Am. J. Hum. Genet. 67:901–925, 2000

Short Tandem-Repeat Polymorphism/*Alu* Haplotype Variation at the PLAT Locus: Implications for Modern Human Origins

S. A. Tishkoff,[1,*] A. J. Pakstis,[2] M. Stoneking,[3] J. R. Kidd,[2] G. Destro-Bisol,[4] A. Sanjantila,[5] R.-b. Lu,[6] A. S. Deinard,[7] G. Sirugo,[8] T. Jenkins,[9] K. K. Kidd,[2] and A. G. Clark[1]

[1]Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, University Park; [2]Department of Genetics, School of Medicine, Yale University, New Haven; [3]Max Planck Institute for Evolutionary Anthropology, Leipzig; [4]Department of Animal and Human Biology, University "La Sapienza" Rome; [5]Department of Forensic Medicine, University of Helsinki, Helsinki; [6]Department of Psychiatry, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan; [7]Department of Anthropology, University of California at Davis, Davis; [8]Human Molecular Genetics Laboratory, Medical Research Council, Fajara, The Gambia; and [9]Department of Human Genetics, School of Pathology, The South African Institute for Medical Research, University of the Witwatersrand, Johannesburg

Two dinucleotide short tandem-repeat polymorphisms (STRPs) and a polymorphic *Alu* element spanning a 22-kb region of the PLAT locus on chromosome 8p12-q11.2 were typed in 1,287–1,420 individuals originating from 30 geographically diverse human populations, as well as in 29 great apes. These data were analyzed as haplotypes consisting of each of the dinucleotide repeats and the flanking *Alu* insertion/deletion polymorphism. The global pattern of STRP/*Alu* haplotype variation and linkage disequilibrium (LD) is informative for the reconstruction of human evolutionary history. Sub-Saharan African populations have high levels of haplotype diversity within and between populations, relative to non-Africans, and have highly divergent patterns of LD. Non-African populations have both a subset of the haplotype diversity present in Africa and a distinct pattern of LD. The pattern of haplotype variation and LD observed at the PLAT locus suggests a recent common ancestry of non-African populations, from a small population originating in eastern Africa. These data indicate that, throughout much of modern human history, sub-Saharan Africa has maintained both a large effective population size and a high level of population substructure. Additionally, Papua New Guinean and Micronesian populations have rare haplotypes observed otherwise only in African populations, suggesting ancient gene flow from Africa into Papua New Guinea, as well as gene flow between Melanesian and Micronesian populations.

Introduction

Haplotypic variation consisting of both fast-evolving short tandem-repeat polymorphisms (STRPs) and more slowly evolving markers such as restriction-fragment-length polymorphisms (RFLPs), single-nucleotide polymorphisms (SNPs), and insertion/deletion polymorphisms (indels) have proved to be useful for both the tracing of population migrations and determining when mutation events occurred (Tishkoff et al. 1996*a,* 1998*a,* 1998*b;* Kidd et al. 1998). STRPs have moderate to high mutation rates (usually 10^{-5}–10^{-2}/generation [Weber and Wong 1993; Tautz and Schlötterer 1994; Chakraborty et al. 1997; Brinkmann et al. 1998]) and are thought to mutate via the "stepwise" gain or loss of single-repeat units, although larger "jumps" in repeat size occasionally do occur (Shriver et al. 1993; Valdes et al. 1993; Di Rienzo et al. 1994; Tishkoff et al. 1998*a*). The instability of STRPs results in the formation of many alleles, and stable flanking markers allow greater certainty in tracing the lineage of each haplotype and in determining the identity by descent of haplotype lineages. Scoring these markers as haplotypes allows analysis both in terms of haplotype frequencies and identity and in terms of linkage disequilibria. Thus, in addition to differing in the frequency of alleles at the individual polymorphic sites, populations may differ in the particular combination of alleles on a chromosome, and a shared pattern of linkage disequilibrium (LD) may be informative for determination of recent common ancestry and for reconstruction of historic migration events (Tishkoff et al. 1996*a,* 1998*a,* 1998*b;* Kidd et al. 1998, 2000). In addition, if one knows or can estimate the mutation rate of an STRP, as well as the recombination rate between the STRP and a stable allele marker, it becomes possible to estimate the age of the stable SNP or indel marker. Such an analysis has been applied to the CD4 gene, as well as to a number of mutations

Received March 13, 2000; accepted for publication July 18, 2000; electronically published September 14, 2000.

Address for correspondence and reprints: Dr. Sarah A. Tishkoff, University of Maryland, Department of Biology, Biology/Psychology Building, #144, University of Maryland, College Park, MD 20742. E-mail: st130@umail.umd.edu

* Present affiliation: Department of Biology, University of Maryland, College Park.

© 2000 by The American Society of Human Genetics. All rights reserved.
0002-9297/2000/6704-0014$02.00

resulting in disease and disease resistance (Serre et al. 1990; Hästbacka et al. 1992; Risch et al. 1995; Bertranpetit and Calafell 1996; Tishkoff et al. 1996*a,* 1998*b;* Rannala and Slatkin 1998; Stephens et al. 1998).

We have examined the global frequency distribution of two dinucleotide STRPs and an *Alu*-insertion polymorphism encompassing a 22-kb region within the tissue-plasminogen–activator locus (PLAT) located on the short arm of chromosome 8 (8p12-q11.2; see fig. 1) (Degen et al. 1986). The polymorphic *Alu* element (often referred to as "TPA *Alu*") is located within intron 8 of the *PLAT* gene (GenBank sequence position 28804 [accession number K03021; see the NCBI GenBank Overview Web site]; see fig. 1) and is a member of the human-specific (HS) (also known as "PV") subfamily of *Alu* elements that recently have retroposed within the human genome (Batzer et al. 1991, 1996). It has been hypothesized that the presence of the *Alu*-insertion allele, herein denoted "*Alu*(+)," at PLAT may predict risk for coronary thrombosis. However, two recent studies (Ridker et al. 1997; Steeds et al. 1998) have observed no significant difference in the frequency of the *Alu*(+) allele in individuals at high risk for myocardial infarction, compared with that in healthy control cases, suggesting that the indel is not a major independent risk factor for coronary thrombosis.

Alu insertions are useful for the study of human evolution because they are unique, stable mutation events and because the ancestral state is known to be the absence of the *Alu* element, herein denoted "*Alu*(−)" (Perna et al. 1992; Batzer et al. 1994, 1996; Tishkoff et al. 1996*b;* Sherry et al. 1997; Stoneking et al. 1997). We have examined haplotypes involving the PLAT *Alu* and the two closely linked STRPs in 1,225–1,375 individuals originating from 30 geographically diverse human populations (fig. 2). The patterns of haplotype variation and LD observed at the PLAT locus support a recent African origin of non-African human populations and suggest that, throughout much of modern human history, sub-Saharan Africa has maintained a large effective population size and a high level of population substructure.

Subjects and Methods

Subjects

Individuals originating from 30 geographically diverse populations were typed for the *Alu,* PLAT $(CA)_n$-1, and PLAT $(CA)_n$-2 polymorphisms. In most cases populations represent well-defined ethnic groups, but in some cases ethnically diverse populations from the same geographic region have been pooled because of small sample sizes. For example, the Bantu-speakers, Somali, and Papua New Guineans each represent here a pooling of ethnically diverse groups. Populations examined include 13 African populations (country of origin is given in parentheses): Wolof (Senegal), Ewondo (Cameroon), Bamileke (Cameroon), Mbuti (Democratic Republic of Congo), Biaka (Central African Republic), Bantu-speakers from various southern-African chieftainships (South Africa), Herero (Namibia), Kwengo (Namibia), Nama (Namibia), Va/Sekele !Kung San (Namibia), Zu/Wasi !Kung San (Namibia), Somali, Ethiopian Jews; two Middle Eastern populations—Yemenite Jews (Yemen) and Druze (Israel); two European populations—Finns and Danes; five Asian populations—Japanese, San Francisco Chinese, Ami (Taiwan), Atayal (Taiwan), and Yakuts (Siberia); three Oceanic populations—Micronesians (assorted islands), Nasioi (Melanesia), and Papua New Guineans; and five Amerindian populations— Cheyenne (Oklahoma), Maya (Yucatan, Mexico), Karitiana (Rondonia, Brazil), Surui (Rondonia, Brazil), and Ticuna (Amazonia, Brazil). Most of the population samples have been described elsewhere, by Nurse et al. (1985), Scozzari et al. (1988, 1999), Stoneking et al. (1990),

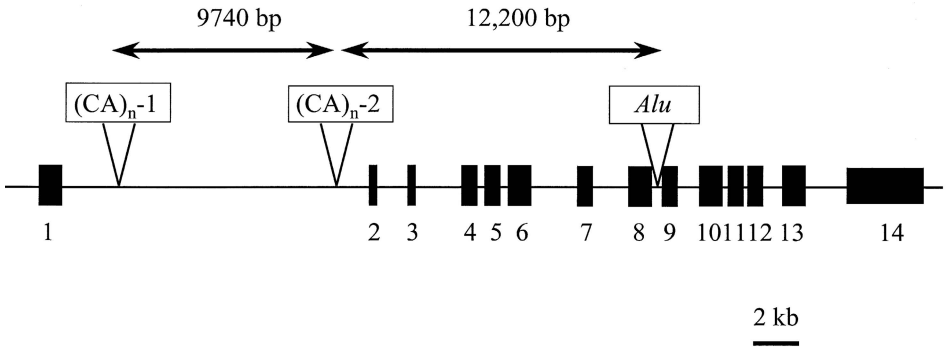

Figure 1 Diagram of PLAT gene structure, showing location of the polymorphic PLAT *Alu* and the $(CA)_n$-1 and $(CA)_n$-2 STRPs used in the haplotype analysis. Exons are shown as blackened boxes.

