Am. J. Hum. Genet. 68:738–752, 2001

Patterns of Ancestral Human Diversity: An Analysis of *Alu*-Insertion and Restriction-Site Polymorphisms

W. S. Watkins,[1] C. E. Ricker,[1] M. J. Bamshad,[1,2] M. L. Carroll,[4] S. V. Nguyen,[4] M. A. Batzer,[4,5] H. C. Harpending,[3] A. R. Rogers,[3] and L. B. Jorde[1]

Departments of [1]Human Genetics and [2]Pediatrics, University of Health Sciences Center, and [3]Department of Anthropology, University of Utah, Salt Lake City; [4]Department of Pathology, Stanley S. Scott Cancer Center, Neuroscience Center of Excellence, Louisiana State University Health Sciences Center, New Orleans; and [5]Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, Baton Rouge

We have analyzed 35 widely distributed, polymorphic *Alu* loci in 715 individuals from 31 world populations. The average frequency of *Alu* insertions (the derived state) is lowest in Africa (.42) but is higher and similar in India (.55), Europe (.56), and Asia (.57). A comparison with 30 restriction-site polymorphisms (RSPs) for which the ancestral state has been determined shows that the frequency of derived RSP alleles is also lower in Africa (.35) than it is in Asia (.45) and in Europe (.46). Neighbor-joining networks based on *Alu* insertions or RSPs are rooted in Africa and show African populations as separate from other populations, with high statistical support. Correlations between genetic distances based on *Alu* and nuclear RSPs, short tandem-repeat polymorphisms, and mtDNA, in the same individuals, are high and significant. For the 35 loci, *Alu* gene diversity and the diversity attributable to population subdivision is highest in Africa but is lower and similar in Europe and Asia. The distribution of ancestral alleles is consistent with an origin of early modern human populations in sub-Saharan Africa, the isolation and preservation of ancestral alleles within Africa, and an expansion out of Africa into Eurasia. This expansion is characterized by increasing frequencies of *Alu* inserts and by derived RSP alleles with reduced genetic diversity in non-African populations.

Introduction

The evolution of modern human populations continues to be a topic of controversy. Evidence from mtDNA, Y-chromosome polymorphisms, autosomal markers, and fossil material supports both the expansion of early modern human populations in Africa and the partial or complete replacement of other hominid groups (Cann et al. 1987; Vigilant et al. 1991; Stoneking et al. 1997; Harpending et al. 1998; Jorde et al. 1998; Krings et al. 1999; Relethford and Jorde 1999; Seielstad et al. 1999; Ovchinnikov et al. 2000). Interpretation of other, primarily fossil, data suggests both an early expansion of hominid lines and multiregional development of modern humans (Hawks et al. 2000; Wolpoff et al. 2000). The increasing number and variety of genetic markers offer additional opportunities for more-detailed analysis of human evolution and of genetic diversity within and between human populations.

Numerous studies utilizing a variety of polymorphic loci suggest an overall pattern of higher gene diversity in African populations compared with that in non-African populations (Merriwether et al. 1991; Vigilant et al. 1991; Bowcock et al. 1994; Deka et al. 1995; Jorde et al. 1997; Stoneking et al. 1997; Kaessmann et al. 1999; Seielstad et al. 1999; Forster et al. 2000; Jorde et al. 2000). These studies have focused primarily on neutral restriction-site polymorphisms (RSPs), short tandem-repeat polymorphisms (STRPs), noncoding autosomal sequences, Y chromosomes, and mtDNA. Analyses of protein-coding regions, including the neurofibromatosis type 1 (NF1), angiotensin-converting enzyme (ACE), myotonic dystrophy (DM), dopamine D2 receptor (DRD2) and fragile X (FMR1) loci, have also shown higher levels of diversity in African populations compared with levels found in non-African populations (Purandare et al. 1996; Kidd et al. 1998; Tishkoff et al. 1998; Rieder et al. 1999; Crawford et al. 2000). Some loci, including the melanocortin 1 receptor (MCR1) and phenylalanine hydroxylase (PAH) loci, do not consistently show patterns of higher diversity in African populations, revealing the potential influence of natural selection on patterns of genetic diversity (Rana et al. 1999; Harding et al. 2000; Kidd et al. 2000). Noncoding DNA sequences on chromosomes 22, 15, and 1 show higher nucleotide-diversity estimates for African populations than for non-African populations, consistent with a re-

Received November 9, 2000; accepted for publication January 17, 2001; electronically published February 16, 2001.

Address for correspondence and reprints: Dr. Lynn B. Jorde, Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT 84112. E-mail: lbj@genetics.utah.edu

0002-9297/2001/6803-0018$02.00

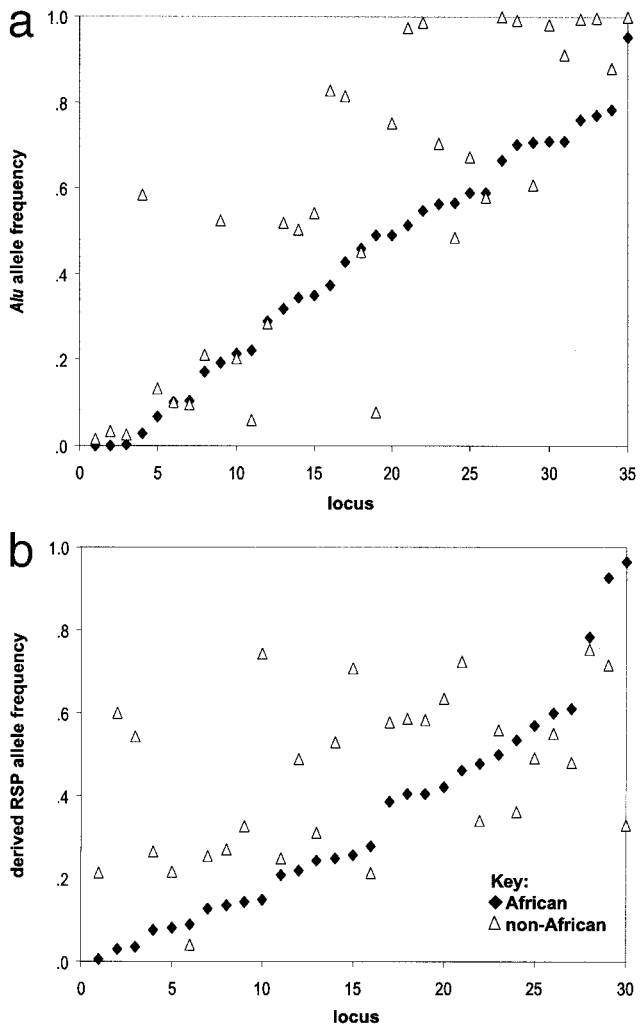

Figure 1 Allele frequencies of the derived alleles at 35 polymorphic *Alu* (*a*) and 30 RSP (*b*) loci, in African and non-African populations, sorted by the insertion frequency in the African population. Note the trend toward higher *Alu*-insertion frequencies in populations outside Africa, whereas ancestral allele frequencies are higher in African populations. A similar trend is observed for 30 derived RSP alleles.

cent human population expansion (Zhao et al. 2000; Yu et al. 2001; L. B. Jorde, W. S. Watkins, M. J. Bamshad, D. Dunn, and R. B. Weiss, unpublished data).

