
[4] Forensic Science Service, Priory House, Gooch Street North, Birmingham B5 6QQ, United Kingdom
[5] Department of Forensic Medicine, University of Bern, Buehlstrasse 20, Bern CH-3012, Switzerland
[6] The Cyprus Institute of Neurology and Genetics, P.O. Box 3462, Nicosia, Cyprus
[7] Department of Biological Sciences, Florida International University, University Park Campus, Miami, FL 33199, USA
[8] Department of Pathology, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112, USA
[9] Department of Biometry and Genetics, Center for Human and Molecular Genetics, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112, USA
[10] Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121, USA
[11] Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA

Received: 1 November 1994 / Accepted: 12 December 1994

Abstract. The Alu family of interspersed repeats is comprised of over 500,000 members which may be divided into discrete subfamilies based upon mutations held in common between members. Distinct subfamilies of Alu sequences have amplified within the human genome in recent evolutionary history. Several individual Alu family members have amplified so recently in human evolution that they are variable as to presence and absence at specific loci within different human populations. Here, we report on the distribution of six polymorphic Alu insertions in a survey of 563 individuals from 14 human population groups across several continents. Our results indicate that these polymorphic Alu insertions probably have an African origin and that there is a much smaller amount of genetic variation between European populations than that found between other population groups.

Key words: Human evolution — African origin — Identical by descent — Polymorphism

Introduction

Short interspersed elements (SINEs) may be found in the genomes of a wide variety of mammals (Deininger and Batzer 1993). The Alu family of SINEs is one of the most successful mobile genetic elements, having arisen to a copy number in excess of 500,000 within the human genome in approximately 65 million years of primate evolution. (For reviews see Deininger 1989; Okada 1991; Schmid and Maraia 1992; Deininger and Batzer 1993.) Alu sequences are thought to be ancestrally derived from the 7SL RNA gene (Ullu et al. 1982) and mobilize through an RNA polymerase III–derived transcript in a process termed ''retroposition'' (Rogers 1983).

Alu sequences within the human genome can be di-

*Present address: Department of Pathology, Stanley S. Scott Cancer Center, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112
Correspondence to: M.A. Batzer

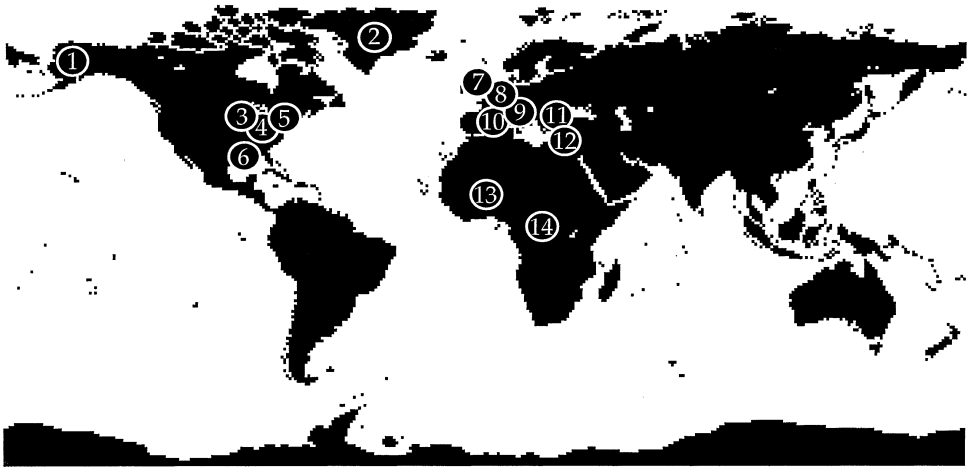

Fig. 1. Geographical map of sampled human population groups. A map of the world with the locations of all the populations sampled in this study denoted by *circles with numbers.* The populations were (*1*) Alaska Natives; (*2*) Greenland Natives; (*3*) European-Americans; (*4*) Hispanics; (*5*) African-Americans; (*6*) French Acadians; (*7*) British Afro-Caribbeans; (*8*) French Bretons; (*9*) Swiss; (*10*) French; (*11*) Greek-Cypriots; (*12*) Turkish-Cypriots; (*13*) Nigerians; (*14*) Central African Republic and Zaire Pygmies.

vided into groups of related elements based upon commonly shared diagnostic mutations. Several groups have independently identified a series of overlapping subfamilies of Alu repeats which appear to be different genetic ages (Slagel et al. 1987; Willard et al. 1987; Britten et al. 1988; Jurka and Smith 1988; Quentin 1988; Deininger and Slagel 1988; Shen et al. 1991; Jurka and Miloslavejevic 1991). These observations have led to the suggestion that the vast majority of Alu amplifications were derived from a small subset of active Alu ''master'' genes (Deininger et al. 1992). However, it is clear that a limited number of Alu subfamilies are currently undergoing amplification from multiple ''master'' genes within the human genome (Matera et al. 1990b; Leeflang et al. 1992; Jurka 1993; Hutchinson et al. 1993; Hammer 1994; Batzer et al. 1995). Here, we will use the nomenclature of Shen et al. (1991) and Jurka (1993) to refer to various Alu subfamilies.