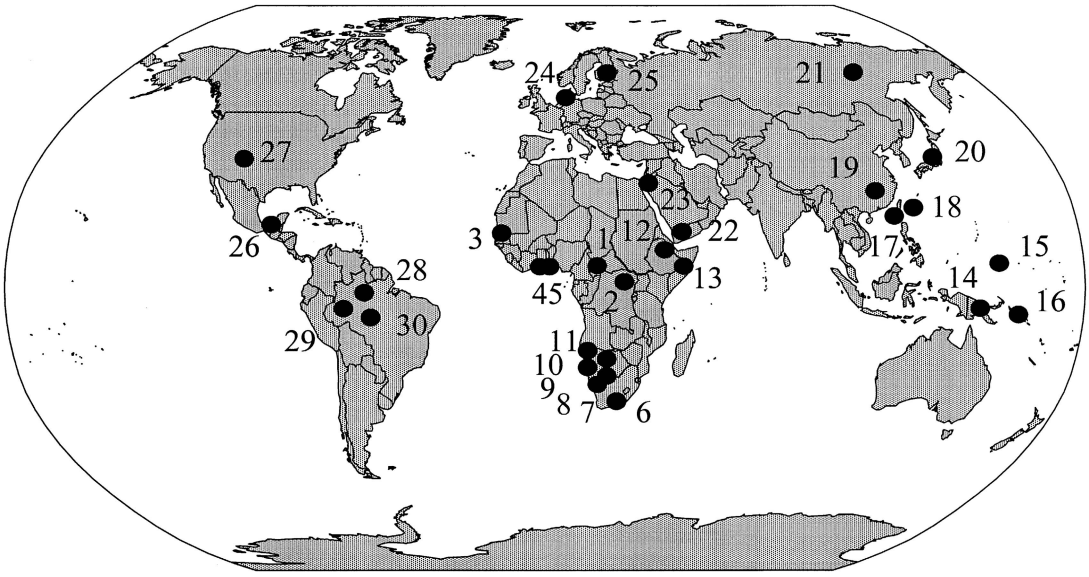

Figure 2 Global distribution of populations included in the PLAT haplotype study. 1 = Biaka; 2 = Mbuti; 3 = Wolof; 4 = Ewondo; 5 = Bamileke; 6 = Bantu-speakers; 7 = Herero; 8 = Zu/Wasi !Kung San; 9 = Kwengo; 10 = Nama; 11 = Va/Sekele !Kung San; 12 = Ethiopians; 13 = Somali; 14 = Papua New Guineans; 15 = Micronesians; 16 = Nasioi Melanesians; 17 = Ami; 18 = Atayal; 19 = Chinese; 20 = Japanese; 21 = Yakut; 22 = Yemenites; 23 = Druze; 24 = Danes; 25 = Finns; 26 = Maya; 27 = Cheyenne; 28 = Ticuna; 29 = Surui; 30 = Karitiana.

Bowcock et al. (1991), Spurdle and Jenkins (1992), Barr and Kidd (1993), Goldman et al. (1993), Lichter et al. (1993), Destro-Bisol et al. (1994), Castiglione et al. (1995), Soodyall et al. (1996), Tishkoff et al. (1996*a*, 1998*a*), Calafell et al. (1998), Kidd et al. (1998), and Spedini et al. (1999). Genomic DNA was extracted from blood or from Epstein-Barr virus–transformed lymphoblastoid cell lines, by standard methods. The great apes sampled for this project include 19 *Pan troglodytes,* 6 *Pan paniscus,* and 4 *Gorilla gorilla* (described in Deinard and Kidd 1999). All blood samples were obtained with informed consent, and typings were done under protocols approved by the human-subjects committees of all universities and research institutions involved in this study.

PCR Methods

The *Alu*-insertion polymorphism was typed with the published primer sequences and methods described by Tishkoff et al. (1996*b*). Amplification produces a 570-bp fragment from chromosomes with the *Alu* insertion and a 260-bp fragment from those without it. PCR products were separated on a 1% agarose gel, were stained with ethidium bromide, and were visualized with UV light.

The two dinucleotide repeats were typed by use of the published primers for $(CA)_n$-1 (Thomas and Drayna 1992) and $(CA)_n$-2 (Sadler et al. 1991). Amplification was performed with 50 ng of genomic DNA in a 25-μl (total volume) reaction mixture. The reaction mixture for $(CA)_n$-1 contained 5 pmol each of fluorescent-labeled primer (PLAT1A 5′-GAC AGC ACA TTC TCT TAG CAA-3-′) and unlabeled primer (PLAT1B 5′-GTG ATG GAG TCA GAC CTT GTC-3′), 200 μM of each dNTP, 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, and .625 U of *Taq* polymerase. Samples were denatured for 1 min at 94°C, followed by 25 cycles of 94°C for 1 min, 57°C for 1 min, and 72°C for 1 min, followed by a 10-min extension at 72°C. The amplification conditions for $(CA)_n$-2 were identical, except for the use of 5 pmol each of fluorescent-labeled primer (PLAT2A 5′-GCC TGG ACA ACA TAG AGA AAC C-3′) and unlabeled primer (PLAT2B 5′-ACT TCA GGC ATG TGC CAC TG-3′) and the addition of 1.5 μl of deionized formamide to the reaction mix. Amplification products were run on a 6% polyacrylamide gel, on either an ABI 373 or ABI 377 DNA sequencer, and fragment sizes were determined with GENESCAN software.

Sequencing

STRP alleles were sequenced either from clones or directly from PCR-amplification products. Cloning was used to sequence large-sized alleles at $(CA)_n$-1, which were present only as heterozygotes with small-sized alleles. STRP alleles were amplified by the method described above and were visualized on a 3.5% Nusieve gel. Bands containing the STRP alleles were excised from the gel, and the DNA was isolated by a Sephaglas DNA

isolation kit (Pharmacia). These purified products were cloned into a PT7Blue-3 plasmid vector by use of a Novagen cloning kit. Minipreps of the plasmids containing the cloned alleles were performed by use of Promega's Wizard Plus SV miniprep kit. Plasmids containing the STRP allele inserts were cycle sequenced in both directions, with fluorescently labeled terminators, by use of either an ABI dideoxy terminator kit or a Beckman CEQ DTCS sequencing kit and T7 and U19 primers specific to the PT7Blue-3 vector. PCR products were run on either an ABI 373 or Beckman CEQ2000 automated DNA sequencer. For sequencing from PCR products from individuals homozygous for STRP alleles, alleles from $(CA)_n$-1 were amplified with primers TpaSeq1A (5′-AAA GCT CAT CCA CCC TGC TC-3′) and TpaSeq1 B (5′-CAT GCC CCT GTA GTC CTA GC-3′), which produce a 302-bp product for a 113-bp allele. Alleles from $(CA)_n$-2 were amplified with primers TpaSeq2A (5′-AAG GAA GGA AAA ATG CTG GG-3′) and TpaSeq2B (5′-GAC TGG AGT GCA GTG GCA TG-3′), which produce a 302-bp product for a 111-bp allele. Amplification was performed by use of 50–100 ng of genomic DNA in a 25-μl (total volume) reaction mixture. The reaction mixture contained 10 pmol of each forward (A) and reverse (B) primer, 200 μM of each dNTP, 50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, and 0.625 U of *Taq* polymerase. For TpaSeq1, samples were denatured for 1 min at 94°C, followed by 25 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min, followed by a 10-min extension at 72°C. For TpaSeq2, samples were denatured for 1 min at 94°C, followed by 25 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min, followed by a 10-min extension at 72°C. The amplified products were purified by use of a Qiagen PCR purification kit and were cycle sequenced by use of a Beckman CEQ DTCS sequencing kit. Products were run and analyzed on a Beckman CEQ2000 automated DNA sequencer.

Allele- and Haplotype-Frequency Estimates

The allele frequencies at the separate sites—$(CA)_n$-1, $(CA)_n$-2, and *Alu*—were estimated by gene counting. Heterozygosities for individual sites and for the haplotypes have been estimated as $n(1 - \Sigma p_i^2)/(n - 1)$, where p_i represents the frequency of the ith allele or haplotype for any given system and where n is the number of chromosomes in the sample. Probability values of Hardy-Weinberg (HW) exact tests and tests for heterozygosity excess or deficiency (Guo and Thompson 1992; Rousset and Raymond 1995; Rousset 1996) as well as values of F_{ST} (Weir and Cockerham 1984) were calculated with GENEPOP software (release 3.1b). Summary statistics for the STRP allele distributions were calculated by software available at the Microsat Web site. The computer program HAPLO (Hawley and Kidd 1995) was used to generate maximum-likelihood estimates of the haplotype frequencies for the $(CA)_n$-1/*Alu* and $(CA)_n$-2/*Alu* haplotypes. No individuals with missing STRP data were included in the data set used to estimate STRP/*Alu* haplotype frequencies. The number of possible three-locus haplotypes, including the two dinucleotide repeats and the *Alu* indel, was >500, and most were rare. In the African populations, nearly every individual had a unique three-locus phenotype. Consequently, haplotype-phase estimation would have very large standard errors, and, therefore, the complete three-locus haplotypes were not estimated; however, frequencies of two-locus haplotypes could be obtained with high statistical accuracy, because there were fewer "ambiguous" multisite heterozygotes. Tishkoff et al. (2000) have demonstrated that frequencies of two-locus STRP/indel haplotypes inferred by the HAPLO program do not differ significantly from frequencies based on gene counting using haplotypes identified unambiguously by molecular haplotyping methods. The inaccuracy in statistically estimated haplotype frequencies is greatest for rare (i.e., population frequency <.05) haplotypes. Because these errors are restricted to rare haplotypes, and because statistics such as haplotype diversity depend on the square of haplotype frequencies, the errors have negligible impact on subsequent calculations of population statistics.

Tests of Population Subdivision and Principal-Components Analysis (PCA)

Genetic differentiation among populations grouped by geographic region were estimated by F_{ST} (Wright 1931), D_{SW} (Shriver et al. 1995), and D_{LR}, the last of which is a likelihood-ratio statistic developed by Paetkau et al. (1997). F_{ST} (Wright 1931) partitions variance into within- and among-population components. D_{SW} (Shriver et al. 1995) is based on the stepwise-mutation model and makes use of the fact that STRP alleles of similar size most likely have a more recent common ancestor than do STRP alleles of grossly divergent size. The likelihood-ratio test D_{LR} (Paetkau et al. 1997) is used for testing the observed data against the null hypothesis that the sample is drawn from a single panmictic population. This measure is particularly sensitive to rare alleles or haplotypes that occur when highly variable STRP systems are examined.

Let f_0 be the probability (averaged over populations) that two identical haplotypes within a population will be drawn, and let $\bar{f}$ be the probability that two identical haplotypes will be drawn across the entire sample; then $F_{ST} = (f_0 - \bar{f})/(1 - \bar{f})$. Shriver et al. (1995) defined the statistic D_{SW} as a measure of population divergence for STRP loci. Let X_{ik} be the frequency of a haplotype with i copies of the repeat at locus k in population X, and let Y_{ik} be the corresponding frequency in population

Y. Letting j be a dummy index, assume that $W_X = \Sigma_k\Sigma_i\Sigma_j \mid i-j \mid x_{ik}x_{jk}$, $W_Y = \Sigma_k\Sigma_i\Sigma_j \mid i-j \mid y_{ik}y_{jk}$, and $W_{XY} = \Sigma_k\Sigma_i\Sigma_j \mid i-j \mid x_{ik}y_{jk}$; then, according to the method of Shriver et al. (1995), $D_{SW} = W_{XY} - [(W_X + W_Y)/2]$. In applying this statistic to haplotypes, we need to consider the distance between haplotypes that differ with respect to the *Alu* indel. Two haplotypes that have the same number of STRP repeats but differ at the *Alu* indel are treated as having a difference of x repeats for purposes of the D_{SW} calculation, where x is allowed to be 1,2,3,...,5. This weighting is arbitrary and, in practice, was found to change the absolute estimates of D_{SW}, but it did not change either the level of significance or the pattern of D_{SW} among regions.