Alu-insertion polymorphisms are ideal markers for human evolutionary studies because retroposition produces infrequent, irreversible, widely distributed insertion events, each with a known ancestral state (Batzer et al. 1994, 1996; Stoneking et al. 1997; Melton et al. 1998; Novick et al. 1998). These insertions are short interspersed repetitive elements (SINEs), and they account for ⩾10% of the human genome (Deininger and Batzer 1993; Smit 1996; Dunham et al. 1999). Each full-length *Alu* element is a dimeric, ~300-bp retroposon that is homologous to the 7SL RNA component of the signal-recognition particle. The Y, Ya, and Yb subfamilies of *Alu* elements are still active and produce new *Alu* insertions that are polymorphic in human populations (Arcot et al. 1995*a;* Batzer et al. 1995, 1996). Since the Y, Ya, and Yb subfamilies did not become active until after the human lineage diverged from the last common ancestor with the nonhuman apes, insertion events of these *Alu*s are restricted to humans, and the ancestral allele is the absence of an insertion. Assignment of the ancestral allele for polymorphic *Alu* loci does not require comparisons with other species such as chimpanzees and provides a unique opportunity for the rooting of phylogenetic networks. At each polymorphic locus, the chromosome containing the *Alu* insert is the derived allele, and different individuals with an *Alu* insert at a given locus share a chromosomal region that is identical by descent. Previous studies using a small number of polymorphic *Alu* insertions have suggested an African origin for modern humans and have argued for an earlier, more extensive expansion of modern humans in tropical regions (Batzer et al. 1994, 1996; Stoneking et al. 1997).

In the present study, we examine 35 *Alu* loci and 30 gene-related RSPs in 31 world populations, to characterize diversity and genetic structure in modern human populations. The ancestral state for each *Alu* and RSP locus has been assigned. This analysis of the largest set of *Alu*-insertion polymorphisms examined to date provides new data on the distribution of ancestral and derived alleles in human populations and is consistent with an African origin of anatomically modern humans. Genetic-distance estimates based on *Alu*-insertion polymorphisms, RSPs, STRPs, and mtDNA, using the same individuals in the same populations, are highly correlated. These data build upon a growing body of evidence characterizing human genetic diversity and population structure.

Subjects, Material, and Methods

Human DNA Samples

The human populations used in this study have been described elsewhere (Jorde et al. 1995, 1997; Bamshad et al. 1998; Watkins et al. 1999). The continental groups, populations, and sample sizes are Africans (155 total)—Alur (12), Biaka Pygmy (5), Hema (18), Mbuti Pygmy (Coriell) (5), Nande (17), Nguni (14), San (15), Sotho/Tswana (22), Tsonga (14), and Zaire Pygmy (Mbuti) (33); Asians (77 total)—Cambodian (12), Chinese (17), Japanese (19), Malay (6), Vietnamese (9), and mixed Asian (14); Europeans (118 total)—Finnish (20), French (20), northern European (68), and Polish (10); and Indians (365 total)—Brahmin (60), Kapu (58),

Table 1

Population Statistics for 35 *Alu*-Insertion Polymorphisms and 30 "Rooted" RSPs

Population (*n*)	Heterozygosity (95% CI)		Observed Homozygosity for *Alu/Alu*	F_{IS}
	Expected under HWE	Observed		
Alu-insertion polymorphisms:				
African (155):				
Alur (12)	.37	.37	.28	−.01
Biaka Pygmy (5)	.29	.25	.30	.16
Hema (18)	.35	.35	.29	.01
Mbuti Pygmy (5)	.29	.27	.26	.08
Nande (17)	.35	.38	.25	−.11
Nguni (14)	.34	.32	.25	.06
San (15)	.31	.29	.25	.06
Sotho, Tswana (22)	.33	.30	.25	.08
Tsonga (14)	.35	.37	.22	−.07
Zaire Pygmy (33)	.32	.34	.24	−.06
Overall	.35 (.28–.43)	.33 (.26–.41)	.26 (.19–.32)	.05
Asian[a](77):				
Cambodian (12)	.24	.26	.46	−.10
Chinese (17)	.24	.25	.44	−.06
Japanese (19)	.25	.26	.43	−.04
Malay (6)	.30	.31	.41	−.05
Vietnamese (9)	.23	.24	.45	−.02
Overall:	.25 (.15–.34)	.25 (.16–.35)	.44 (.33–.55)	−.02
European (118):				
Finnish (20)	.21	.21	.46	.02
French (20)	.21	.22	.46	−.02
Northern European (68)	.21	.20	.45	.06
Polish (10)	.19	.19	.46	.01
Overall:	.22 (.14–.29)	.21 (.13–.28)	.46 (.37–.51)	.05
Indian (365):				
Brahmin (60)	.25	.22	.42	.09
Irula[b] (34)	.27	.29	.41	−.06
Kapu (58)	.26	.24	.42	.05
Khonda Dora[b] (27)	.24	.25	.41	−.04
Kshatriya (11)	.26	.26	.43	.00
Maiga (29)	.25	.25	.42	−.01
Mala (26)	.26	.26	.42	.01
Maria Gond[b] (22)	.25	.22	.41	.15
Relli (19)	.27	.27	.41	.01
Santal[b] (16)	.24	.22	.42	.09
Vysya (10)	.23	.21	.44	.09
Yadava (53)	.25	.26	.44	−.03
Overall	.26 (.22–.31)	.25 (.20–.29)	.42 (.37–.47)	.04
			Observed Homozygosity for 1/1[c]	
"Rooted" RSPs:				
African (75):				
Biaka Pygmy (5)	.33	.35	.18	−.07
Mbuti Pygmy (5)	.24	.21	.21	.15
Nguni (14)	.34	.33	.18	.04
San (15)	.28	.25	.21	.13
Sotho, Tswana (22)	.33	.33	.19	.00
Tsonga (14)	.34	.38	.17	−.12
Overall	.32 (.22–.43)	.31 (.21–.42)	.19 (.11–.29)	.03

(*continued*)

Table 1 *Continued*

Population (*n*)	Heterozygosity (95% CI)		Observed Homozygosity for 1/1[c]	F_{IS}
	Expected under HWE	Observed		
Asian[a] (78):				
Cambodian (12)	.40	.38	.28	.06
Chinese (17)	.38	.34	.25	.12
Japanese (19)	.36	.36	.26	.00
Malay (6)	.40	.40	.28	−.02
Vietnamese (9)	.36	.36	.28	.02
Overall	.38 (.27–.47)	.36 (.25–.47)	.27 (.17–.37)	.05
European (121):				
Finnish (20)	.43	.46	.24	−.07
French (20)	.44	.41	.26	.08
Northern European (70)	.43	.43	.25	.00
Polish (11)	.42	.43	.21	−.03
Overall	.43 (.35–.52)	.43 (.34–.52)	.25 (.17–.32)	.01

[a] 14 and 15 individuals of mixed Asian ancestry omitted from subpopulations.

[b] Tribal sample.

[c] Allele 1 represents the derived allele.

Kshatriya (11), Madiga (29), Mala (26), Relli (19), Vysya (10), Yadava (53), Irula (34), Khonda Dora (27), Maria Gond (22), and Santal (16). DNA was extracted from blood lymphocytes or cell lines by standard phenol/chloroform-based or salt-based extraction methods (Puregene) and was suspended in 10 mM Tris, 0.1 mM EDTA for *Alu* genotyping. This research has been approved by the institutional review board at the University of Utah.

Ascertainment and Typing of Alu *Markers*

The *Alu* loci used in this study were ascertained by direct sequencing, library screening, or database searching, as described elsewhere (Roy et al. 1999, 2000). The cell lines used initially to isolate and characterize human-specific *Alu* insertions were human HeLa (ATCC CCL2), chimpanzee (*Pan troglodytes*) Wes (ATCC CRL1609), and gorilla (*Gorilla gorilla*) Ggo-1 (primary gorilla fibroblasts) (Arcot et al. 1995*b*). For library-ascertained *Alu* loci, genomic phage libraries, constructed from HeLa or other genomic DNA, were screened with [^{32}P]-labeled oligonucleotides specific for Ya5 or Ya8 *Alu* sequences (Arcot et al. 1995*b*, 1998). Purified phage DNA was subcloned and sequenced by standard methods. To assess heterozygosity, primers flanking the *Alu* insert were used to amplify a test panel of 80 individuals (160 total chromosomes) from four populations (20 African Americans, 20 Greenland natives, 20 Egyptians, and 20 Europeans). To confirm that an *Alu* insert was human specific, one common chimpanzee and one pygmy chimpanzee were included in the typing panel. All *Alu*-insertion polymorphisms used in the study were absent from orthologous positions in the genomes of nonhuman primates.