One of the most recently formed groups of Alu elements within the human genome has been termed human-specific (HS) (Batzer et al. 1990; Batzer and Deininger 1991) or predicted variant (PV) (Matera et al. 1990a,b) and was derived from the CS subfamily (Shen et al. 1991) of Alu repeats. There are an estimated 500–2,000 HS Alu elements which are mostly (Batzer and Deininger 1991), but not exclusively (Leeflang et al. 1992, 1993), restricted to the human genome. In parallel, a second subfamily which is an independent derivative of the CS lineage of Alu sequences, termed Sb2 (Jurka 1993; Hutchinson et al. 1993), has also expanded in the human genome (Batzer et al. 1995; Zietkiewicz et al. 1994).

Some HS and Sb2 Alu elements have retroposed so recently that they have not fixed within the human genome (Batzer and Deininger 1991; Batzer et al. 1991; Hammer 1994; Kass et al. 1994; Batzer et al. 1995). The distribution of these elements varies in geographically distinct human population groups (Batzer et al. 1991; Batzer and Deininger 1991; Perna et al. 1992; Hammer 1994; Kass et al. 1994; Batzer et al. 1994). The recent Alu insertions provide a novel set of highly informative nuclear DNA markers for the study of human population genetics since they represent relatively stable polymorphisms that are identical by descent, and the ancestral state of the polymorphism is known (Batzer et al. 1994). Here, we report on the distribution of six polymorphic Alu insertions in a survey of 563 individuals from 14 population groups across several continents. Our results indicate that these polymorphic Alu insertions probably have an African origin and that there is a much smaller amount of genetic variation between European populations than that found between other population groups.

Methods

DNA Samples. DNA samples from individuals were isolated from peripheral blood lymphocytes as described (Ausabel et al. 1987). The geographic origin of each population group from this study is shown in Fig. 1. The European-American group consisted of United States individuals with predominantly northern European ancestry. The African-American, Hispanic (largely of Mexican descent), and European-Americans were all collected in Michigan. The Cypriots and Alaska Natives were previously described (Batzer et al. 1994). Swiss samples were obtained in Switzerland. French Acadian samples were collected from individuals located in the Acadian triangle region near Lafayette, Louisiana. The French and Breton samples were previously described (Monson et al., in press). Greenland Natives were collected from Greenland. DNA samples from two groups of African Pygmies (Zaire and the Central African Republic [CAR]) were kindly provided by L. Cavalli-Sforza, while Nigerian DNA samples were kindly provided by J. Wainscoat. Afro-Caribbean samples were collected from the British Isles and were comprised of individuals with African heritage.

PCR Amplification. Amplification of DNA samples was carried out in 50-μl reactions using 50 ng of target DNA, 375 ng of each oligonucleotide, 200 μM dNTPs in 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM

Table 1. Oligonucleotide primers, annealing temperatures, and location of polymorphic *Alu* repeats[a]

Repeat	Sub-family	Primer sequences		Annealing temperature (°C)	Chromosomal location
		5′ primer	3′ primer		
TPA 25	HS-2	5′-GTAAGAGTTCCGTAACAGGACAGCT-3′	5′-CCCCACCCTAGGAGAACTTCTCTTT-3′	58	8
PV 92	HS-1	5′-AACTGGGAAAATTTGAAGAGAAAGT-3′	5′-TGAGTTCTCAACTCCTGTGTGTTAG-3′	54	16
FXIIIB	HS-1	5′-TCAACTCCATGAGATTTTCAGAAGT-3′	5′-CTGGAAAAAATGTATTCAGGTGAGT-3′	56	1
D1	Sb2	5′-TGCTGATGCCCAGGGTTAGTAAA-3′	5′-TTTCTGCTATGCTCTTCCCTCTC-3′	70	3
APO	HS-1	5′-AAGTGCTGTAGGCCATTTAGATTAG-3′	5′-AGTCTTCGATGACAGCGTATACAGA-3′	50	11
ACE	HS-1	5′-CTGGAGACCACTCCCATCCTTTCT-3′	5′-GATGTGGCCATCACATTCGTCAGAT-3′	58	17

[a]Subfamily nomenclature as described in Batzer et al. (1990) and Shen et al. (1991) for HS-1 and HS-2 and in Jurka (1993) and Hutchinson et al. (1993) for Sb2. Oligonucleotide primers as reported in references below except as noted TPA 25 (Batzer and Deininger 1991), FXIIIB (Kass et al. 1994), APO (Batzer et al. 1994). Chromosomal locations were previously reported for TPA 25 (Yang-Feng et al. 1986), PV 92 (Batzer et al. 1994), FXIIIB (Webb et al. 1989), D1 (Batzer et al. 1995), APO (Karathanasis 1985) and ACE (Tiret et al. 1992).