Define L_{ij} as the likelihood that haplotype i will be drawn in population j, and let L_{iT} be the likelihood that haplotype i will be drawn in any population other than j. A slight modification of the method of Paetkau et al. (1997) gives

$$D_{LR} = \sum_{i=1}^{i_{max}}\left[\frac{n_j}{n_{total}}\sum_{j=1}^{j_{max}}\log\left(\frac{L_{ij}}{L_{iT}}\right)\right],$$

where n is the number of chromosomes in the sample. Tail probabilities for D_{SW} and D_{LR} were generated as described by Hudson et al. (1992). Null distributions were obtained by randomly permuting the genotypes across the populations and calculating the test statistics over 10,000 such randomized data sets.

PCA

PCA was performed by defining each population as a vector of frequencies of two-locus haplotypes (composed of either the $(CA)_n$-1 or $(CA)_n$-2 repeat and the *Alu* indel). Frequencies were arcsine transformed and were treated as measures from a multivariate Gaussian distribution. PCA was performed on the correlation matrix of these scores, by Minitab Statistical Software (version 12).

LD Analysis

For each STRP allele, a 2 × 2 table was constructed for counts of that allele versus counts of all other alleles pooled, and for *Alu*(+) alleles versus *Alu*(−) alleles. The standardized, pairwise LD value D' (Lewontin 1964) was calculated for each such 2 × 2 table, and the null hypothesis of LD ($D' = 0$) was tested by Fisher's exact test (Sokal and Rohlf 1995, pp. 730–736). The significance of LD estimated by a likelihood-ratio test gave results that were virtually identical to those of Fisher's exact tests, so only the latter results are reported.

Results

Alu *Polymorphism*

The full sample typed for the *Alu* indel represents 1,375 individuals originating from 30 geographically diverse populations (fig. 2 and table 1). Allele frequencies for the *Alu* polymorphism are shown in table 1. The *Alu* insertion is polymorphic, with moderate (⩾.30) heterozygosity levels in most populations, with the exceptions of the Nasioi and Papua New Guinean populations, which have low frequencies (7% and 16%, respectively) of the *Alu*(+) allele, and the South American Ticuna population, which has a very high frequency (91%) of the *Alu*(+) allele. All three of these populations are small and isolated and have experienced high levels of genetic drift, which likely played a role in the establishment of one allele at very high frequency.

In general, African populations have low frequencies of the *Alu*(+) allele, in the range of .18–.38, with the exception of the Wolof (.44) and the Somali (.47). Non-African populations have high frequencies of the *Alu*(+) allele, in the range of .27–.91, with the exception of the Nasioi (.07) and Papua New Guineans (.16). These results are consistent with those of previous studies of the PLAT *Alu* polymorphism in a smaller sample of populations (Perna et al. 1992; Batzer et al. 1994; Tishkoff et al. 1996*b*; Stoneking et al. 1997; Novick et al. 1998). None of the populations exhibit a significant departure from HW equilibrium, with the exception of the Bamileke from Cameroon ($P < .04$), the Atayal from Taiwan ($P < .02$), and the Yakuts from Siberia ($P < .01$). However, these probability values are for individual tests, and none of the departures from HW expectation were significant at the experiment-wide level after the Bonferroni correction for multiple tests was applied. Only the *Alu*(−) allele was detected in 29 nonhuman primates examined (19 common chimpanzees, 6 pygmy chimpanzees, and 4 gorillas). Both the absence of the HS *Alu* in the nonhuman primates and the high heterozygosity of the *Alu* indel in most human populations support the hypothesis that the *Alu*-insertion event occurred after the divergence of humans from the great apes, ~5 million years ago (Tishkoff et al. 1996*b*).

$(CA)_n$ *Polymorphisms*

The $(CA)_n$-1 repeat is located ~21,940 bp from the *Alu* polymorphism, at sequence position 7173 of the GenBank sequence (accession number K03021; fig. 1). Published primers flanking the repeat produce a PCR product that is 99–173 bp in size (Thomas and Drayna 1992). The GenBank sequence, which represents a 113-bp allele, is actually a compound repeat consisting of the sequence $(GT)_{14}(AT)_{12}$. The $(CA)_n$-2 repeat is located ~12,200 bp from the *Alu* polymorphism at sequence

Table 1

Allele Frequencies of the *Alu*-Insertion (+)/Deletion (−) Polymorphism

Population (No. of Chromosomes)	Frequency			Heterozygosity	HW Probability[a]
	Alu(+)	*Alu*(−)	Standard Error		
Wolof (122)	.44	.56	.045	.49	.122
Ewondo (68)	.38	.62	.059	.47	.477
Bamileke (62)	.29	.71	.058	.41	.034*
Mbuti (78)	.26	.74	.050	.38	1.000
Biaka (138)	.22	.78	.035	.34	.731
Bantu-speakers (84)	.29	.71	.050	.41	1.000
Herero (84)	.31	.69	.050	.43	.276
Kwengo (74)	.27	.73	.052	.39	.398
Nama (50)	.30	.70	.065	.42	.144
Va/Sekele !Kung San (60)	.18	.82	.050	.30	.551
Zu/Wasi !Kung San (58)	.34	.66	.062	.45	1.000
Somali (58)	.47	.53	.066	.50	.467
Ethiopian Jews (120)	.27	.73	.041	.39	.467
Yemenite Jews (82)	.35	.65	.053	.46	1.000
Druze (142)	.54	.46	.042	.50	1.000
Finns (68)	.65	.35	.058	.46	.056
Danes (102)	.50	.50	.050	.50	.053
Japanese (100)	.56	.44	.050	.49	1.000
Chinese (102)	.46	.54	.049	.50	.574
Atayal (84)	.27	.73	.048	.39	.019*
Ami (80)	.52	.48	.056	.50	.751
Micronesians (58)	.55	.45	.065	.50	1.000
Nasioi (44)	.07	.93	.038	.13	1.000
Papua New Guineans (186)	.16	.84	.027	.27	1.000
Yakuts (102)	.53	.47	.049	.50	.011*
Cheyenne (110)	.59	.41	.047	.48	.401
Maya (102)	.63	.37	.048	.47	1.000
Karitiana (110)	.65	.35	.045	.46	1.000
Ticuna (134)	.91	.09	.025	.16	.417
Rondonian Surui (88)	.50	.50	.053	.50	.555

[a] Tail probability of χ^2 test of HW equilibrium.

* $P < .05$.

position 16911 of the GenBank sequence (fig. 1). Published primers flanking the repeat produce a PCR product that is 105–167 bp in size (Sadler et al. 1991). The GenBank sequence, which represents a 111-bp allele, is also a compound repeat consisting of the sequence $(CA)_{15}(CT)_{10}$. The $(CA)_n$-1 and $(CA)_n$-2 alleles in our analysis are based on the size of the PCR product when primers flanking the compound repeat are used, and therefore there is some possibility of additional heterogeneity at the sequence level.

In order to examine the possibility of sequence heterogeneity, $(CA)_n$-1 and $(CA)_n$-2 alleles were sequenced in a set of Bantu-speakers and Papua New Guinean individuals as well as in several chimpanzees and a gorilla; results are shown in table 2. In general, alleles of similar size in humans have a similar number of STRP repeats. But sequencing of multiple independently cloned alleles from the same individual indicates that slippage of *Taq* polymerase during PCR amplification prior to cloning can result in expansion or contraction of either of the two microsatellites that compose the "compound" STRP at $(CA)_n$-1 and $(CA)_n$-2; however, comparison of multiple alleles from several human individuals indicates that alleles of similar size at $(CA)_n$-1 and $(CA)_n$-2 are similar at the sequence level in Africans and Papua New Guineans. By contrast, $(CA)_n$-1 and $(CA)_n$-2 alleles of similar size in chimpanzees and gorillas are quite distinct at the sequence level, both from humans and from each other. Compared with humans, who have perfect, uninterrupted repeats, both the gorilla and the chimpanzee have highly compound repeats at $(CA)_n$-1. In addition, the gorilla sequence at $(CA)_n$-1 diverges from both the human sequence and the chimp sequence, in the region immediately flanking the STRP (table 2), because of deletions in the sequence farther from the STRP but within the 105-bp distance spanning the primer sites used for GENESCAN analysis (data not shown). At $(CA)_n$-2, chimpanzees and humans differ slightly in the number of $(AC)_nCT_n$ repeats, and chimpanzees have a TTC duplication absent in humans.

Table 2

Sequence Analysis of $(CA)_n$-1 and $(CA)_n$-2 Alleles in Selected Individuals Representing Various Human Populations and in Nonhuman Primates

Source or Individual	Allele Size (bp)	No. of Independent Sequences	Sequence
$(CA)_n$-1:			
GenBank (accession number K03021)	113	1	TATGTGTTTGTA$(TG)_{14}(TA)_{12}$ATTTAGAGACAAGGTCT
Papua New Guinea 1	119	1	$(TG)_{16}(TA)_{13}$ [a]
		2	$(TG)_{17}(TA)_{12}$ [b]
		1	$(TG)_{15}(TA)_{13}$ [c]
Papua New Guinea 2	121	2	$(TG)_{18}(TA)_{12}$ [d]
		1	$(TG)_{19}(TA)_{11}$ [a]
		1	$(TG)_{18}(TA)_{11}$ [e]
		2	$(TG)_{18}(TA)_{10}$ [f]
Papua New Guinea 3	121	2	$(TG)_{18}(TA)_{12}$ [g]
		1	$(TG)_{18}(TA)_{11}$ [e]
Papua New Guinea 4	121	1	$(TG)_{19}(TA)_{11}$ [h]
Papua New Guinea 5	121	1	$(TG)_{18}(TA)_{12}$ [h]
Papua New Guinea 6	121	1	$(TG)_{18}(TA)_{12}$ [h]
Papua New Guinea 7	121	1	$(TG)_{18}(TA)_{12}$ [h]
Papua New Guinea 8	121	1	$(TG)_{18}(TA)_{12}$ [h]
Bantu 1	121	5	$(TG)_{18}(TA)_{12}$ [i]
Bantu 2	123	1	$(TG)_{18}(TA)_{13}$ [h]
		1	$(TG)_{18}(TA)_{12}$ [c]
Bantu 3	125	1	$(TG)_{17}(TA)_{14}$ [a]
Bantu 4	139	1	$(TG)_{22}(TA)_{12}$ [c]
Papua New Guinea 9	161	1	$(TG)_{28}(TA)_{19}$ [e]
Chimp 1	125	1	TATGTGTTTG$(TA)_2(TG)_{17}(TA)_1(TG)_4(GG)_1(TA)_8$TTTAGAG––AAGGTCT [h]
Chimp 2	121	1	$(TA)_2(TG)_{15}(TA)_1(TG)_4(GG)_1(TA)_8$ [h]
Chimp 3	121	1	$(TA)_2(TG)_{15}(TA)_1(TG)_4(GG)_1(TA)_8$ [h]
Gorilla 1	105	1	TATGTGTTTG$(TA)_3(TG)_3(TA)_1(TG)_{10}(TA)_8(GA)_2$CAGAGGTCT [j]
$(CA)_n$-2:			
GenBank (accession number K03021)	111	1	TTGTCTCT$(AC)_{16}(TC)_{10}$AGCCAGGCACAGT
Papua New Guinea 1	113	1	$(AC)_{15}(TC)_{12}$ [h]
Papua New Guinea 2	115	1	$(AC)_{16}(TA)_{12}$ [h]
Papua New Guinea 3	125	1	$(AC)_{17}(TC)_{15}$ [c]
Chimp 1	115	1	$(AC)_{13}(TC)_{11}(TTC)_2$ [h]

[a] One sequence from a plasmid clone.
[b] Two sequences from PCR amplification.
[c] One sequence from PCR amplification, which is smaller than would be expected on the basis of allele size as determined by GENESCAN analysis.
[d] Two sequences from a plasmid clone.
[e] One sequence from a plasmid clone, which is smaller than would be expected on the basis of allele size as determined by GENESCAN analysis.
[f] One sequence from a plasmid clone and one sequence from PCR amplification, which are smaller than would be expected on the basis of allele size as determined by GENESCAN analysis.
[g] One sequence from a plasmid clone and one sequence from PCR amplification
[h] One sequence from PCR amplification.
[i] Four sequences from a plasmid clone and one sequence from PCR amplification.
[j] One sequence from PCR amplification, which is larger than would be expected on the basis of allele size as determined by GENESCAN analysis.