For *Alu* loci beginning with the prefix NBC, screening of nonredundant and high-throughput genomic sequence (HTGS) databases was performed through use of the Basic Local Alignment Search Tool (BLAST [GenBank]). The database was searched for exact complements to the oligonucleotide 5′-CCATCCCGGCTAAAACGGTG-3′, which is an exact match of that portion of the *Alu* Ya5-subfamily consensus sequence that contains unique diagnostic mutations. Sequences that were exact complements of the oligonucleotide were then subjected to more-detailed analysis. A region of 1,000 bases, which was directly adjacent to the sequences identified from the databases and matched the initial GenBank BLAST query, was analyzed through use of either the REPEATMASKER2 or the CENSOR program. PCR primers were designed from unique flanking DNA sequences adjacent to individual *Alu* elements, through use of PRIMER3 software (Whitehead Institute for Biomedical Research). The PCR primers were screened, against the GenBank nonredundant, Alu, and HTGS databases, for the presence of repetitive elements or duplications, through use of BLAST. The sequences of the oligonucleotide primers and PCR conditions are available from the Eccles Institute of Human Genetics Web site. The chromosomal location of *Alu* repeats identified from clones that had not been mapped previously was determined by PCR amplification of National Institute of General Medical Sciences human/rodent somatic cell hybrid mapping panel 2 (Coriell Institute for Medical Research).

Polymorphic *Alu* loci were genotyped by amplifica-

Table 2

Hierarchical *F* Statistics for 35 *Alu* Loci

	F_{IS}	F_{IT}[a]	F_{ST}[a]
World subdivided into:			
Africa, Asia, Europe, India	.038	.155	.122
31 World subpopulations	.010	.121	.112
Asia, Europe, India	.017	.064	.048
Continent subdivided into:			
Africa (10 populations)	−.005	.052	.057
Asia (5 populations)	−.055	−.036	.018
Europe (4 populations)	.033	.053	.020
India (12 populations)	.024	.047	.024

[a] Underlined values are significantly different from 0 ($P < .025$).

tion of 25 ng of genomic DNA in a standard 30-cycle, three-step PCR. Appropriate annealing temperatures and additives were optimized for each system. For most systems, samples were amplified with 5 μl of cresol red loading buffer (34% sucrose, 0.02% cresol red) to eliminate the need to add dye to the samples before gel loading. After PCR, the samples were loaded onto multiple-combed 3% Nusieve agarose (3:1) gels and were electrophoresed at 175 V for 2 h. Ethidium bromide–stained gels were visualized by UV and were documented. The methodology and genotypes for the STRP, RSP, and hypervariable segment 1 (HVS1) data have been described elsewhere (Jorde et al. 1995, 1997).

The ancestral allele of each RSP polymorphism was determined by comparison with that of a common chimpanzee and that of a pygmy chimpanzee. Human and chimpanzee DNA was amplified with primers for each of the 30 RSP loci. The locations, primer sequences, and PCR conditions for each RSP locus are provided by The Genome Database and are listed at the Eccles Institute of Human Genetics Web site. PCR products were digested with 4 units of the appropriate restriction endonuclease in PCR buffer for 2 h and then were visualized by agarose-gel electrophoresis and ethidium bromide staining. The ancestral status was assigned to the allele that was shared by the chimpanzees and humans. No polymorphic variation was seen in the chimpanzees. Of the 30 RSPs used here 11 are documented single-nucleotide polymorphisms (SNPs), whereas 19 have not been characterized at the sequence level.

Data Analysis

Alu-insertion frequencies and genotype frequencies were obtained by gene counting. The *Alu*-insertion frequencies are provided in the Appendix. Unbiased heterozygosity estimates were calculated as $h = (N/N - 1)(1 - \Sigma p_i^2)$, where N is the number of chromosomes sampled and p is the frequency of the ith allele. Evaluation of Hardy-Weinberg equilibrium (HWE) by Fisher's exact test (using a random-permutation method with a Bonferroni correction for multiple comparisons), and estimates of F statistics for all populations were obtained through use of the Genetic Data Analysis program (Lewis and Zaykin 2000). Genetic distances between populations were calculated with GENDIST (Felsenstein 1993). Between-population distances are based on an infinite-alleles model of evolution and are expressed as Nei's distance, $D = -\ln(I)$ and $I = (\Sigma p_i q_i)/\sqrt{(\Sigma p_i^2 \Sigma q_i^2)}$, where p_i and q_i are the allele frequencies of the ith allele in populations p and q, and summation occurs over all loci (Nei 1987). Neighbor-joining trees were produced through use of the NEIGHBOR program, 1,000 bootstrap replicates were generated with SEQBOOT, and a consensus tree was built with CONSENSE as implemented in the PHYLIP program package (Felsenstein 1993; PHYLIP Home Page). Correlations between the genetic-distance matrices and their associated significance levels were obtained through use of the Mantel test (100,000 permutations) (Smouse et al. 1986).

Results

Of the 35 polymorphic *Alu* insertions, 33 are present in all continental groups. Ya5NBC54 and Ya5NBC135 insertions were not found in African populations. No *Alu* insertion was fixed in African populations, but two, five, and six insertions have achieved fixation in Indian, Asian, and European populations, respectively. Most *Alu* loci are in HWE. For the four major population groups, five, four, five, and nine loci produced significant values ($P < .05$) in African, Asian, European, and Indian populations, respectively, but these values fell to two, four, two, and three when a Bonferroni correction was applied. The loci that deviated from HWE differed from population to population. Ten *Alu* polymorphisms were examined for Mendelian segregation in two three-generation CEPH families. Normal codominant segregation patterns were observed (data not shown).

Figure 1*a* shows the *Alu* allele-frequency profile for African and non-African populations. This plot displays the frequencies of the *Alu* inserts at each locus, sorted according to their frequency in the African population. Two patterns are discernible: (1) African populations show an overall trend toward lower *Alu*-insertion frequencies, and (2) loci that have a low (<.3) frequency in non-Africans also have low frequencies in Africans, whereas loci with higher frequencies in non-Africans tend to have lower frequencies in Africans. These patterns were also observed when the continental groups were divided into subpopulations. African subpopulations showed trends toward lower *Alu*-insertion frequencies, whereas the highest insertion frequencies were found in southeastern Asian populations. The mean *Alu*-insertion frequencies for the major population groups are as follows: Africans, .4226; Asians, .5690; Europeans, .5604; and Indians, .5453. Allele-frequency distributions in the four major population groups are