Tris-HCl pH 8.4, and AmpliTaq DNA polymerase (2.5 U) according to the supplier's (Roche Molecular Diagnostics) instructions. Each sample was subjected to the following amplification conditions: 1 min at 94°C (denaturation), 2 min at the appropriate annealing temperature, and 2 min at 72°C (extension) for 32 cycles. The annealing temperature and oligonucleotide primer sequences for each Alu insertion are shown in Table 1. Twenty microliters of each PCR reaction was analyzed by electrophoresis through a 2% agarose gel containing 0.5 μg/ml ethidium bromide, and the reaction products were directly visualized using UV fluorescence.

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 (1987). The CONTML program in PHYLIP 3.4 was used to estimate a maximum-likelihood tree (Felsenstein 1981) directly from the allele frequencies. The maximum-likelihood tree was rooted by setting the frequency of each insertion to zero as previously described (Batzer et al. 1994). In addition, a series of genetic distance matrices were constructed using the GENDIST program in PHYLIP 3.4 according to the methods of Reynolds et al. (1983), Nei (1972), and Cavalli-Sforza and Edwards (1967), and each distance matrix was used to construct a neighbor-joining tree (rooted as outlined above) using the program NEIGHBOR.

To assess the relative amount of gene flow experienced by each population, the heterozygosity of each population was plotted against the distance of the population from the centroid, as described by Harpending and Ward (1982), where heterozygosity is the usual expected heterozygosity under Hardy-Weinberg, and the distance from the centroid r_i for a population i is:

$$r_i = (p_i - P)^2/(P)(1 - P)$$

where p_i and P are the frequency of the Alu insertion in population i and in the total population, respectively. According to Harpending and Ward (1982), under an island model of population structure, the theoretical expectation is that there should exist a linear relationship between heterozygosity and distance from the centroid:

$$h_i = H(1 - r_i)$$

where h_i and H are the heterozygosities of population i and the total population, respectively. Of particular interest in this analysis are the outliers: populations that have experienced more gene flow than average will fall above the theoretical prediction, while populations that have experienced less gene flow than average will fall below the theoretical prediction.

Results

Genetic Variation Within Populations

The distribution of six individual polymorphic Alu insertions was determined in a total of 563 unrelated individuals that comprised 14 population groups and is summarized in Table 2. Each Alu insertion was polymorphic in all of the populations except for the D1 repeat, which was not found within a small sample of Nigerians (n = 11). A total of 84 tests for Hardy-Weinberg equilibrium were performed, and only two significant departures from Hardy-Weinberg equilibrium were found (French Acadians for D1 and French for factor 13B [FXIIIB]) in cases where sufficient numbers of each genotype were present. This number of departures is not surprising since approximately 4 of the 84 tests should be significant at the 5% level based upon chance alone. However, the French Acadians also exhibited a nearly significant departure from Hardy-Weinberg equilibrium at the TPA 25 locus as well ($\chi^2 = 3.72$, $df = 1$, $0.05 < p < 0.1$). For both D1 and TPA 25 there were fewer heterozygotes observed than expected, which might indicate inbreeding within this group.

The heterozygosity for each population, averaged across the six Alu insertions, was fairly substantial, ranging from a high of 0.436 in African-Americans to a low of 0.296 in Nigerians. The heterozygosity values of each marker, averaged across all of the populations, were also quite high, with all of the Alu insertions except for APO having heterozygosity values in excess of 0.4. This is impressive given that each Alu insertion is a bi-allelic polymorphism with a maximum heterozygosity of 0.5.

Genetic Differentiation Among Populations

To examine the amount of genetic differentiation among populations, G_{st} values (a measure of the interpopulation variability) for each Alu insertion were determined. The G_{st} values ranged from a high of 0.236 for the factor 13B Alu insertion to a low of 0.039 for the ACE Alu repeat;