Summary statistics of the size range, frequency, and diversity of alleles of the $(CA)_n$-1 and $(CA)_n$-2 STRPs are shown in table 3. In general, heterozygosity and allele-size ranges are highest in African populations, lower in Middle Eastern, European, and Asian populations, and lowest in Oceanic and New World populations (table 3). For the $(CA)_n$-1 polymorphism, the only populations that exhibit genotype proportions that deviate significantly from HW expectations are the Central African Republic Biaka ($P < .0001$), Namibian Kwengo ($P < .0001$), Namibian Va/Sekele !Kung San ($P < .0001$), Namibian Zu/Wasi !Kung San ($P < .002$), North American Cheyenne ($P < .004$), and Rondonian Surui ($P < .02$) (table 3). All six of these populations have an excess of homozygosity. For the $(CA)_n$-2 polymorphism, the only populations that exhibit genotype proportions that de-

Table 3

Allele-Size Statistics for PLAT $(CA)_n$-1 and $(CA)_n$-2 STRP Markers, Based on PCR-Fragment Size

Population (No. of Chromosomes)	Minimum	Maximum	Mean	Median	Mode	Variance	Heterozygosity
$(CA)_n$-1:							
African:							
Wolof (122)	115	169	131.20	127	123	168.62	.892
Ewondo (70)	115	159	134.43	129	129	187.61	.912
Bamileke (64)	117	169	134.22	131	121	203.89	.920
Mbuti (68)	99	159	130.47	125	123	157.09	.863
Biaka (94)	99	171	131.17	129	129	192.92	.902
Bantu-speakers	113	159	133.03	113	159	173.74	.917
Herero (86)	115	167	137.72	129	121	269.36	.905
Kwengo (74)	115	157	126.51	125	129	59.49	.870
Nama (48)	115	157	127.50	125	121	107.49	.853
Va/Sekele !Kung San (60)	115	147	124.90	123	121	49.55	.854
Zu/Wasi !Kung San (58)	117	145	126.41	125	125	35.23	.859
Somali (60)	119	157	125.40	123	119	68.18	.738
Ethiopian Jews (58)	111	157	127.52	121	121	136.64	.707
Middle East/Europe:							
Yemenite Jews (56)	119	133	122.04	121	121	7.56	.392
Druze (110)	117	131	122.05	121	121	7.46	.485
Finns (68)	115	133	121.59	121	121	5.14	.644
Danes (100)	109	139	121.82	121	121	16.19	.530
Asia:							
Yakut (66)	109	125	120.91	121	121	3.25	.383
Japanese (82)	117	125	121.15	121	121	2.45	.574
Chinese (84)	119	145	121.62	121	121	9.44	.622
Atayal (72)	119	127	121.69	121	121	4.75	.736
Ami (62)	119	123	120.71	121	121	2.34	.649
Oceania:							
Micronesians (52)	113	169	125.42	121	121	209.31	.717
Nasioi (46)	119	123	121.96	121	121	.98	.392
Papua New Guineans (190)	115	173	123.66	121		134.65	.642
New World:							
Cheyenne (106)	119	125	121.27	121	121	1.56	.489
Maya (102)	115	125	120.92	121	121	2.92	.597
Karitiana (106)	119	123	121.00	121	121	.16	.075
Ticuna (126)	117	131	120.43	121	121	2.74	.481
Rondonian Surui (90)	121	125	121.09	121	121	.35	.043
$(CA)_n$-2:							
African:							
Wolof (124)	107	153	126.97	131	115	190.18	.898
Ewondo (64)	109	159	127.34	131	141	97.81	.899
Bamileke (62)	109	157	130.42	131	143	221.56	.905
Mbuti (76)	113	155	133.79	139	115	153.50	.888
Biaka (130)	109	151	127.49	125	123	134.20	.911
Bantu-speakers (84)	109	165	127.05	125	113	222.36	.893
Herero (86)	109	167	128.88	129	113	202.10	.906
Kwengo (72)	109	145	125.97	129	129	120.22	.875
Nama (50)	109	145	125.44	125	113	99.96	.891
Va/Sekele !Kung San (58)	109	157	126.45	127	129	144.67	.908
Zu/Wasi !Kung San (58)	109	157	125.14	125	113	129.94	.885
Somali (54)	109	145	119.70	115	113	80.48	.826
Ethiopian Jews (118)	109	151	125.83	125	113	121.05	.711
Middle East/Europe:							
Yemenite Jews (142)	105	139	120.68	123	113	54.68	.770
Druze (94)	109	139	118.53	117	113	34.27	.766
Finns (70)	113	131	116.97	117	113	13.48	.711
Danes (100)	105	141	118.94	117	113	48.11	.797
Asia:							
Yakut (82)	107	139	115.51	113	113	25.86	.673

(*continued*)

Table 3 Continued

Population (No. of Chromosomes)	Minimum	Maximum	Mean	Median	Mode	Variance	Heterozygosity
Japanese (84)	109	131	116.71	113	113	16.68	.624
Chinese (84)	109	139	118.40	115	113	21.24	.787
Atayal (80)	113	127	118.88	115	113	33.15	.754
Ami (72)	113	149	118.08	115	113	70.19	.733
Oceania:							
Micronesians (72)	113	131	118.58	117	113	35.37	.793
Nasioi (46)	113	143	120.70	117	115	50.48	.798
Papua New Guineans (180)	113	145	123.33	127	127	41.54	.774
New World:							
Cheyenne (86)	113	127	114.35	113	113	12.27	.345
Maya (78)	109	127	113.79	113	113	10.77	.448
Karitiana (100)	113	115	113.16	113	113	.30	.147
Ticuna (90)	113	123	113.13	113	113	1.15	.044
Rondonian Surui (68)	113	147	113.56	113	113	17.18	.058

viate significantly from HW expectations are the South African Bantu-speakers ($P < .0001$), Namibian Nama ($P < .0001$), Namibian Kwengo ($P < .02$), Namibian Va/Sekele !Kung San ($P < .008$), Namibian Zu/Wasi !Kung San ($P < .0001$), Papua New-Guineans ($P < .04$), and the Rondonian Surui ($P < .01$) (table 3). All seven of these populations have an excess of homozygosity. With the exception of the Cheyenne and Bantu-speakers, populations exhibiting a deviation from HW expectations for both STRPs have small population sizes. Additionally, the Rondonian Surui are a small Amazonian Amerindian tribe with high levels of consanguinity within the sample (Calafell 1999), which could account for deviation from HW expectations. It is also possible that the deficit of heterozygotes in some of these populations may indicate the presence of "null alleles" that, because of either mutations within the primer sequences or preferential amplification of an allele, we are not detecting; however, the fact that the Kwengo, Va/Sekele, Zu/Wasi, and Surui have a deficit of heterozygotes for both the $(CA)_n$-1 and $(CA)_n$-2 polymorphisms makes this possibility less likely and suggests that the deviation from HW expectations may be due to population-level effects (substructure, admixture, or inbreeding) or to chance deviations, rather than to the presence of "null alleles."

$(CA)_n$-1 and $(CA)_n$-2 allele sizes for nonhuman primates are within the range observed in humans but are less variable (table 4). The increased variability in humans may be due to the larger sample size examined, the bias resulting from selection of these STRPs on the basis of high heterozygosity in humans, and/or greater instability of the STRPs in humans, who have larger numbers of perfect repeats than do chimpanzees and gorillas (table 2).

Haplotype Diversity

Each of the $(CA)_n$ repeats was analyzed separately as two-site haplotypes with the *Alu* polymorphism. Haplotype frequencies for both $(CA)_n$-1/*Alu* and $(CA)_n$-2/*Alu* haplotype systems, as well as haplotype-diversity values, are in Appendices A and B at the Kidd Lab Home Page and the Tishkoff lab Web page. Histograms of haplotype frequencies from a subset of geographically diverse populations are shown in figure 3. The total number of haplotypes and haplotype-diversity statistics for each major geographic region are given in table 5. For both $(CA)_n$-1/*Alu* and $(CA)_n$-2/*Alu* haplotype systems, both the total number of haplotypes and haplotype diversity are greatest in African populations, lower in Middle Eastern, European, and Asian populations, and lowest in Pacific Island and New World populations (fig. 3 and table 5). There are many more haplotypes specific to African populations than to non-African populations (table 5). In general, non-African populations have a subset of the haplotype diversity observed in Africa. This pattern of variation is consistent with results from the CD4 (Tishkoff et al. 1996*a,* 1998*b*), DM (Tishkoff et al. 1998*a;* S. A. Tishkoff, A. G. Clark, and T. Jenkins, unpublished data), DRD2 (Kidd et al. 1998), and PAH (Kidd et al. 2000) loci. In addition, African populations have highly divergent patterns of haplotype variation, whereas the non-African populations share a more similar pattern of haplotype variation. However, genetic drift has resulted in some divergent haplotype distributions in isolated populations from all regions of the world (fig. 3; also see Appendices A and B at the Kidd Lab Home Page and the Tishkoff lab Web page).