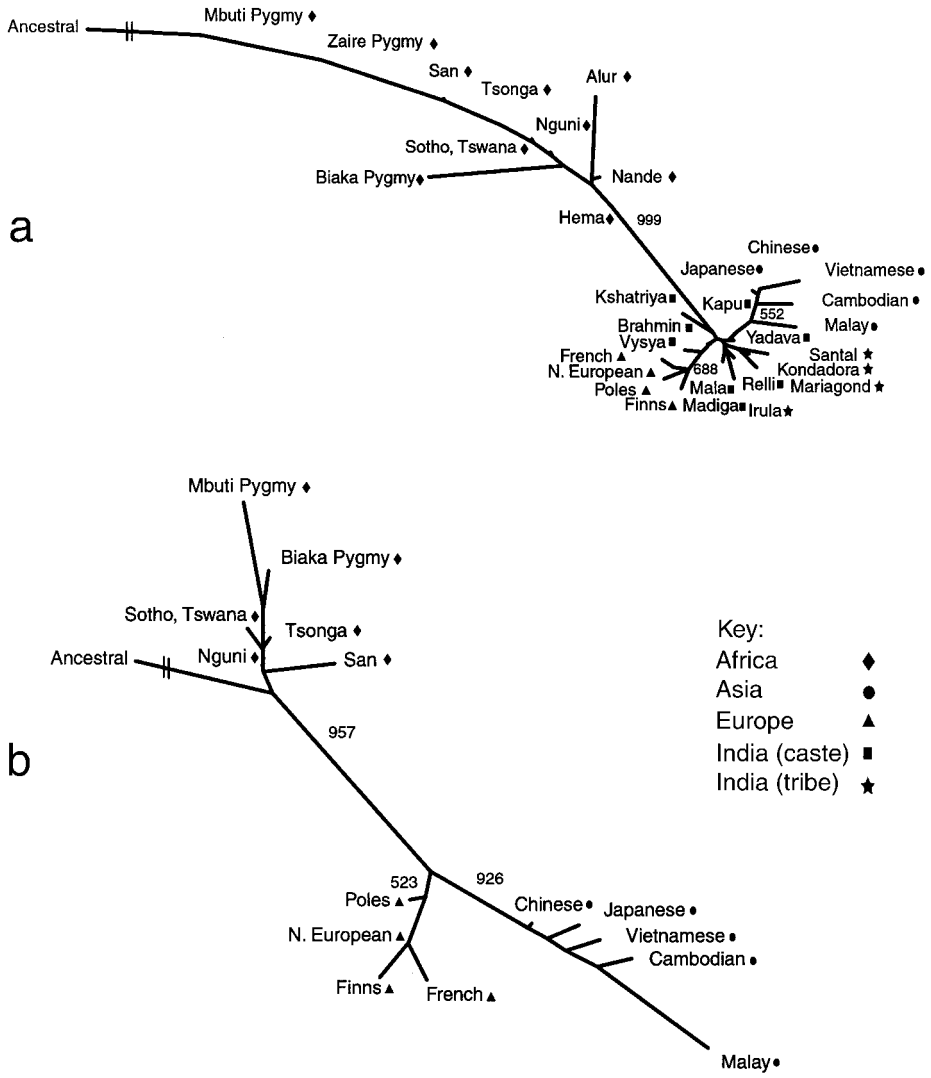

Figure 2 Neighbor-joining networks of genetic distances, based on 35 polymorphic *Alu* loci (*a*) and 30 RSP loci (*b*). African Mbuti Pygmy populations are nearest to the ancestral outgroup. African populations are clustered together, with long branches. All African populations and the ancestral population are separated from the Asian, European, and Indian populations, with very high bootstrap support (999/1,000). Asian and European populations cluster into their respective continents, with moderate support (552/1,000 and 688/1,000). Indian populations are distributed in the non-African portion of the network. The RSP network based on derived alleles places an ancestral population nearest to African populations, and these populations are separated from non-African populations, with very high bootstrap support (957/1,000). Populations from southeastern Asia have the greatest genetic distance from the ancestral population.

significantly different between Africans and Asians ($P < .001$, by a Wilcoxon signed-rank test), Africans and Europeans ($P < .003$), and Africans and Indians ($P < .002$) but are not significant different between non-African groups.

We also evaluated the frequency of *Alu* insertions in a data set, published elsewhere (Stoneking et al. 1997), consisting of eight *Alu* loci in 34 independently ascertained world populations. Average *Alu*-insertion frequencies over all loci in this data set show a similar trend toward an increasing frequency of *Alu* insertions in populations outside Africa (Africa, .374; western Asia, .435; Europe, .464; southeastern Asia, .529; and Americas, .543). Dividing the polymorphic *Alu* insertions into groups based on the method of ascertainment (directed HeLa library screening [Arcot et al. 1995*b*, 1998], random library screening, or *in silico* database mining) showed lower *Alu*-insertion frequencies in Africans, for all ascertainment methods.

To compare ancestral-allele and derived-allele patterns based on *Alu* elements versus those seen for other polymorphic markers, the derived-allele–frequency profiles for 30 randomly chosen, gene-related, unlinked RSPs were examined in a subset of 274 individuals (Jorde et al. 1995). The ancestral allele for each of these previously published loci was determined by typing the

Table 3

Correlations between Genetic Distance Matrices

	Alu	RSPs	STRPs
RSPs (autosomal)	.850		
STRPs (autosomal)	.789	.750	
HVS1 (mtDNA)	.816	.677	.681

NOTE.—$P < .0001$ for all comparisons.

system in chimpanzees, and new analyses were performed with the "rooted" RSP data. The allele-frequency profiles for the 30 RSPs in African, Asian, and European populations show trends toward higher frequencies of derived alleles in Asia and Europe and lower frequencies of derived alleles in Africa (fig. 1*b*). Thus, through use of either widely distributed *Alu* markers or RSP markers from the same individuals, an increase in the overall frequencies of derived alleles can be observed in continental populations outside Africa.

Alu-heterozygosity and -homozygosity estimates for the continental groups and their subpopulations are consistent with the *Alu*-frequency distributions (table 1*a*). African populations have higher overall *Alu* gene diversity (heterozygosity) than do other populations. This finding is concordant with results from other marker systems, including the mitochondrial genome, autosomal STRPs, and the Y chromosome (Deka et al. 1995; Jorde et al. 2000). The observed homozygosity for the *Alu* insert (the derived state) is significantly higher ($P < .05$) in all non-African groups than it is in sub-Saharan Africans. All subpopulations outside Africa have higher observed levels of *Alu* homozygosity, as averaged over the 35 *Alu* loci, than do all populations within Africa. Comparisons of the major population groups and their subpopulations, through use of rooted RSP loci, show a similar trend, outside Africa, toward increasing frequency and homozygosity of derived alleles (table 1*b*). Despite the high levels of heterozygosity in Europe that are caused by ascertainment bias (Bowcock et al. 1991; Rogers and Jorde 1996), observed levels of RSP homozygosity of derived alleles, over all loci, is lowest in Africa and its subpopulations, higher in Europe, and highest in populations of southeastern Asia.

Using hierarchical *F* statistics, we examined the effect of population subdivision of individuals with respect to the total population (F_{IT}) and of subdivision of the subpopulations with respect to the total population (F_{ST}), at the worldwide and continental levels, for the 35 *Alu* loci (table 2). Dividing the total population into the four major population groups and into 31 subpopulations yields statistically significant F_{ST} estimates of 12.2% and 11.2%, respectively. Omitting Africans from the continental comparison reduces the F_{ST} to 4.8%. The F_{ST} for populations within Africa is 5.7%, more than twofold greater than those for populations within Asia (1.8%), Europe (2.0%), and India (2.4%). The F_{ST} for the five populations within Asia is not significantly different from 0. For most subpopulations with larger sample sizes, the fixation index of the individual with respect to its own subdivision (F_{IS}) is relatively low, for both *Alu* and RSP markers (shown in table 1).

A neighbor-joining tree of *Alu* genetic distances illustrates the relationships among the 31 world populations (fig. 2*a*). A hypothetical population with an *Alu* allele frequency of 0 for each locus is included as an "ancestral" population (Batzer et al. 1994). The ancestral population is located closest to the Mbuti and Zaire Pygmy populations. African populations are closer to the ancestral population than are non-African populations. The ancestral population and all African populations are separated from all non-African populations, with very high bootstrap support. The populations of Asia and Europe cluster by continent, and Indian populations are dispersed among non-African groups. Tribal Indians of close geographic proximity to one another (Maria Gond, Khonda Dora, and Santal) cluster together. The network shows greater separation among the 10 African populations than among Asian, European, and Indian populations.

A neighbor-joining network based on the 30 RSPs shows similar relationships among a subset of the populations shown in the *Alu* network (fig. 2*b*). A derived–RSP-allele frequency (0) at each locus was used to create the ancestral population. The ancestral population for the RSP network falls very close to the African cluster. African populations branch off first, and bootstrap support for the non-African cluster is very high. Populations from Asia and Europe cluster into their respective continents. The populations of southeastern Asia have the greatest distance from the ancestral population, which reflects a higher overall frequency of derived RSP alleles in Vietnamese, Cambodian, and aboriginal Malay populations.

To compare genetic distances based on *Alu* loci versus those based on other marker systems, we created four matrices of genetic distances for 15 populations from Africa, Asia, and Europe, based on 35 *Alu* markers, 30 STRPs, 30 RSPs, and the mtDNA control-region HVS1. This subset of 15 populations was selected for the analysis because it contains data for all populations, loci, and marker types. Each matrix represents the between-population genetic distances derived from the same 245 individuals from 15 populations, for a given genetic system. Correlation coefficients between *Alu* distances and distances based on other marker systems are high and significant ($P < .0001$), with the highest correlations seen between the *Alu* and RSP systems (table 3).