Table 2. Distribution of polymorphic Alu insertions

		TPA 25			PV 92			APO		
Population	*n*	Frequency of Alu	Het	SE	Frequency of Alu	Het	SE	Frequency of Alu	Het	SE
European-Americans	45	0.556	0.499	0.014	0.178	0.296	0.052	0.944	0.106	0.043
African-Americans	43	0.302	0.427	0.040	0.209	0.335	0.051	0.570	0.496	0.017
Hispanics	44	0.625	0.474	0.027	0.523	0.505	0.009	0.920	0.148	0.049
Afro-Caribbeans	42	0.286	0.413	0.043	0.143	0.248	0.055	0.500	0.506	0.008
Swiss	43	0.453	0.502	0.013	0.198	0.321	0.052	0.942	0.111	0.045
Bretons	45	0.556	0.499	0.014	0.267	0.396	0.044	0.900	0.182	0.05
French Acadians	45	0.433	0.497	0.016	0.178	0.296	0.052	0.922	0.145	0.048
Greek Cypriots	50	0.530	0.503	0.009	0.250	0.379	0.044	0.950	0.096	0.039
Turkish Cypriots	33	0.576	0.496	0.021	0.333	0.451	0.040	0.985	0.030	0.029
Nigerians	11	0.409	0.506	0.050	0.091	0.173	0.101	0.500	0.524	0.033
Pygmies	34	0.221	0.349	0.057	0.309	0.433	0.044	0.794	0.332	0.058
French	44	0.557	0.499	0.014	0.227	0.355	0.049	0.989	0.023	0.022
Alaska Natives	42	0.298	0.423	0.041	0.619	0.477	0.026	0.917	0.155	0.050
Greenland Natives	42	0.333	0.450	0.035	0.607	0.483	0.024	0.940	0.113	0.046
Average heterozygosity (Het)			0.494			0.423			0.239	
Standard error (SE)			0.003			0.011			0.015	
G_{st}			0.055			0.132			0.113	

all of the G_{st} values differed significantly from zero, as judged by contingency chi-square analysis of the allele frequencies. Hence, there are significant differences among human populations with respect to the frequencies of all of these Alu insertion polymorphisms.

Contingency chi-square analysis for heterogeneity in allele frequencies was also performed separately for the European, Amerindian (Alaska and Greenland Natives), and African (Nigerian and Pygmy) populations; the populations known to be admixed (Hispanic, African-American, and British Afro-Caribbean) were excluded from this analysis. The two African populations differed significantly in allele frequencies at three loci and the two Amerindian populations differed significantly at one locus, while the seven European populations did not differ significantly at any of the six Alu insertion loci. This analysis indicates that there is little genetic differentiation among European populations, including recently derived European populations such as European-Americans and French Acadians.

To investigate the evolutionary relationships of these populations, a maximum-likelihood tree was constructed directly from the allele frequency data in Table 2 and is depicted in Fig. 2. The topology of the maximum-likelihood tree consisted of five major branches: Nigerians (upper left); Pygmies (lower left); African-Americans and British Afro-Caribbeans (middle); European populations (upper right); and Hispanics and Amerinds (lower right). In addition, three types of genetic distance measures were calculated from the allele frequencies in Table 2, followed by the construction of a neighbor-joining tree from each distance measure. All of the major branches present in the maximum-likelihood tree in Fig. 2 were also present in the neighbor-joining trees; the only differences among the trees were minor changes in the branching order of the European populations (data not shown).

Previously the TPA 25 (Batzer and Deininger 1991) APO, ACE and PV92 (Batzer and Stoneking et al. 1994), D1 (Batzer et al. 1995), and factor 13B (Kass et al. 1994) Alu repeats have been shown to be absent from the genomes of nonhuman primates and located on different human chromosomes as outlined in Table 1. Since the direction of mutation for Alu insertions is the insertion rather than the deletion of each Alu element, the root of the tree was derived by the inclusion of a hypothetical ancestor which did not contain any of the polymorphic Alu insertions (i.e., the allele frequencies for each locus were set to zero). The hypothetical ancestral population connected to the maximum-likelihood network in the African branch (Fig. 2), as was found previously for a larger sample of populations examined for four of these loci (Batzer et al. 1994).

Gene Flow Within Populations

To determine the relative amount of gene flow experienced by each population, we compared the heterozygosity of each group to the genetic distance from the centroid, as described previously (Harpending and Ward 1982). A plot of heterozygosity vs distance from the centroid for the worldwide sample of 14 populations shows that the British Afro-Caribbeans, African-Americans, Hispanics, Greenland Natives, Alaska Natives, and Nigerians are above the predicted values (Fig. 3A), while all of the European populations (cluster of 7 data points) as well as the Pygmies fall below the line (Fig. 3A). Since the sampling of African and Amerind populations in this study is much less extensive than the sampling of European populations, the African and Am-