The Papua New Guinean and Micronesian populations have low frequencies (8% and 13% frequency,

Table 4

Frequencies of PLAT $(CA)_n$-1 and $(CA)_n$-2 Alleles in Nonhuman Primates

	No. of Chromosomes in		
STRP Allele Size	*Pan troglodytes*	*Pan paniscus*	*Gorilla gorilla*
$(CA)_n$-1:			
105	0	0	6
108	0	0	2
111	1	0	0
119	6	6	0
121	7	4	0
123	4	0	0
125	4	0	0
127	2	0	0
129	4	0	0
131	3	0	0
139	3	0	0
141	2	0	0
Total	36	10	8
$(CA)_n$-2:			
103	1	0	0
107	1	0	0
109	4	1	3
111	7	5	0
113	6	1	0
115	6	0	1
117	1	3	0
119	1	0	4
121	2	0	0
123	5	0	0
Total	34	10	8

respectively) of haplotypes containing large-sized (149–173 bp) $(CA)_n$-1 alleles, which are absent in all other non-African populations studied (in which the maximum allele size is 145 bp). Such large alleles, however, are present at low to moderate frequencies in sub-Saharan African populations, which have a broader and more continuous distribution of alleles ranging up to 171 bp in size (fig. 3 and table 3). The very large number of dinucleotide repeats in the large-sized alleles made cloning and sequencing analysis difficult. However, sequence analysis of a 161-bp allele from a Papua New Guinean individual and a 139-bp allele in a Bantu-speaking individual demonstrates that both contain perfect repeats with no interruptions (table 2). Thus, it is not possible to distinguish whether these alleles are identical by descent or arose independently because of recurrent mutation.

PCA was used to provide a visual representation of population clustering based on haplotype variation (fig. 4). The first principal component (*X*-axis) accounts for 55.1% of the variance, and the second principal component (*Y*-axis) accounts for 10.5% of the variance. In general, populations cluster by geographic origin. The most distinct separation is between African and non-African populations. The northeastern-African—that is, the Ethiopian and Somali—populations are located centrally between sub-Saharan African and non-African populations. Among the non-Africans, populations cluster by geographic region, with some exceptions; the Siberian Yakut and the Japanese populations cluster with the North American populations and the Papua New Guinean and Taiwanese Ami populations have identical coordinates (fig. 4). On the PCA plot, the three Amazonian Amerindian populations (Ticuna, Surui, and Karitiana) and the Nasioi Melanesian population appear to be the groups that are most isolated from other geographically nearby populations, possibly because of high levels of genetic drift among these small and isolated populations.

Population Subdivision

The amount of population subdivision was quantified with three statistics that capture different aspects of the data as described in the Subjects and Methods section. These estimates were obtained from populations grouped by geographic region; results are shown in table 6. F_{ST} values are lowest for European populations for the $(CA)_n$-1/*Alu* haplotype system and are lowest for the New World populations for the $(CA)_n$-2/*Alu* haplotype system; F_{ST} values are highest for the Oceanic and New World populations for the $(CA)_n$-1/*Alu* haplotype system and for the Asian and Oceanic populations for the $(CA)_n$-2/*Alu* haplotype system. However, F_{ST} is highly biased by within-population diversity levels. As can be seen by the formula for estimation of Wright's F_{ST} (see the Subjects and Methods section), F_{ST} values will always be low when within-population diversity levels (i.e., heterozygosity levels) are high. Therefore, F_{ST} is not a good measure to use for comparison of levels of subdivision for highly variable STRP systems that vary widely in heterozygosity levels across geographic regions. The fact that African populations have F_{ST} values lower than those in other regions of the world is likely a result of the very high levels of STRP diversity within these populations, rather than a reflection of population history, subdivision, and migration.

Two measures of population subdivision that are particularly useful for highly variable STRP systems are D_{SW}, which takes into account the stepwise-mutation process of STRPs (Shriver et al. 1995), and a likelihood-ratio test—D_{LR}—described by Paetkau et al. (1997). Values for both D_{SW} and D_{LR} are lowest for the Oceanic populations and are highest for the African populations. These results demonstrate considerably higher haplotype heterogeneity among African populations than among populations from Europe and the Middle East, Asia, Oceania, or the New World.

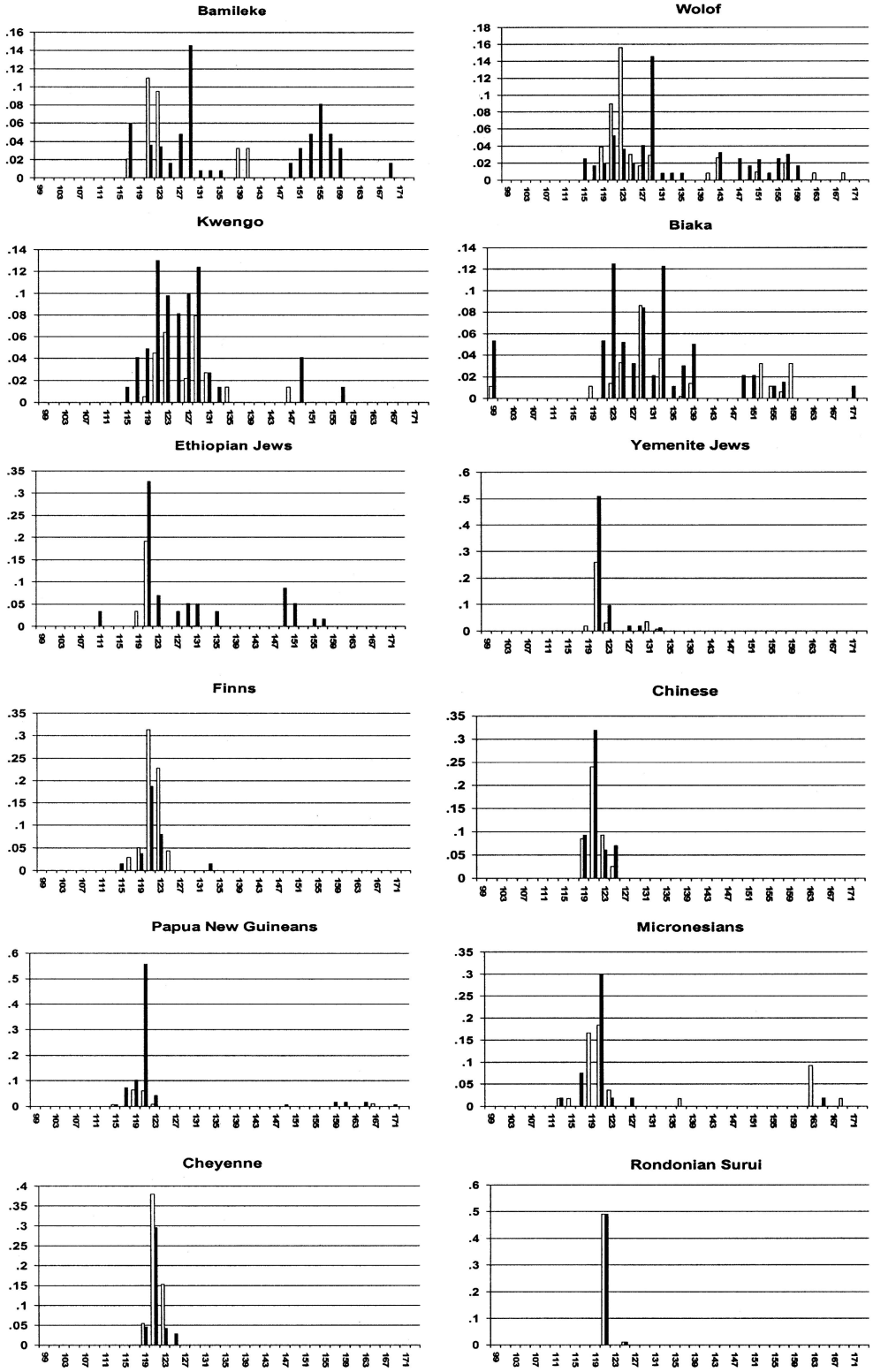

Figure 3 Haplotype-frequency distributions of $(CA)_n$-1 STRPs (*this page*) and $(CA)_n$-2 STRPs (*following page*), on *Alu*(+) and *Alu*(−) chromosomes from a globally diverse subset of the populations included in this study. STRP allele sizes are shown on the *X*-axis. *Alu*(+) haplotypes are shown as unblackened bars, and *Alu*(−) haplotypes are shown as blackened bars.

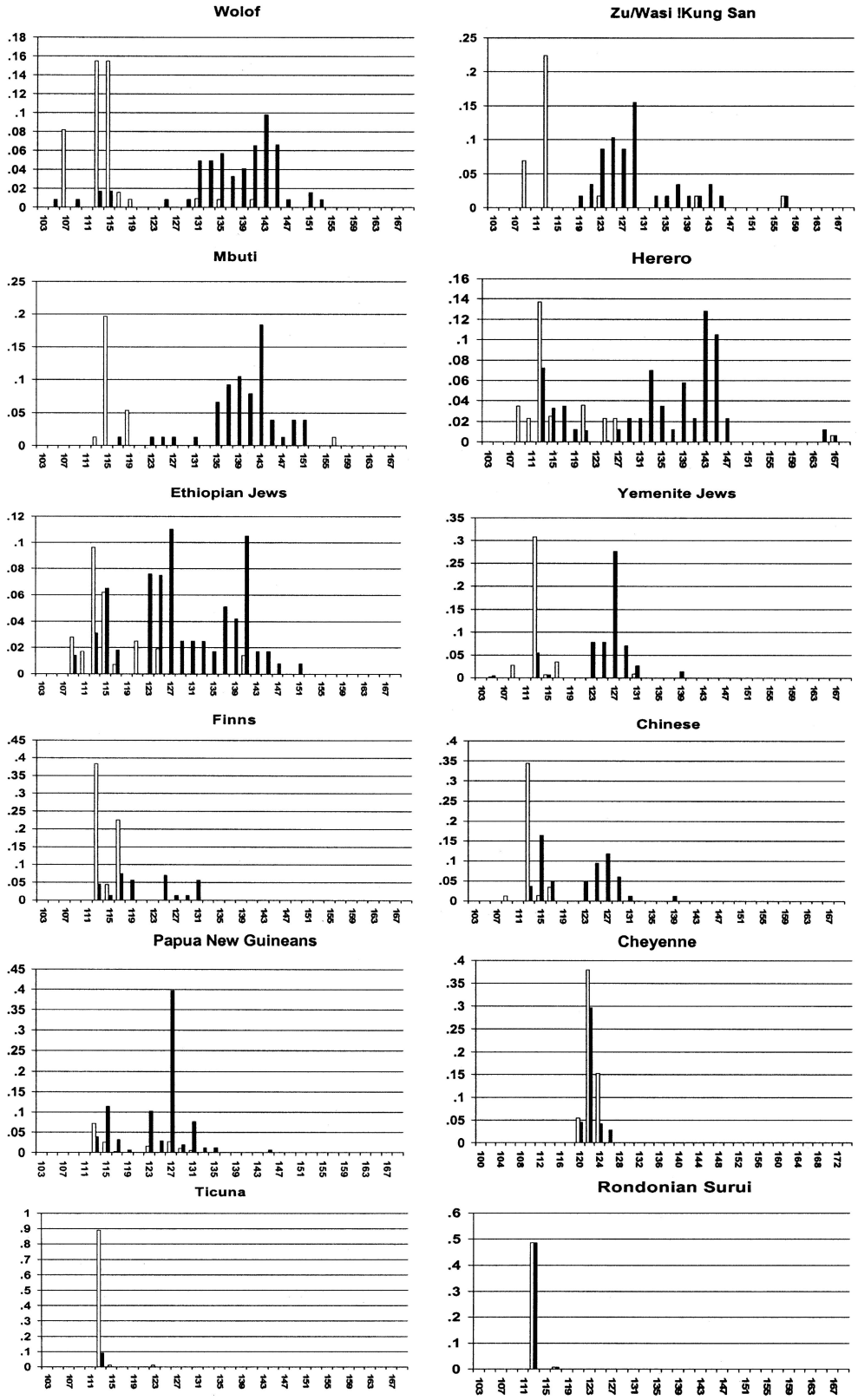
Wolof
Zu/Wasi !Kung San
Mbuti
Herero
Ethiopian Jews
Yemenite Jews
Finns
Chinese
Papua New Guineans
Cheyenne
Ticuna
Rondonian Surui