Discussion

Both the higher frequency of ancestral *Alu* loci and the higher genetic diversity in Africans are consistent with (1) the emergence of modern *Homo sapiens* within Africa, (2) the exodus and rapid expansion of a limited subset of humans into Eurasia, and (3) limited migration back into Africa. Higher African diversity has been observed for many neutral marker systems, including those for autosomal RSPs and STRPs (Bowcock et al. 1991, 1994; Deka et al. 1995; Jorde et al. 1995, 1997), mtDNA (Merriwether et al. 1991; Vigilant et al. 1991; Penny et al. 1995), and the Y chromosome (Seielstad et al. 1999; Forster et al. 2000). The diversity patterns at several genes, including ACE (Rieder et al. 1999), DM (Tishkoff et al. 1998), and the plasma alpha (1,3) fucosyltransferase gene (FUT6) (Pang et al. 1999), among others, also show high diversity in African populations. In addition to high African diversity, our analysis shows a directional increase in the overall frequency and homozygosity of *Alu* and derived RSP alleles. This finding is consistent with a rapid expansion and/or a bottleneck in early populations leaving Africa. A continued expansion of populations into Asia and southeastern Asia may account for the higher frequency of derived alleles in these populations. Additional work, using tightly linked markers flanking each *Alu* insert, will provide relative ages of the *Alu* insertions, which will allow better resolution of the potential mechanisms generating these patterns.

The F_{ST} estimates based on the 35 *Alu*-insertion polymorphisms for the four major population groups are also consistent with an African origin and rapid expansion. The continental-level *Alu* F_{ST} of 12.2% is similar to previous estimates made on the basis of autosomal RSPs and STRPs (Bowcock et al. 1991; Deka et al. 1995; Barbujani et al. 1997; Jorde et al. 2000), but removal of the African populations reduces the F_{ST} to 4.8%. Thus, substantially more *Alu* variation occurs between African and non-African continental populations than occurs between all non-African populations. Low F_{ST} estimates for European and Asian populations are consistent with rapid and recent expansion with relatively little time for population differentiation.

The *Alu*-insertion polymorphism and the RSP consensus neighbor-joining networks place the ancestral population among and near Africans, respectively. This result is consistent with the allele frequencies for both *Alu* and RSP markers and reflects a higher frequency of ancestral alleles in African populations. Recent studies of SNPs indicate that the level of sequence-verified polymorphism sharing between chimpanzees or gorillas and humans is low (<1%) (Hacia et al. 1999), which suggests that rooted trees based on RSP/SNP are subject to only limited error due to convergent mutations. Using data sets for RFLPs, microsatellite data, protein polymorphisms, and a limited number of *Alu* loci, Nei and Takezaki (1996) consistently place a root, for human phylogenetic trees, between African and non-African populations. Despite initial methodological concerns, a substantial body of evidence indicates a root, in African populations, for the maternally inherited mtDNA tree (Cann et al. 1987; Vigilant et al. 1991; Stoneking and Soodyall 1996; Ingman et al. 2000). Additionally, a human mtDNA "molecular fossil" inserted into chromosome 11 has been used as an independent outgroup for the rooting of mtDNA trees (Zischler et al. 1995). The topology of the tree, although statistically weak, suggests an African origin. Like *Alu*-insertion frequencies, the insertion frequency of the mtDNA "fossil" is low in Africans and higher in non-Africans. Recent studies using independent sets of STRP and biallelic markers on the Y chromosome suggest that the earliest branches for male founder(s) of our species can be traced to ancestors of the African San population (Forster et al. 2000). In the *Alu* and RSP networks, the San individuals examined in this analysis fall near the ancestral populations.

The eight *Alu* loci reported by Stoneking et. al. (1997) show a pattern of increasing *Alu*-insertion frequencies in populations outside Africa, with the highest insertion frequencies being found in American Indians. An exception to this trend was a low frequency of inserts in the aboriginal populations of Papua New Guinea and Australia. Using a principal-components analysis, those authors suggest an early tropical expansion of modern humans, because these populations' distance to the root is similar to that of Africans. The tribal populations of the Indian subcontinent may also represent remnants of an early Paleolithic expansion out of Africa (Cavalli-Sforza et al. 1994). However, the Irula, Khonda Dora, Maria Gond, and Santal tribes of (sub)tropical India that are examined here do not show an excess of ancestral alleles. The analysis of more-isolated "Negrito" populations of the tropics, including the Kadar (India), Jarawa (Andaman Islands), and Onge (Andaman Islands), may reveal signals of early hominid expansion in tropical regions between Africa and Australia.

Patterns of higher *Alu* frequencies in non-African populations could be produced solely by *in silico* ascertainment of *Alu* markers from mostly non-African DNA sequences (Europeans or Asians). Several results argue against this possibility. First, *Alu* markers either ascertained in African American–derived genomic HeLa libraries or ascertained by other screening methods produce frequency patterns similar to those ascertained *in silico*. Second, gene-diversity estimates for the *Alu* loci are highest in Africa, as is the case with unbiased mtDNA and STRP markers (Jorde et al. 2000). Third, the highest reported *Alu* frequencies occur in popula-

tions from the Americas rather than in populations from Europe or Asia. Fourth, the large continental sample sizes used here diminish the effects of potential ascertainment bias. If ascertainment bias for *Alu* loci exists, the bias is a function of $1/n$, where n is the haploid sample size, under the assumption of rapid population growth (A.R.R., unpublished data). Thus, it is unlikely that the results seen here are simply due to marker-ascertainment methods.

A more likely explanation for the observed pattern stems from the fact that allele frequencies are driven toward fixation (frequencies of either 0 or 1) in populations that have undergone bottlenecks. If a bottleneck accompanied the exodus of human populations out of Africa, then non-African populations should exhibit a more strongly U-shaped allele-frequency distribution than should African populations (Sherry et al. 1997). Because *Alu* polymorphisms are ascertained on the basis of a limited number of individuals, low-frequency polymorphisms are more likely to be missed and high-frequency inserts are likely to be detected. Thus, the observed pattern of insertion frequencies is influenced strongly by the right half of the allele-frequency distribution, in which non-African populations should exhibit an excess of insertion frequencies that are closer to 1. Indeed, this pattern is evident in figure 1*a*. The higher *Alu*-insertion frequencies in European and Asian populations are consistent with a pronounced bottleneck having occurred in these populations.

The similarity between the *Alu*, RSP, STRP, and mtDNA genetic-distance patterns is notable, especially in light of the differences in mutation rates and in mechanisms producing each type of variation. The *Alu*-insertion polymorphism and RSP genetic distances analyzed in the present study are based on the frequencies of loci with known polarity, and both show a gradient of derived alleles that increases from Africa into Europe, eastern Asia, and southeastern Asia. These results suggest a distribution pattern for ancestral alleles and for genetic diversity, in world populations. Further characterization of many loci should provide greater insight into genetic variation and allele-distribution patterns in the human species.

Acknowledgments

The authors thank J. Hawks and B. V. R. Prasad for their helpful comments and suggestions for the manuscript. We are appreciative for the DNA samples provided through collaborations with P. Fischer, T. Jenkins, K. Kidd, J. Kidd, J. Kere, J. M. Naidu, and B. B. Rao. This research was supported by National Institutes of Health grants GM-59290 and RR-00064 and National Science Foundation grants SBR-9514733, SBR-9512178, and SBR 9818215.