Table 2. Extended

Population	ACE Frequency of Alu	ACE Het	ACE SE	FXIIIB Frequency of Alu	FXIIIB Het	FXIIIB SE	D1 Frequency of Alu	D1 Het	D1 SE	Average Het ± SE
European-Americans	0.511	0.505	0.008	0.467	0.503	0.011	0.444	0.499	0.014	0.402 ± 0.068
African-Americans	0.488	0.506	0.009	0.221	0.348	0.050	0.488	0.506	0.009	0.436 ± 0.032
Hispanics	0.545	0.502	0.013	0.705	0.421	0.040	0.364	0.468	0.029	0.420 ± 0.056
Afro-Caribbeans	0.524	0.505	0.010	0.310	0.433	0.039	0.405	0.488	0.022	0.432 ± 0.040
Swiss	0.372	0.473	0.028	0.477	0.505	0.010	0.337	0.452	0.034	0.394 ± 0.063
Bretons	0.478	0.505	0.009	0.400	0.485	0.022	0.389	0.481	0.024	0.425 ± 0.051
French Acadians	0.511	0.505	0.008	0.478	0.505	0.009	0.422	0.493	0.018	0.407 ± 0.062
Greek Cypriots	0.390	0.481	0.023	0.616	0.479	0.026	0.267	0.396	0.045	0.389 ± 0.062
Turkish Cypriots	0.333	0.451	0.040	0.394	0.485	0.028	0.348	0.461	0.037	0.396 ± 0.073
Nigerians	0.273	0.416	0.090	0.083	0.159	0.094	0.000	0.000	0.000	0.296 ± 0.088
Pygmies	0.221	0.349	0.057	0.015	0.029	0.028	0.338	0.454	0.038	0.324 ± 0.062
French	0.477	0.505	0.009	0.420	0.493	0.019	0.455	0.502	0.013	0.396 ± 0.078
Alaska Natives	0.583	0.492	0.020	0.917	0.155	0.050	0.415	0.491	0.020	0.366 ± 0.067
Greenland Natives	0.548	0.501	0.013	0.786	0.341	0.051	0.452	0.501	0.013	0.398 ± 0.062
Average heterozygosity (Het)		0.498			0.499			0.475		
Standard error (SE)		0.002			0.002			0.007		
G_{st}		0.039			0.236			0.069		

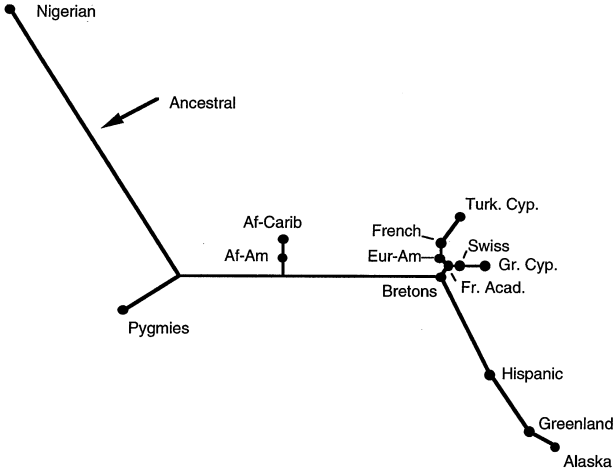

Fig. 2. Maximum-likelihood tree of population relationships. This tree was derived directly from the allele frequencies of six polymorphic Alu repeats (TPA 25, PV 92, APO, ACE, FXIIIB, and D1) in a total of 563 unrelated individuals (Table 2) using PHYLIP 3.4; the log-likelihood of this tree was 135.76. The genetic distance between populations is proportional to the branch lengths on the tree. Addition of a hypothetical ancestor which does not contain any of the Alu repeats results in a branch which connects with the tree at the position denoted by the *arrow* in the African branch.

erind populations were removed and the analysis was repeated (Fig. 3B). The Hispanics, African-Americans, and British Afro-Caribbeans all fall well above the theoretical line, in accordance with the expectation that these admixed populations have received gene flow from other sources, while the European populations have not. When only the European populations are analyzed (Fig. 3C), the fit between the observed and predicted values is quite close, suggesting that the model applies to European populations.

Discussion

Polymorphic Alu insertions represent a unique source of nuclear genetic variability for the study of human population genetics. Each polymorphic Alu insertion arose within the human population as a unique event in human evolutionary history, making the Alu repeats identical by descent from a common ancestor, as opposed to other polymorphisms which are merely identical by state (Batzer and Deininger 1991; Batzer et al. 1991, 1993, 1994; Deininger and Batzer 1995). Alu elements appear to be stable integrations into the genome which rarely delete from a location (Sawada et al. 1985; Sawada and Schmid 1986; Bailey and Shen 1993); even when a rare deletion occurs, a signature of the original insertion event is left behind (Edwards and Gibbs 1992), as an exact excision would be an extremely low-probability event. In addition, Alu elements are subject to very limited amounts of gene conversion (Kass et al. 1995; Batzer et al. 1995). Furthermore, the direction of mutation that results in an Alu polymorphism is known to be forward (i.e., the insertion of the element), facilitating an accurate estimation of the root in trees of population relationships (Batzer et al. 1994).