Table 5

Haplotype Diversity Statistics for Populations Grouped by Geographic Region

	African	Middle Eastern/ European	Asian	Oceanic	New World	Non-African[a]
$(CA)_n$-1:						
No. of populations	13	4	5	3	5	17
No. of chromosomes	924	352	386	284	540	1,562
Mean heterozygosity	.903	.739	.772	.639	.614	.694
No. of haplotypes	54	21	12	22	11	35
Mean no. of haplotypes/population	21.54	11	7.6	11.33	6.20	8.64
No. of population-specific haplotypes	26	0	0	4	0	4
$(CA)_n$-2:						
No. of populations	13	4	5	3	5	17
No. of chromosomes	1,030	410	422	296	488	1,616
Mean heterozygosity	.909	.8	.767	.809	.506	.705
No. of haplotypes	48	22	24	21	14	37
Mean no. of haplotypes/population	21.69	14.25	12	13.667	5.80	11
No. of population-specific haplotypes	15	1	1	1	0	3

[a] Includes pooled data from Europe, Asia, Oceania, and the New World.

Patterns of LD

Fisher's exact test (Sokal and Rohlf 1995, pp. 730–736) was used to test for significance of LD between each STRP allele and the *Alu* indel polymorphism (table 7). If allele frequencies are sufficiently low, some alleles may fail to ever show statistically significant LD (Lewontin 1995), and these cases are indicated by blank spaces in table 7. Significance values in table 7 are not corrected for multiple comparisons. For the $(CA)_n$-1/*Alu* haplotype system, with markers located ~22 kb apart, there are sporadic cases of LD in both African and non-African populations (table 7). However, the pattern of LD in Africans and non-Africans is distinct. In 6 of the 13 African populations, there is an association between *Alu*(+) and either 121- or 123-bp alleles. However, in the non-African populations, only 3 of 18 populations exhibit this association, and, in 6 populations, the *Alu*(+) allele is most strongly associated with a 119-bp allele. The latter association is never observed in African populations. *Alu*(−) chromosomes are most frequently associated with 117–125-bp alleles in non-Africans (but rarely in Africa), whereas the *Alu*(−) chromosomes are most frequently associated with a 129-bp allele in Africans and in Middle Eastern Druze (this allele is rare or absent in all other non-African populations). For the $(CA)_n$-2/*Alu* haplotype system, with markers located ~12.2 kb apart, we observe a strong positive association between the 113-bp and *Alu*(+) alleles in 21 of the 31 African and non-African populations (table 7). However, for many other STRP alleles, Africans and non-Africans have distinct patterns of both allelic variation and LD (see table 7).

Overall, there are fewer significant LD values (either positive or negative) in African populations than in non-African populations, relative to the number of alleles present at frequencies high enough to allow detection of significance. Also, the number of pairwise associations is considerably greater for the $(CA)_n$-2/*Alu* haplotype system than for the $(CA)_n$-1/*Alu* haplotype system. This pattern of LD is also demonstrated by the histograms of haplotype frequencies, shown in fig. 3; both the *Alu*(+) and *Alu*(−) chromosomes are associated with a wide range of STRP alleles for the $(CA)_n$-1/*Alu* haplotype system. By contrast, for the $(CA)_n$-2/*Alu* haplotype system, the *Alu*(+) allele tends to be associated predominantly with small-sized alleles (107–127-bp alleles, with 113 bp being most common) whereas the *Alu*(−) is associated with many alleles in the 105–169-bp range.

Discussion

These results demonstrate the usefulness of haplotype analysis of STRPs, in combination with more-stable indels, for the reconstruction of human population history. Recently retroposed *Alu* elements are valuable tools for the reconstruction of historical population and demographic events, since they are unique, unidirectional, and stable mutation events for which the ancestral state (i.e., absence of the *Alu* element) is known. Additionally, analysis of *Alu* elements as haplotypes with highly variable STRP repeats makes it possible to estimate the time of insertion of the *Alu* element, on the basis of erosion of LD (A. G. Clark and S. A. Tishkoff, unpublished data). Knowledge of insertion times can be important for inference of the timing of historical evolutionary events and for reconstruction of human demographic history (Sherry et al. 1997).

Haplotype History at the PLAT Locus

STRP/*Alu* haplotype analysis can be useful for inference of the evolutionary history of a locus of interest. The absence of the HS *Alu* in chimpanzees and gorillas,

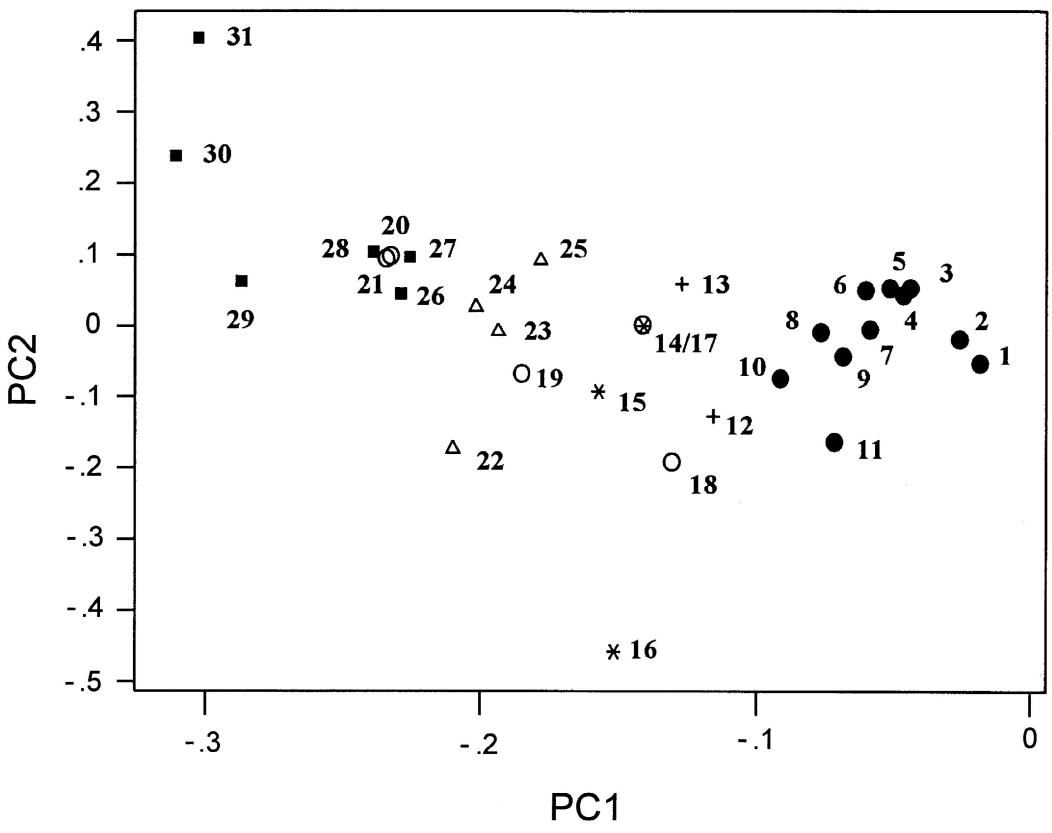

Figure 4 PCA plot of populations, based on haplotype-frequency variation. ● = Sub-Saharan African populations; + = northeastern-African populations; △ = European and Middle Eastern populations; * = Oceanic populations; ○ = Asian populations; ■ = New World populations; 1 = Biaka; 2 = Mbuti; 3 = Bantu-speakers; 4 = Ewondo; 5 = Bamileke; 6 = Wolof; 7 = Herero; 8 = Zu/Wasi !Kung San; 9 = Kwengo; 10 = Nama; 11 = Va/Sekele !Kung San; 12 = Ethiopians; 13 = Somali; 14= Papua New Guineans; 15 = Micronesians; 16 = Nasioi Melanesians; 17 = Ami; 18 = Atayal; 19 = Chinese; 20 = Japanese; 21 = Yakut; 22 = Yemenites; 23 = Danes; 24 = Druze; 25 = Finns; 26 = Maya; 27 = Cheyenne; 28 = Ticuna; 29 = Surui; 30 = Karitiana.

as well as the fact that the *Alu* element is highly variable for presence/absence in populations from globally diverse regions, suggests that the insertion of this retroposable element occurred after the divergence of humans from the great apes, which occurred ~5 million years ago (Wilson and Sarich 1969; Kumar and Hedges 1998), but prior to the divergence of modern human populations, which occurred during the past 150,000 years (Stringer and Andrews 1988; Harpending et al. 1993; Cavalli-Sforza et al. 1994; Goldstein et al. 1995; Nei 1995; Tishkoff et al. 1996*a,* 1996*b,* 1998*b;* Stoneking et al. 1997). The two dinucleotide repeats $(CA)_n$-1 and $(CA)_n$-2, which are located ~22 and ~12.2 kb, respectively, from the polymorphic *Alu,* are variable in common chimpanzees (*P. troglodytes*), pygmy chimpanzees (*P. paniscus*) and gorillas (*G. gorilla*). The allele sizes observed in nonhuman primates are similar to those observed in humans. However, sequence analysis of the STRPs in humans, chimpanzees, and gorillas demonstrates that alleles that (on the basis of PCR-product length) are found to be of similar size are heterogeneous at the sequence level (table 2). These results confirm and extend previous studies demonstrating sequence heterogeneity of STRP alleles (Blanquer-Maumont and Crouau-Roy 1995; Garza and Freimer 1996; Grimaldi and Crouau-Roy 1997; Rosenbaum and Deinard 1998; Deinard and Kidd 1999). Sequence analysis of $(CA)_n$-1 in the gorilla also demonstrates that deletions outside the repeat unit can affect the size of the PCR product. Thus, one must be cautious about making inferences about the evolutionary history of different taxa based on sizes of PCR amplicons of STRPs without sequence analysis of alleles. Additionally, our results demonstrate that slippage by *Taq* polymerase during PCR amplification can result in expansion or contraction of either of the two repeats composing a "compound" STRP. Thus, multiple clones from multiple individuals must be sequenced, to obtain a "consensus sequence" of the dinucleotide composition of compound STRPs. Such knowledge could be important for determining whether