Appendix

Table A1

***Alu*-Insertion Frequencies in World Populations**

Population	APO	B65	Col3A1	HS2.43	HS4.14	HS4.32	HS4.65
Africa:							
Alur	.7083	.7500	.4167	.0000	.6667	.5000	.0833
Biaka Pygmy	.7500	.5000	.0000	.0000	.4000	.2000	.0000
Hema	.7778	.5000	.3889	.0000	.5278	.6471	.1667
Mbuti Pygmy	.8000	.8000	.0000	.0000	.2000	.2000	.0000
Nande	.7500	.6765	.3235	.0000	.7059	.3667	.1765
Nguni	.5714	.5417	.1538	.0385	.5000	.2273	.0769
San	.8214	.6538	.1667	.0000	.3929	.3214	.1154
Sotho, Tswana	.7750	.2647	.0000	.0000	.5750	.1333	.0000
Tsonga	.6818	.5000	.1250	.0000	.5714	.2308	.0714
Zaire Pygmy	.9394	.6094	.2813	.0000	.3030	.3871	.1515
Overall	.7823	.5679	.2226	.0034	.4901	.3504	.1047
Asia:							
Cambodian	.7917	.4167	.0833	.0000	.9167	.5455	.0417
Chinese	.8824	.4706	.0294	.0000	.9688	.4375	.2188
Japanese	.8438	.4118	.1563	.0000	.8235	.4375	.0556
Malay	.8333	.5000	.0833	.1667	.8333	.5000	.0000
Vietnamese	.9444	.4444	.0556	.0000	.9444	.5556	.0000
Overall	.8562	.4333	.0890	.0130	.8986	.4722	.0867

Europe:							
Finnish	.9737	.4211	.0250	.0250	.7895	.5789	.0250
French	.9500	.6250	.0500	.0750	.8500	.7000	.0526
Northern European	.9697	.6186	.0224	.0530	.6716	.5859	.0000
Polish	.9500	.3500	.0000	.0000	.8333	.7000	.0000
Overall	.9652	.5602	.0256	.0474	.7348	.6150	.0129
India:							
Brahmin	.8167	.4831	.0583	.0167	.5508	.5690	.0167
Irula[a]	.8088	.4394	.1618	.0294	.6618	.7647	.3529
Kapu	.8103	.4052	.0517	.0345	.6552	.5439	.0345
Kondadora[a]	.8704	.5370	.0192	.0000	.6852	.4630	.4074
Kshatriya	.7778	.3182	.2727	.0000	.6818	.7273	.2727
Madiga	.8571	.3103	.0000	.0000	.5690	.4655	.2931
Mala	.8077	.4231	.0385	.0000	.5385	.5000	.2692
Maria Gond[a]	.6591	.6364	.0455	.0000	.6667	.5238	.2727
Relli	.7895	.5789	.0000	.0000	.6111	.4211	.2895
Santal[a]	.7308	.3750	.0000	.0000	.4615	.5417	.2308
Vysya	.8000	.6000	.1500	.1000	.5000	.5000	.0000
Yadava	.8774	.4717	.0755	.0377	.6961	.5000	.1792
Overall	.8148	.4596	.0637	.0193	.6190	.5435	.1892
All populations	.8372	.4949	.0933	.0199	.6399	.5088	.1312
	HS4.75	Sb19.12	Sb19.3	PV92	TPA25	Ya5 NBC 27	Ya5 NBC 35
Africa:							
Alur	.5455	.5000	.7000	.1250	.2500	.5000	.6667
Biaka Pygmy	.2500	.7000	.2000	.3000	.0000	.2000	.7000
Hema	.7500	.4444	.3824	.3333	.3056	.1111	.6071
Mbuti Pygmy	.8750	.2500	.5000	.6000	.0000	.1000	.7000
Nande	.7941	.4118	.3750	.2667	.2353	.2941	.6818
Nguni	.8182	.4615	.4615	.2500	.0833	.0909	.5455
San	.6071	.3929	.2333	.3000	.2000	.2308	.9091
Sotho, Tswana	.8750	.5000	.2250	.3235	.2857	.2647	.7647
Tsonga	.6786	.5000	.3929	.3077	.2083	.1429	.7308
Zaire Pygmy	.9394	.6034	.3871	.3750	.1212	.0172	.7273
Overall	.7711	.4896	.3733	.3178	.1931	.1715	.7076
Asia:							
Cambodian	1.0000	.0417	.7500	1.0000	.5417	.2083	.5909
Chinese	1.0000	.0000	.8824	.8529	.4412	.4118	.5882
Japanese	1.0000	.0000	.8333	.8571	.5000	.3529	.5385
Malay	1.0000	.0833	.6667	.5000	.2500	.2500	.5833
Vietnamese	1.0000	.0000	.6875	.8750	.2778	.5000	.5625
Overall	1.0000	.0137	.7606	.8571	.3973	.3446	.5746
Europe:							
Finnish	1.0000	.0000	.9500	.1563	.4444	.3500	.7000
French	1.0000	.1500	.8000	.2750	.7250	.0000	.5500
Northern European	1.0000	.1385	.9470	.2537	.5522	.0000	.6032
Polish	1.0000	.1000	.9000	.1250	.7500	.0000	.9000
Overall	1.0000	.1182	.9181	.2342	.5826	.0625	.6372
India:							
Brahmin	.9833	.0169	.8475	.4083	.5185	.2364	.6864
Irula[a]	.9853	.2833	.7941	.4853	.7879	.1818	.5294
Kapu	.9828	.1091	.7946	.4828	.4818	.3103	.6552
Kondadora[a]	1.0000	.0000	.6923	.4038	.7115	.0000	.5185
Kshatriya	1.0000	.0909	.8636	.1364	.4545	.1500	.5455
Madiga	1.0000	.1207	.7759	.4483	.6552	.3214	.6207
Mala	1.0000	.1154	.7308	.5000	.6154	.3077	.5962
Maria Gond[a]	1.0000	.0000	.7273	.3864	.5714	.0000	.6190
Relli	1.0000	.1842	.7895	.6316	.5263	.3158	.5789
Santal[a]	1.0000	.0000	.8750	.6250	.6538	.0000	.5000
Vysya	1.0000	.0556	.9000	.4444	.6500	.0000	.5000