Each of the polymorphic Alu insertions analyzed here displayed a significant amount of interpopulation differentiation with G_{st} values that ranged from 0.039 for the ACE Alu insertion to 0.236 for the factor 13B Alu repeat. Although the populations studied were different and the G_{st} values are not strictly comparable, Batzer et al. (1994) found a range 0.097–0.283 in G_{st} values for the TPA 25, APO, ACE, and PV92 Alu insertions. In contrast, the G_{st} values in this report were slightly lower, presumably due to the fact that a number of closely related populations of European descent were analyzed (as

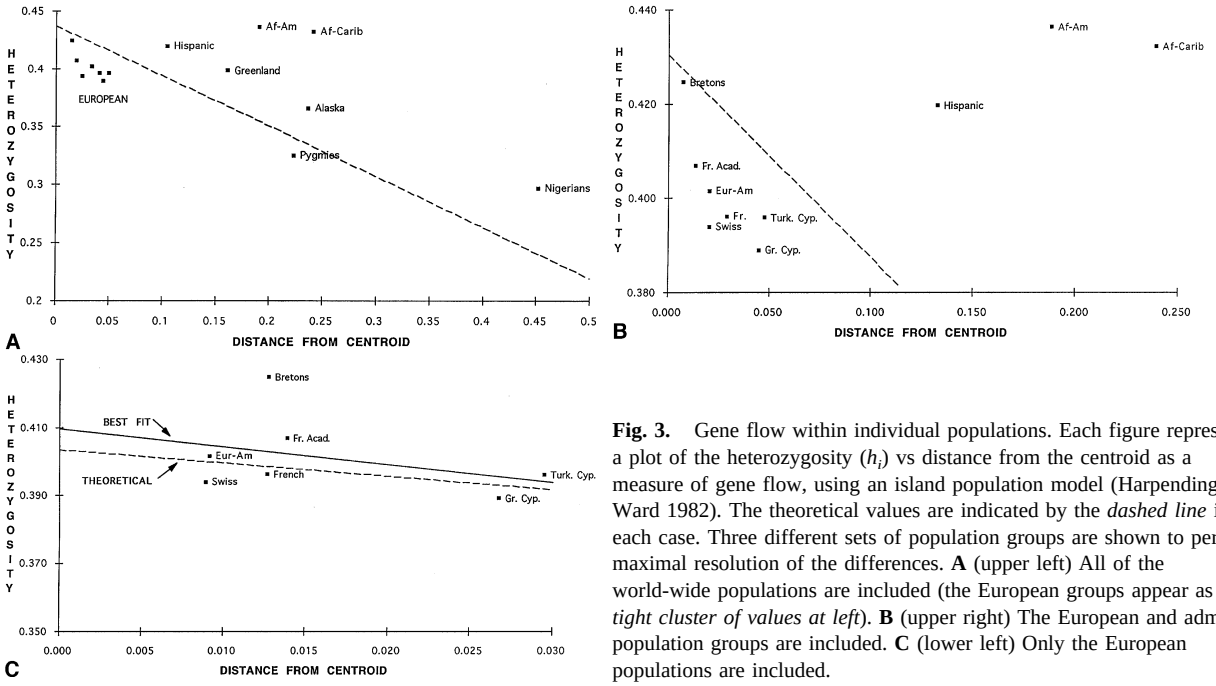

Fig. 3. Gene flow within individual populations. Each figure represents a plot of the heterozygosity (h_i) vs distance from the centroid as a measure of gene flow, using an island population model (Harpending and Ward 1982). The theoretical values are indicated by the *dashed line* in each case. Three different sets of population groups are shown to permit maximal resolution of the differences. **A** (upper left) All of the world-wide populations are included (the European groups appear as the *tight cluster of values at left*). **B** (upper right) The European and admixed population groups are included. **C** (lower left) Only the European populations are included.

discussed below). By way of comparison with other nuclear DNA markers, a survey of 42 bi-allelic polymorphisms by Bowcock et al. (1987) found 23 F_{st} values (similar to G_{st} values) of 0.097 or greater, and only four F_{st} values exceeding 0.283. Therefore, the six polymorphic Alu insertions reported here show a considerable amount of interpopulation differentiation.

The relationships between populations are shown by the topology of the maximum likelihood tree. The structure of the tree consists of five main branches: African (two branches), African-American/British Afro-Caribbean, Europeans, and Hispanic/Amerind. This structure is consistent with what is known about the history of these populations, in that populations from the same continent tend to cluster, and admixed populations tend to be intermediate in position between the source populations: The Hispanic group is on a branch that leads from the Europeans out to the two Amerind populations at the termini of the branch, while the African-Americans and British Afro-Caribbeans reside on a branch that is between the cluster of European groups and the Africans. The admixed nature of Hispanic, African-American, and British Afro-Caribbean groups was also apparent from the plot of heterozygosity vs distance from the centroid (Fig. 3); these three groups consistently had higher heterozygosities than predicted, indicative of the greater level of gene flow experienced by these groups. The placement of the African-Americans is consistent with previous estimates that the African-American gene pool is comprised of 10–30% European alleles (Chakraborty et al. 1992). The present analysis places the African-Americans closer to Africa than a previous study involving only four Alu insertions (Batzer et al. 1994); presumably the increased number of loci in the present study is leading to a more accurate placement of this admixed population on the tree.