Table 6

Estimates of Genetic Heterogeneity within Geographic Regions

	F_{ST} (P)	D_{SW} (P)	D_{LR}(P)
$(CA)_n$-1/*Alu:*			
Africa	.050 (.000)	1.974 (.000)	3.175 (.000)
Europe	.029 (.001)	.139 (.002)	2.312 (.041)
Asia	.045 (.000)	.242 (.000)	2.088 (.055)
Oceania	.072 (.020)	.084 (.000)	.340 (.000)
New World	.078 (.000)	.399 (.000)	1.932 (.000)
Global	.094 (.000)	12.541 (.000)	7.170 (.000)
$(CA)_n$-2/*Alu:*			
Africa	.042 (.000)	1.968 (.000)	3.280 (.000)
Europe	.041 (.000)	.153 (.000)	1.215 (.016)
Asia	.074 (.000)	.268 (.000)	1.660 (.039)
Oceania	.062 (.000)	.101 (.000)	.391 (.000)
New World	.033 (.009)	.576 (.000)	2.616 (.014)
Global	.106 (.000)	15.202 (.000)	8.641 (.000)

NOTE.—*P* values are the tail probability that a heterogeneity estimate greater than that observed in a panmictic population will be obtained. Empirical null distributions under panmixia within each region were obtained by a permutation test (Hudson et al. 1992).

alleles are identical by descent. However, even if STRP alleles are identical at the sequence level, this is not absolute proof of identity by descent, since identical alleles could have arisen by independent convergent mutation. In such a case, identity by descent can be determined only by detailed analysis of genetic diversity in the sequence flanking the STRP.

Initially, the *Alu* insertion into intron 8 of the *TPA* gene occurred on a chromosome containing a single allele at each of the neighboring dinucleotide markers. The "TPA *Alu*" likely inserted into a chromosome containing a small-sized $(CA)_n$-2 allele (possibly the 113-bp allele, which is most common in modern populations). The close proximity of the *Alu* marker and the $(CA)_n$-2 marker (which are 12.2 kb apart) has resulted in the maintenance of strong LD between the *Alu*(+) allele and the 113-bp allele. The unimodal distribution of $(CA)_n$-1 and $(CA)_n$-2 alleles on *Alu*(+) and *Alu*(−) haplotype backgrounds in most populations is consistent with a stepwise-mutation model in which mutations result in expansion or contraction by one or a few repeat units, as has been observed at many other STRP repeats in humans (Shriver et al. 1993; Valdes et al. 1993; Weber and Wong 1993; Di Rienzo et al. 1994, 1998). The greater variance of STRP alleles on *Alu*(−) chromosomes compared with *Alu*(+) chromosomes (table 3) is consistent with a recent origin of the *Alu*-insertion allele. LD between $(CA)_n$-1 and $(CA)_n$-2 can erode by mutation as fast as by recombination. The variance of STRP alleles is similar for both $(CA)_n$-1 STRPs and $(CA)_n$-2 STRPs (table 3), suggesting that their mutation rates may be similar. Additionally, at the sequence level, the numbers of perfect repeats in a 121-bp $(CA)_n$-1 allele and those in a 113 bp $(CA)_n$-2 allele (both of which are the alleles most common on *Alu*(+) chromosomes) are nearly identical, which makes it less likely that the differential pattern of LD observed could be due to fewer perfect repeats producing a lower mutation rate of the 113-bp $(CA)_n$-2 allele. Thus, it seems likely that recombination has resulted in low levels of LD between $(CA)_n$-1 and the *Alu* polymorphism (which are 22 kb apart), whereas high levels of LD have been maintained between $(CA)_n$-2 and the *Alu* polymorphism (which are 12.2 kb apart). Because of the loss of LD between $(CA)_n$-1 and the PLAT *Alu*, the haplotype history has become obscured, and it has become difficult to infer the ancestral $(CA)_n$-1/*Alu* haplotype. However, the narrower distribution of $(CA)_n$-1 alleles on *Alu*(+) chromosomes compared with *Alu*(−) chromosomes, with most alleles being 119–129 bp in size, suggests that the *Alu*-insertion event occurred on a chromosome containing a $(CA)_n$-1 allele in this smaller size range, possibly on a chromosome containing a 121-bp allele, which is the most common *Alu*(+) haplotype in Africa (fig. 3). Analysis of variation of the two dinucleotide repeats on *Alu*(+) and *Alu*(−) haplotypes, as well as of the breakdown of disequilibrium between the markers, suggests that the *Alu* insertion occurred during the past 500,000 years (A. G. Clark and S. A. Tishkoff, unpublished data), whereas the global distribution of the *Alu*(+) allele suggests that the insertion event occurred prior to migration of modern humans out of Africa, ~100,000 years ago.

Human Evolutionary History

The $(CA)_n$-1/*Alu* and $(CA)_n$-2/*Alu* haplotype systems are highly informative for reconstructing historical migration and population-differentiation events, as demonstrated by the PCA plot of population-clustering based on haplotype-frequency variation (fig. 4). These results are consistent with anthropological knowledge, results from studies of classical markers (Nei and Roychoudhury 1993; Cavalli-Sforza et al. 1994; Nei 1995), and results from molecular markers from autosomes (Bowcock et al. 1991, 1994; Jorde et al. 1995, 1997; Nei 1995; Armour et al. 1996; Tishkoff 1996*a*, 1998*a*, 1998*b*; Harding et al. 1997; Stoneking et al. 1997; Zietkiewicz 1997, 1998; Calafell et al. 1998; Kidd et al. 1998, 2000; Harris and Hey 1999), mtDNA (Cann et al. 1987; Vigilant et al. 1989, 1991; Merriwether 1991; Penny et al. 1995), and Y chromosome DNA (Hammer 1995; Hammer et al. 1997, 1998; Underhill et al. 1997). These studies suggest a recent and primary subdivision between African and non-African populations, high levels of divergence among African populations, and a recent shared common ancestry of non-African populations, from a population originating in Africa.

Table 7

Fisher's Exact Test of Allele-Specific Pairwise LD, Estimated by Comparisons of Each Microsatellite Allele and *Alu*(+) and *Alu*(−) Alleles

Haplotype System and Population	Results of Fisher's Exact Test of *D* for STRP and *Alu* Alleles[a]																														
	99	109	111	113	115	117	119	121	123	125	127	129	131	133	135	137	139	141	143	145	147	149	151	153	155	157	159	161	163	165
$(CA)_n$-1/*Alu*:																														
African:																														
Biaka (94)	.							.	.	.	.	+*		.		.	.							.	.		.			
Mbuti (68)							.	.	−*	.		−*										.		.		.				
Herero (86)								.	.	.	.	−*	.	.								−*			.	.			+***	
Kwengo (74)						.	.	.	.	.	.	.	.									.								
Nama (48)					.	.		.		.		.	+*		.										.					
Sekele (60)						.	.	−*	.	.		+*	.								.									
Zu/Wasi (58)							.	.	.	.	.	+**																		
Bantu (70)						.	.	+***	+**		.	−*	−*	.							.	.		.	.		.			
Ewondo (68)						.		.	+***	.	.	−**								.		.		.	.	.				
Bamileke (62)						.		+**	+*		.	.	.	.	.									.	.	.				
Wolof (120)					.	.	.	.	+***	.	.	−**							.		.	.	.		.	.	.				
Somali (58)							.	+**	.			−*																		
Ethiopians (58)								+**	.			.	.									.	.							
Non-African:																														
Yemenites (74)								.	.				.																	
Druze (110)							.	.	.			−**	.																	
Danes (100)						−*	+*	.	+*			.		.	.															
Finns (68)							.	.	.	.																				
Chinese (84)							.	.	.	.																				
Japanese (82)							+**	.	.	−**																				
Yakuts (78)							+*	.	.																					
Ami (70)							.	.	.																					
Atayal (72)							.	−***	+**	.																				
Micronesians (58)				.		−*	+**	−**	.																				+*	
Nasio (48)							+*	.	.																					
Papua New Guineans (178)					.	−***	.	.	.																		.	.		.
Cheyenne (108)							.	.	+*	.																				
Maya (102)					.		+***	−*	.	.																				
Karitiana (104)								.																						
Surui (92)								.		.																				
Ticuna (134)						.	.	.	.																					

	103	105	107	109	111	113	115	117	119	121	123	125	127	129	131	133	135	137	139	141	143	145	147	149	151	153
(CA)$_n$-2/*Alu*:																										
African:																										
Biaka (130)				+***		.	.	+*	+*		−**	.	.		.		.		.	.	.			.		
Mbuti (76)							+***		+**								.	.	.	.	−**	.		.	.	
Herero (86)				+*		+***	.	.		+*		.	.			.	.		.		−*	−*				
Kwengo (74)				+*		+***	+*			.	.			−**	.	.		.	.		.	.				
Nama (50)						+***		.			.		.			.			.		.					
Sekele (58)				+*	+**	+***					.		.	.		.			.							
Zu/Wasi (58)				+*		+***					.	.	.	−*												
Bantu (84)				.		+***	+*	.					.			.	.		.	.	.	.				.
Ewondo (62)			+***			+***	.								.	.	.		.	−**		.				
Bamileke (60)			+*		+**	+***								.	.					.	−*	.				
Wolof (122)			+***			+***	+***								.	.	−*	.	.	−*	−**	−**				
Somali (52)			.			+***	−*	.			.			−*												
Ethiopians (118)				.	.	+***	+*	.		.	.	.	−*	.	.	.	.	.	.	.	.	.				
Non-African:																										
Yemenites (146)				.		.	.	.			.	.	.	.	.				.							
Druze (94)						+***	.	+**	.		.	−***														
Danes (100)		.				+***	+**	.			−*	−*	−***	− **												
Finns (70)						+***	.	.	−*			−*			−*											
Chinese (84)						+***	−**	.			.	−*	**	.												
Japanese (84)						+***	−***	.	.		−*		−***		.											
Yakuts (102)			+*			+***	.	−***	−*	.		−*														
Ami (72)						−*	+**	−*				.	.							.						
Atayal (80)						+***	.	.				−***	.													
Micronesians (76)						+***	−***	+***			.	.	−***	.												
Nasioi (48)						.	.	.				.		.	.											
Papua New Guineans (172)						.	.	−***			−***	−***	.	.	−***	−***	−***									
Cheyenne (86)						.	.	.					−*													
Maya (78)				.		.	.	.					.													
Karitiana (100)						+***	−***																			
Surui (90)						.																				
Ticuna (134)						.																				

[a] A dot (.) denotes that the allele is present at a frequency high enough to allow a valid test of LD to be performed (Lewontin 1995); blank cells denote either allele absence or allele frequency so low that a significant test could not be performed; a plus sign (+) indicates that there is an overrepresentation of that STRP allele on an *Alu*(+) haplotype background, and a minus sign (−) indicates that there is an overrepresentation of that STRP allele on an *Alu*(−) haplotype background.