Yadava	1.0000	.1887	.8529	.5472	.6132	.2981	.6538
Overall	.9931	.1060	.8014	.4662	.5957	.2251	.6117
All populations	.9493	.1797	.7260	.4398	.4868	.2000	.6296
	Ya5 NBC 45	Ya5 NBC 51	Ya5 NBC 54	Ya5 NBC 61	Ya5 NBC 102	Ya5 NBC 123	Ya5 NBC 132
Africa:							
Alur	.7778	.5000	.0000	.6000	.5556	.7000	.7778
Biaka Pygmy	.8000	.6000	.0000	.9000	.4000	.3000	1.0000
Hema	.8214	.5278	.0000	.6471	.2778	.5588	.8056
Mbuti Pygmy	.2000	.8000	.0000	.5000	.5000	.6000	.3000
Nande	.7500	.5294	.0000	.4688	.4667	.3824	.7647
Nguni	.4091	.5909	.0000	.6923	.2778	.6667	.7500
San	.4583	.2308	.0000	.7143	.4231	.7667	.9231
Sotho, Tswana	.5588	.6176	.0000	.4375	.2273	.5556	.6667
Tsonga	.3571	.4643	.0000	.7143	.3214	.3929	.6786
Zaire Pygmy	.4375	.7188	.0000	.5000	.2391	.7333	.3594
Overall	.5488	.5634	.0000	.5912	.3443	.5909	.6655
Asia:							
Cambodian	1.0000	.8750	.1250	.6818	.8333	.6667	1.0000
Chinese	.9706	.8529	.0000	.6000	.3824	.6765	1.0000
Japanese	1.0000	.9063	.0833	.6875	.2667	.7059	1.0000
Malay	1.0000	.6667	.1667	.6667	.9167	.3333	1.0000
Vietnamese	1.0000	.9444	.0000	.4444	.5556	.7222	1.0000
Overall	.9929	.8784	.0682	.6357	.5362	.6419	1.0000
Europe:							
Finnish	1.0000	.6250	.0000	.7632	.5250	.7750	1.0000
French	1.0000	.4750	.0500	.6579	.3250	.8750	1.0000
Northern European	.9779	.4779	.0074	.6418	.3788	.8358	1.0000
Polish	.9500	.7222	.0000	.6000	.5500	.8125	1.0000
Overall	.9831	.5214	.0127	.6609	.4095	.8304	1.0000
India:							
Brahmin	1.0000	.7203	.0175	.3704	.4500	.6293	1.0000
Irula[a]	.9848	.6912	.0294	.4091	.6176	.5606	1.0000
Kapu	.9914	.7500	.0263	.4211	.6121	.5088	1.0000
Kondadora[a]	.9423	.7222	.0556	.4074	.6111	.4615	1.0000
Kshatriya	.9091	.6818	.0000	.5455	.5455	.7727	1.0000
Madiga	1.0000	.7414	.0172	.4655	.6897	.6724	1.0000
Mala	.9615	.7500	.0200	.3654	.6923	.5769	1.0000
Maria Gond[a]	.9773	.7273	.0682	.4318	.6136	.2727	1.0000
Relli	1.0000	.7105	.0000	.5789	.5526	.6053	1.0000
Santal[a]	.9231	.8077	.0333	.6538	.4231	.4091	1.0000
Vysya	1.0000	.5000	.0000	.5000	.4444	.7778	1.0000
Yadava	1.0000	.6887	.0094	.4811	.5288	.5000	1.0000
Overall	.9831	.7175	.0237	.4433	.5694	.5501	1.0000
All populations	.9039	.6700	.0216	.5304	.4970	.6160	.9298
	Ya5 NBC 135	Ya5 NBC 147	Ya5 NBC 148	Ya5 NBC 150	Ya5 NBC 157	Ya5 NBC 159	Ya5 NBC 171
Africa:							
Alur	.0000	.0833	.2500	.4545	.9500	.5909	.5000
Biaka Pygmy	.0000	.3000	.4000	.3000	1.0000	.9000	.0000
Hema	.0000	.2500	.1563	.5556	1.0000	.8333	.3333
Mbuti Pygmy	.0000	.0000	.4000	.1000	1.0000	.7500	.2000
Nande	.0000	.0882	.2059	.6176	1.0000	.5313	.0938
Nguni	.0000	.1154	.1786	.6923	.9167	.7143	.2308
San	.0000	.0000	.3214	.5000	.7692	.6538	.1538
Sotho, Tswana	.0000	.1429	.2619	.6471	.9706	.6667	.1875
Tsonga	.0000	.0385	.1667	.5714	.8929	.8077	.2143
Zaire Pygmy	.0000	.0606	.4697	.3906	1.0000	.7097	.2031
Overall	.0000	.1026	.2886	.5137	.9542	.7028	.2143

Asia:							
Cambodian	.0417	.0000	.4091	1.0000	1.0000	1.0000	.4167
Chinese	.0000	.0882	.4118	1.0000	1.0000	.9706	.2059
Japanese	.0278	.1389	.3333	1.0000	1.0000	1.0000	.1944
Malay	.0000	.3333	.5000	1.0000	1.0000	1.0000	.3333
Vietnamese	.0000	.1667	.6111	1.0000	1.0000	1.0000	.0556
Overall	.0132	.1118	.4200	1.0000	1.0000	.9932	.2566
Europe:							
Finnish	.0000	.1250	.3250	1.0000	1.0000	1.0000	.0500
French	.0250	.0250	.1500	1.0000	1.0000	1.0000	.1579
Northern European	.0441	.0956	.1765	.9242	1.0000	1.0000	.1471
Polish	.0000	.2000	.2000	1.0000	1.0000	1.0000	.0000
Overall	.0315	.0975	.1992	.9537	1.0000	1.0000	.1197
India:							
Brahmin	.0083	.1017	.2250	.9576	1.0000	1.0000	.2373
Irula[a]	.0294	.0441	.1765	.9265	1.0000	1.0000	.1563
Kapu	.0000	.1140	.2845	1.0000	1.0000	1.0000	.2636
Kondadora[a]	.0000	.0741	.2115	.9423	1.0000	1.0000	.1852
Kshatriya	.0000	.0455	.3182	.9545	1.0000	1.0000	.5000
Madiga	.0000	.0690	.1250	.9821	1.0000	.9643	.2321
Mala	.0000	.0962	.2115	.9423	1.0000	1.0000	.2692
Maria Gond[a]	.0000	.0909	.1591	.9773	1.0000	.7381	.3095
Relli	.0000	.1316	.2368	.9474	1.0000	1.0000	.2105
Santal[a]	.0000	.0313	.0313	1.0000	1.0000	.9444	.3846
Vysya	.0000	.0556	.1667	1.0000	1.0000	1.0000	.1111
Yadava	.0000	.1321	.3774	.9808	1.0000	1.0000	.1765
Overall	.0042	.0925	.2320	.9681	1.0000	.9802	.2365
All populations	.0086	.0976	.2585	.8724	.9904	.9271	.2142

	Ya5 NBC 208	Ya5 NBC 212	Ya5 NBC 221	Ya5 NBC 237	Ya5 NBC 239	Ya5 NBC 241	Ya5 NBC 242
Africa:							
Alur	.4500	.7500	.7917	.5500	.0000	.0000	.4444
Biaka Pygmy	.7000	1.0000	.6000	.9000	.0000	.2000	.5000
Hema	.5667	.8056	.8611	.8333	.0833	.0833	.4167
Mbuti Pygmy	.3000	.3000	.8750	1.0000	.1000	.0000	.7000
Nande	.7059	.8000	.8235	.5667	.0625	.0000	.4118
Nguni	.5000	.8214	.7500	.7143	.0714	.0000	.5556
San	.1154	.9286	.5000	.8750	.0769	.0000	.3077
Sotho, Tswana	.4722	.7955	.6316	.6471	.0833	.0000	.4167
Tsonga	.5714	.7308	.6923	.8214	.0714	.0385	.4643
Zaire Pygmy	.2031	.4063	.6563	.8281	.0667	.0313	.5323
Overall	.4286	.7100	.7103	.7606	.0667	.0284	.4604
Asia:							
Cambodian	.7917	1.0000	.8182	1.0000	.2000	.6250	.2727
Chinese	.7500	1.0000	.8529	.9706	.0882	.4545	.2692
Japanese	.6765	1.0000	.8947	1.0000	.2000	.4444	.3235
Malay	.7500	1.0000	.8333	1.0000	.2500	.5000	.4000
Vietnamese	.5556	1.0000	.8889	1.0000	.1667	.4444	.4375
Overall	.7162	1.0000	.8618	.9865	.1620	.5072	.3358
Europe:							
Finnish	.8421	1.0000	.8750	1.0000	.2000	.8000	.5250
French	.8250	1.0000	.9500	1.0000	.1750	.7750	.6500
Northern European	.8657	1.0000	.9470	1.0000	.1397	.7222	.6417
Polish	.8000	1.0000	.9000	1.0000	.0500	.6000	.6000
Overall	.8491	1.0000	.9310	1.0000	.1483	.7345	.6182
India							
Brahmin	.8729	1.0000	.8333	1.0000	.1017	.4917	.5000
Irula[a]	.8281	.6912	1.0000	1.0000	.1061	.6029	.2647
Kapu	.8448	1.0000	.8534	1.0000	.0603	.5000	.3772

Kondadora[a]	.9259	1.0000	.9815	1.0000	.1296	.4074	.3519
Kshatriya	.9545	1.0000	1.0000	1.0000	.0909	.5000	.4091
Madiga	.8448	.8966	1.0000	.9655	.0517	.3966	.5172
Mala	.8654	1.0000	1.0000	1.0000	.0769	.5000	.4038
Maria Gond[a]	.8636	1.0000	1.0000	1.0000	.0909	.5952	.2500
Relli	.8684	.7368	1.0000	1.0000	.0789	.5000	.2895
Santal[a]	.9615	1.0000	.8182	1.0000	.2308	.5000	.3462
Vysya	1.0000	1.0000	1.0000	1.0000	.2222	.4444	.5556
Yadava	.9135	1.0000	.9906	1.0000	.0673	.6058	.4804
Overall	.8792	.9489	.9403	.9971	.0922	.5127	.4048
All populations	.7646	.9110	.8824	.9470	.1041	.4483	.4443