The root of the maximum likelihood tree of population relationships resides within Africa near the Nigerians. The placement of the root involves a very minimal number of assumptions in comparison to the placement of roots along trees derived from other classical polymorphic markers. This suggests that these polymorphic Alu insertions arose within Africa and is consistent with a previous study of four recent Alu insertions (Batzer et al. 1994) as well as a study of a polymorphic Sb2 Alu repeat located on the Y chromosome (Hammer 1994). In addition, these data support the African-origin hypothesis for modern humans. The placement of the ancestral human population in Africa compares favorably with a number of previous studies involving nuclear DNA (Bowcock et al. 1991, 1994; Wainscoat et al. 1986; Batzer et al. 1994), mitochondrial DNA (Merriweather et al. 1991; Cann et al. 1987; Vigilant et al. 1991), and protein markers (Cavalli-Sforza et al. 1988; Nei and Roychoudhury 1993). Therefore, these data provide additional, compelling support for the African origin of modern humans.

The amount of variation within each Alu insertion was quite large between European and non-European population groups. However, the variation between the groups of European origin was much smaller, as indicated by the lack of significant heterogeneity in allele frequencies, the tight clustering of the European populations on one branch of the maximum likelihood tree, and the good fit between the observed and predicted relationship between the heterozygosity and distance of each population from the centroid. These data compare favorably with previous studies that suggest that Europeans

are of relatively recent origin, and hence show relatively small genetic differences (Cann et al. 1987; DiRienzo and Wilson 1991; Cavalli-Sforza et al. 1993; Piazza 1993; Torroni et al. 1994).

In conclusion, this study supports the utility of polymorphic Alu insertions for the accurate dissection of the evolutionary history of individual population groups. The mobilization and insertion of new Alu elements is a constant process (Wallace et al. 1991; Vidaud et al. 1993). The insertion of new Alu repeats at different times during human evolutionary history provides different amounts of information about populations: Alu repeats may be unique to an individual (Wallace et al. 1991; Vidaud et al. 1993), family (Muratani et al. 1991), or species or genus of primates (Batzer et al. 1993; Deininger and Batzer 1995). Thus, Alu repeats of different evolutionary ages provide different perspectives on the evolutionary history of humans; as Alu repeats continue to expand within the human genome, so does the information they provide on human population relationships.

Acknowledgments. We thank L.L. Cavalli-Sforza, J. Wainscoat, and J.-P. Moisan for samples. J.W.P. was the recipient of an Associated Western Universities student fellowship. 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., (RR 08205) to R.J.H., and (HL-42082) to W.D.S.; 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.

References

Ausabel FM, Brent R, Kingston RE, More DD, Seidman JG, Smith JA, Struhl K (eds) (1987) Current protocols in Molecular Biology. John Wiley and Sons, New York

Bailey AD, Shen C-KJ (1993) Sequential insertion of Alu family repeats into specific genomic sites of higher primates. Proc Natl Acad Sci USA 90:7205–7209

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, Hoppens 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 PL (1994) African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91:12288–12292

Batzer MA, Rubin CM, Hellman-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW (1995) Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. J Mol Biol 247:418–427

Batzer MA, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Boudreau DA, Scheer WD, Herrera RJ, Stoneking M, Deininger PL (1993) Alu repeats as markers for human population genetics. In: Proceedings of the fourth international symposium on human identification. Promega Publishing, Madison, WI, pp 49–57

Bowcock AM, Bucci C, Hebert JM, Kidd JR, Kidd KK, Friedlaender JS, Cavalli-Sforza, LL (1987) Study of 47 DNA markers in five populations from four continents. Gene Geography 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

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Britten RJ, Baron WF, Stout DB, Davidson EH (1988) Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci USA 85:4770–4774

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Evolution 32:550–570

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci USA 85:6002–6006

Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. Science 259:639–646

Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE (1992) Caucasian genes in American blacks: new data. Am J Hum Genet 50: 145–155

Deininger PL (1989) SINEs: short interspersed repeated DNA elements in higher eucaryotes. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 619–636

Deininger PL, Batzer MA (1993) Evolution of Retroposons. Evol Biol 27:157–196

Deininger PL, Batzer MA (1995) SINE master genes and population biology. In: Maraia RJ (ed) The impact of short interspersed elements (SINEs) on the host genome. RG Landes, Georgetown, TX, pp 43–60

Deininger PL, Slagel VK (1988) Recently amplified Alu family members share a common parental Alu sequence. Mol Cell Biol 8:4566–4569

Deininger PL, Batzer MA, Hutchison III CA, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311

DiRienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88: 1597–1601

Edwards MC, Gibbs RA (1992) A human dimorphism resulting from loss of an Alu. Genomics 14:590–597

Felsenstein J (1981) Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution 35:1229–1242

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761

Harpending HC, Ward RH (1982) Chemical systematics and human populations. In: Nitecki M (ed) Biochemical aspects of evolutionary biology. University of Chicago Press, Chicago, pp 213–256

Hutchinson GB, Andrew SE, McDonald H, Goldberg YP, Graharn R, Rommens JR, Hayden MR (1993) An Alu element retroposition in two families with Huntington disease defines a new active Alu subfamily. Nucleic Acids Res 21:3379–3383

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

Jurka J, Milosavljevic A (1991) Reconstruction and analysis of human Alu genes. J Mol Evol 32:105–121

Jurka J, Smith T (1988) A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci USA 85:4775–4778

Karathanasis SK (1985) Apolipoprotein multigene family: tandem or-

ganization of apolipoprotein AI, CIII, and AIV genes. Proc Natl Acad Sci USA 82:6374–6378

Kass DH, Batzer MA, Deininger PL (1995) Gene conversion as a secondary mechanism of SINE evolution. Mol Cell Biol 15:19–25

Kass DH, Aleman C, Batzer MA, Deininger PL (1994) Identification of a human specific Alu insertion in the Factor XIIIB gene. Genetica 94:1–8

Leeflang EP, Liu W-M, Hashimoto C, Choudary PV, Schmid CW (1992) Phylogenetic evidence for multiple Alu source genes. J Mol Evol 35:7–16

Leeflang EP, Liu WM, Chesnokov IN, Schmid CW (1993) Phylogenetic isolation of a human Alu founder gene: drift to new subfamily identity. J Mol Evol 37:559–565

Matera AG, Hellmann U, Schmid CW (1990a) A transpositionally and transcriptionally competent Alu subfamily. Mol Cell Biol 10:5424–5432

Matera AG, Hellmann U, Hintz MF, Schmid CW (1990b) Recently transposed Alu repeats result from multiple source genes. Nucleic Acids Res 18:6019–6023

Merriweather DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry S T, Wallace DW (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33:543–555

Monson KL, Moisan JP, Pascal O, McSween M, Aubert D, Giusti A, Budowle B, Lavergne L (in press) Description and analysis of alleles distribution for four VNTR markers in French and French Canadian populations. Hum Hered

Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K (1991) Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. Proc Natl Acad Sci USA 88:11315–11319

Nei M (1972) Genetic distance between populations. Am Naturalist 106:283–292

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. Mol Biol Evol 10:927–943

Okada N (1991) SINEs. Curr Opin Genet Dev 1:498–504

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

Piazza A (1993) Who are the Europeans? Science 260:1767–1769

Quentin Y (1988) The Alu family developed through successive waves of fixation closely connected with primate lineage history. J Mol Evol 27:194–202

Reynolds JB, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779

Rogers J (1983) Retroposons defined. Nature 301:460

Sawada I, Schmid CW (1986) Primate evolution of the α-globin gene cluster and its Alu-like repeats. J Mol Biol 192:693–703

Sawada I, Willard C, Shen C-KJ, Chapman B, Wilson AC, Schmid CW (1985) Evolution of Alu family repeats since the divergence of human and chimpanzee. J Mol Evol 22:316–322

Schmid CW, Maraia R (1992) Transcriptional regulation and transpositional selection of active SINE sequences. Curr Opin Genet Dev 2:874–882.

Shen M-R, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). J Mol Evol 33:311–320

Slagel V, Flemington E, Traina-Dorge V, Bradshaw H, Deininger PL (1987) Clustering and sub-family relationships of the Alu family in the human genome. Mol Biol Evol 14:19–29

Tiret L, Riget B, Visvikis S, Breda C, Corvol P, Cambien F, Soubrier F (1992) Evidence from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51:197–205

Torroni A, Lott MT, Cabell MF, Chen Y-S, Lavergne L, Wallace DC (1994) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet 55:760–776

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, Sanchez SG, Laurin Y, Meyer D, Goossens M, Lavergne JM (1993) Hemophilia B due to a *de novo* insertion of a Human-Specific Alu subfamily member within the coding region of the factor IX gene. Eur J Hum Genet 1:30–36

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

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, Lynch JR, Falusi AG, Weatherall DJ, Clegg JB (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491–493

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

Webb GC, Coggan M, Ichinose A, Board PG (1989) Localization of the coagulation factor XIII B subunit gene (F13B) to chromosome bands 1q31–32.1 and restriction fragment polymorphism at the locus. Hum Genet 81:157–160

Willard C, Nguyen HT, Schmid CW (1987) Existence of at least three distinct Alu subfamilies. J Mol Evol 26:180–186

Yang-Feng TL, Opdenakker G, Volckaert G, Franke U (1986) Human tissue-type plasminogen activator gene located near chromosomal breakpoint in myeloproliferative disorder. Am J Hum Genet 39: 79–87

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