[a] Tribal sample.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Eccles Institute of Human Genetics, http://www.genetics.utah.edu/~swatkins/pub/Alu_primers.html (listing of sequences of *Alu* oligonucleotide primers) and http://www.genetics.utah.edu/~swatkins/pub/RSP_links.html (listing of RSP locations and PCR conditions)

Genbank, http://www.ncbi.nlm.nih.gov/Genbank/ (for BLAST screening)

Genome Database, The, http://www.gdb.org/

PHYLIP Home Page, http://evolution.genetics.washington.edu/phylip.html

References

Arcot SS, Adamson AW, Risch GW, LaFleur J, Robichaux MB, Lamerdin JE, Carrano AV, Batzer MA (1998) High-resolution cartography of recently integrated human chromosome 19–specific Alu fossils. J Mol Biol 281:843–856

Arcot SS, Shaikh TH, Kim J, Bennett L, Alegria-Hartman M, Nelson DO, Deininger PL, Batzer MA (1995*a*) Sequence diversity and chromosomal distribution of "young" Alu repeats. Gene 163:273–278

Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995*b*) Alu repeats: a source for the genesis of primate microsatellites. Genomics 29:136–144

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BV, Rasanayagam A, Hammer MF (1998) Female gene flow stratifies Hindu castes. Nature 395:651–652

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. Proc Natl Acad Sci USA 94:4516–4519

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M (1996) Genetic variation of recent Alu insertions in human populations. J Mol Evol 42: 22–29

Batzer MA, Rubin CM, Hellmann-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, 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

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

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

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton

Crawford DC, Schwartz CE, Meadows KL, Newman JL, Taft LF, Gunter C, Brown WT, Carpenter NJ, Howard-Peebles PN, Monaghan KG, Nolin SL, Reiss AL, Feldman GL, Rohlfs EM, Warren ST, Sherman SL (2000) Survey of the fragile X syndrome CGG repeat and the short-tandem-repeat and single-nucleotide-polymorphism haplotypes in an African American population. Am J Hum Genet 66:480–493

Deininger PL, Batzer MA (1993) Evolutionary biology. In: Hect M, MacIntyre RJ, Clegg M (eds) Evolution of retroposons. Vol 27. Plenum Press, New York, pp 157–196

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995) Population genetics of dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56:461–474

Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, Collins JE, Bruskiewich R, et al (1999) The DNA sequence of human chromosome 22. Nature 402:489–495

Felsenstein J (1993) PHYLIP (phylogeny inference package), version 3.5c. Department of Genetics, University of Washington, Seattle

Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat–based phylogeny for the human Y chromosome. Am J Hum Genet 67:182–196

Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS (1999) De-

termination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet 22:164–167

Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL (2000) Evidence for variable selective pressures at MC1R. Am J Hum Genet 66:1351–1361

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Hawks J, Oh S, Hunley K, Dobson S, Cabana G, Dayalu P, Wolpoff MH (2000) An Australasian test of the recent African origin theory using the WLH- 50 calvarium. J Hum Evol 39:1–22

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713

Jorde LB, Bamshad M, Rogers AR (1998) Using mitochondrial and nuclear DNA markers to reconstruct human evolution. Bioessays 20:126–136

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, Soodyall H, Jenkins T, Rogers AR (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66:979–988

Kaessmann H, Heissig F, von Haeseler A, Pääbo S (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22:78–81

Kidd JR, Pakstis AJ, Zhao H, Lu R-B, Okonofua FE, Odunsi A, Grigorenko E, Bonne-Tamir B, Friedlaender J, Schulz LO, Parnas J, Kidd KK (2000) Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. Am J Hum Genet 66: 1882–1899

Kidd KK, Morar B, Castiglione CM, Zhao H, Pakstis AJ, Speed WC, Bonne-Tamir B, Lu RB, Goldman D, Lee C, Nam YS, Grandy DK, Jenkins T, Kidd JR (1998) A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus. Hum Genet 103:211–227

Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S (1999) DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. Proc Natl Acad Sci USA 96:5581–5585

Lewis PO, Zaykin D (2000) Genetic Data Analysis: computer program for the analysis of allelic data. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs

Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet 63:1807–1823

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

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

Nei M, Takezaki N (1996) The root of the phylogenetic tree of human populations. Mol Biol Evol 13:170–177

Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ (1998) Polymorphic Alu insertions and the Asian origin of Native American populations. Hum Biol 70: 23–39

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404: 490–493

Pang H, Koda Y, Soejima M, Schlaphoff T, du Toit ED, Kimura H (1999) Allelic diversity of the human plasma alpha(1,3)fucosyltransferase gene (FUT6). Ann Hum Genet 63: 277–284

Penny D, Steel M, Waddell PJ, Hendy MD (1995) Improved analyses of human mtDNA sequences support a recent African origin for Homo sapiens. Mol Biol Evol 12:863–882

Purandare SM, Cawthon R, Nelson LM, Sawada S, Watkins WS, Ward K, Jorde LB, Viskochil DH (1996) Genotyping of PCR-based polymorphisms and linkage-disequilibrium analysis at the *NF1* locus. Am J Hum Genet 59:159–166

Rana BK, Hewett-Emmett D, Jin L, Chang BH, Sambuughin N, Lin M, Watkins S, Bamshad M, Jorde LB, Ramsay M, Jenkins T, Li WH (1999) High polymorphism at the human melanocortin 1 receptor locus. Genetics 151:1547–1557

Relethford JH, Jorde LB (1999) Genetic evidence for larger African population size during recent human evolution. Am J Phys Anthropol 108:251–260

Rieder MJ, Taylor SL, Clark AG, Nickerson DA (1999) Sequence variation in the human angiotensin converting enzyme. Nat Genet 22:59–62

Rogers AR, Jorde LB (1996) Ascertainment bias in estimates of average heterozygosity. Am J Hum Genet 58:1033–1041

Roy AM, Carroll ML, Kass DH, Nguyen SV, Salem AH, Batzer MA, Deininger PL (1999) Recently integrated human Alu repeats: finding needles in the haystack. Genetica 107: 149–161

Roy AM, Carroll ML, Nguyen SV, Salem AH, Oldridge M, Wilkie AO, Batzer MA, Deininger PL (2000) Potential gene conversion and source genes for recently integrated Alu elements. Genome Res 10:1485–1495

Seielstad M, Bekele E, Ibrahim M, Toure A, Traore M (1999) A view of modern human origins from Y chromosome microsatellite variation. Genome Res 9:558–567

Sherry ST, Harpending HC, Batzer MA, Stoneking M (1997) Alu evolution in human populations: using the coalescent to estimate effective population size. Genetics 147:1977–1982

Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743–748

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematics Zool 35:627–632

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS,

Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071

Stoneking M, Soodyall H (1996) Human evolution and the mitochondrial genome. Curr Opin Genet Dev 6:731–736

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK (1998) A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402

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

Watkins WS, Bamshad M, Dixon ME, Bhaskara Rao B, Naidu JM, Reddy PG, Prasad BV, Das PK, Reddy PC, Gai PB, Bhanu A, Kusuma YS, Lum JK, Fischer P, Jorde LB (1999) Multiple origins of the mtDNA 9-bp deletion in populations of South India. Am J Phys Anthropol 109:147–158

Wolpoff MH, Hawks J, Caspari R (2000) Multiregional, not multiple origins. Am J Phys Anthropol 112:129–136

Yu N, Zhao Z, Fu Y-X, Sambuughin N, Ramsay M, Jenkins T, Leskinen L, Patthy L, Jorde LB, Kuromori T, Li W-H (2001) Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol Biol Evol 18:214–222

Zhao Z, Jin L, Fu YX, Ramsay M, Jenkins T, Leskinen E, Pamilo P, Trexler M, Patthy L, Jorde LB, Ramos-Onsins S, Yu N, Li W-H (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci USA 97:11354–11358

Zischler H, Geisert H, von Haeseler A, Pääbo S (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature 378:489–492