* $P = .05$.

** $P = .01$.

*** $P = .001$.

The intermediate position, between African and non-African populations, that the Ethiopian Jews and Somalis occupy in the PCA plot also has been observed in other genetic studies (Ritte et al. 1993; Passarino et al. 1998) and could be due either to shared common ancestry or to recent gene flow. The fact that the Ethiopians and Somalis have a subset of the sub-Saharan African haplotype diversity—and that the non-African populations have a subset of the diversity present in Ethiopians and Somalis—makes simple-admixture models less likely; rather, these observations support the hypothesis proposed by other nuclear-genetic studies (Tishkoff et al. 1996*a,* 1998*a,* 1998*b;* Kidd et al. 1998)—that populations in northeastern Africa may have diverged from those in the rest of sub-Saharan Africa early in the history of modern African populations and that a subset of this northeastern-African population migrated out of Africa and populated the rest of the globe. These conclusions are supported by recent mtDNA analysis (Quintana-Murci et al. 1999). In light of the high variance expected on the basis of the stochastic effects of genetic drift at a single locus, it is particularly striking that the PCA plot of populations so accurately reflects population relationships as predicted on the basis of common ancestry as well as on the basis of geographic and linguistic similarity. For example, in the PCA plot of African populations (fig. 4), the Khoisan speakers (Nama, Va/Sekele, Zu/Wasi, and Kwengo) all cluster together, the Biaka and Mbuti populations cluster together, and the western-African, central-African, and southern-African Bantu-speaking populations cluster together. The latter cluster is predicted in light of linguistic evidence and the archeological record that suggest a Bantu migration from central Africa during the past 3,000 years (Clark 1959; Guthrie 1962; Greenberg 1963; Nurse et al. 1985). Additionally, the three Amazonian populations (Surui, Karitiana, and Ticuna) cluster together, as do the two North and Central American populations (Cheyenne and Maya). Interestingly, the Siberian Yakut and Japanese populations cluster with these North American Indian populations. This is consistent with the hypothesis, based on linguistic, archeological, and genetic data, that modern Amerindians may have originated from one or more populations migrating from Siberia or other regions in northeastern Asia (Kidd et al. 1991; Kidd and Kidd 1996; Kolman et al. 1996; Calafell et al. 1998; Starikovskaya et al. 1998; Karafet 1999; Schurr et al. 1999). The fact that these haplotype systems are so informative for the inference of population history is likely due to the combined use of highly variable STRP markers with the stable *Alu*-insertion polymorphism.

Tests of population subdivision that consider the high variability of STRP haplotypes (D_{SW} and D_{LR}) demonstrate that Africa has levels of population subdivision that are higher than those of any other geographic region (table 6). The high heterozygosity levels, high STRP variance, and high haplotype-diversity levels, within and between African populations, compared with populations from all other geographic regions (table 5), suggest that African populations have an older population history and have maintained both a large effective population size and high levels of population subdivision. These results are consistent with results of studies of other STRP haplotype systems (Tishkoff et al. 1996*a,* 1998*a,* 1998*b;* Kidd et al. 1998; S. A. Tishkoff, A. G. Clark, and T. Jenkins, unpublished data), *Alu* elements (Sherry et al. 1997; Stoneking 1997; Harpending et al. 1998), STRPs (Shriver et al. 1997; Calafell et al. 1998; Kimmel et al. 1998; Reich and Goldstein 1998; Relethford and Jorde 1999), Y-chromosome DNA (Hammer et al. 1997, 1998; Pritchard et al. 1999; Scozzari et al. 1999), mtDNA (Rogers and Harpending et al. 1992; Sherry et al. 1994; Rogers and Jorde 1995), and craniometric data (Relethford and Harpending 1994; Relethford 1995), all of which suggest that African populations (*a*) may have expanded in size earlier than have non-African populations and (*b*) have maintained a larger effective population size. Africans have many more region-specific haplotypes, and, in general, non-Africans have a subset of the haplotypes present in Africa. Compared with African populations, the non-African populations have less haplotype diversity, more-extensive LD, and a similar pattern of haplotype variation across geographic regions (e.g., Asia, Europe, Oceania, and the New World) (fig. 3 and table 7). These results are consistent with both an appreciable founder effect associated with migration out of Africa and a recent shared common ancestry of non-African populations that has been followed by rapid population expansion. In addition, there is no indication that non-African populations have descended from multiple migrations from different source populations in Africa; rather, the shared pattern of variation among geographically diverse non-African populations supports previous findings (Tishkoff et al. 1996*a,* 1998*a,* 1998*b;* Calafell et al. 1998; Kidd et al. 1998) that indicate that all non-African populations have descended from a single source population, most likely from northeastern Africa (although subsequent gene flow from Africa to Australo-Melanesia remains a possibility, as is discussed below).

The pattern of LD between the $(CA)_n$-1 and $(CA)_n$-2 STRPs and the *Alu* polymorphism is distinct in Africans vis-à-vis non-Africans (table 7). With respect to the number of alleles present at high frequency, the African populations have more haplotypes, more-divergent patterns of LD, and, overall, fewer significant LD values than do non-Africans. These results are consistent with the pattern of LD observed at several loci: CD4 (Tishkoff et al. 1996*a,* 1998*b*), DM (Tishkoff et al. 1998*a*), DRD2 (Kidd et al. 1998), and PAH (Kidd et al. 2000). Levels

and patterns of LD depend on a number of factors, including initial conditions (e.g., population size), population structure, founder effect, admixture, and the dynamics of molecular processes of mutation and recombination. If mutation and recombination rates are assumed to have remained constant across populations, then the global pattern of LD observed at these loci is a reflection of the population and demographic histories of these populations. The pattern of LD in African populations is consistent with a larger, more subdivided population structure in Africa. The divergent pattern of LD in non-African populations relative to African populations is likely the result of a founding event by one or more small populations emerging from Africa during the past 100,000 years and expanding and spreading throughout the rest of the world (Stringer and Andrews 1988; Stringer 1993). During this founding event, the particular pattern of pairwise allelic association may differ from that in the parental African population, because of the stochastic effects of drift during population founding. The pattern of LD established at the time of population founding will be preserved during subsequent rapid population expansion, because of the decreased effects of genetic drift. However, there are some exceptions. For the $(CA)_n$-2/*Alu* haplotype system, the Oceanic and several New World populations lack the strong association, present in most other populations, between the 113-bp allele and the *Alu*(+) allele. This is due to an increase in the frequency of 113/*Alu*(−) chromosomes, possibly because of the effects of genetic drift in small populations that have recently undergone founding events; for example, in the Rondonian Surui sample, the 113/*Alu*(+) and 113/*Alu*(−) haplotypes exist at identical frequencies, resulting in complete linkage equilibrium (fig. 3). Interestingly, in the Amazonian Ticuna sample, only the 113/*Alu*(+) haplotype exists at high frequency (fig. 3), demonstrating the strong effects of genetic drift within these small, isolated tribal populations. These results also demonstrate that genetic drift can result in either the establishment of higher levels of LD, as likely occurred during migration out of Africa, or a decrease in disequilibrium, as occurred in some of the Oceanic and New World populations, for the $(CA)_n$-2/*Alu* haplotype system. Thus, because of the stochastic effects of genetic drift, we anticipate that there will be some heterogeneity among loci, in their global patterns of LD (Peterson et al. 1999). In the case of common SNPs in autosomal genes, in which the variation may predate the expansion out of Africa, patterns of fine-scale LD among sites within genes may show less geographic variation (Clark et al. 1998). Genetic variation that arose after the time of population founding would exhibit the highest variability. A systematic study of patterns of LD of SNPs, indels, and STRPS at multiple loci among geographically diverse populations will thus provide a more precise reconstruction of modern human population and demographic history.

Implications for Oceanic-Island Population History

The high variability of STRP/*Alu* haplotype systems can be particularly informative for the reconstruction of historic migration events. For example, Papua New Guinean and Micronesian populations have haplotypes containing large-sized (149–173 bp) $(CA)_n$-1 STRP alleles that have not been observed in any other non-African populations but that have been observed in sub-Saharan African populations, the latter of which have a continuous distribution of $(CA)_n$-1 alleles up to 171 bp in size. These alleles are present on both *Alu*(+) and *Alu*(−) haplotype backgrounds. This observation is intriguing in light of hypotheses based on archeological, morphological, and genetic evidence that suggest that there could have been early colonization of Papua New Guinea by a distinct wave of migration out of Africa across southern and eastern Asia and into Australo-Melanesia (Birdsell 1967; 1993, pp. 22–23; Lahr and Foley 1995; Harpending et al. 1996; Stoneking et al. 1997). Archeological data indicate that Australia and New Guinea, which, during the Pleistocene, were a single landmass referred to as "Sahul," were occupied as early as 40,000–65,000 years ago (Roberts and Jones 1994; Jones 1995; O'Connell and Allen 1998; Johnson et al. 1999; Miller et al. 1999; Redd and Stoneking 1999). Papua New Guinea highland populations are thought to be the descendants of the earliest migration into this region, whereas coastal Papua New Guinea populations are thought to be admixed with Austronesian speakers originating from Southeast Asia who colonized coastal Papua New Guinea during the past 5,000 years (Bellwood 1989; Redd and Stoneking 1999). The hypothesis of ancient ties between Papua New Guinean and African populations is supported by craniometric studies (Howells 1989, pp. 37–79), as well as by genetic studies of mtDNA variation (Redd and Stoneking 1999), autosomal sequence variation (Kofler et al. 1995), and polymorphic *Alu* loci (Harpending et al. 1996; Stoneking et al. 1997), all of which show Papua New Guinea populations as clustering near the ancestral African root of phylogenetic trees constructed from the data. However, analyses of classical genetic polymorphisms (Cavalli-Sforza et al. 1994) and of haplotype variation at the CD4 (Tishkoff et al. 1996*a*, 1998*b*) and DRD2 loci (Kidd et al. 1998) do not find evidence for a closer relationship of Papua New Guinea populations to Africans than to non-Africans. Thus, these nuclear-haplotype and classical-genetic data sets do not support an independent origin of Australo-Melanesians from a different source population in Africa, but they do not rule out the possibility of an ancient migration event from Africa, for

which the genetic trail has been largely, but not completely, erased by subsequent migrations.

The presence of large-sized $(CA)_n$-1 alleles in Papua New Guinean, Micronesian, and African populations could indicate either shared common ancestry, gene flow from African populations into these Oceanic populations, and/or independent mutation into the large-sized–allele range in African and Oceanic populations. Sequence analysis of a large-sized allele from a Papua New Guinea individual and of a large-sized allele from a Bantu-speaking individual indicate that both alleles consist of perfect repeats. Thus, in the absence of a more detailed analysis of sequence variation in the regions flanking the STRP, it is not possible to distinguish whether they are identical by descent or arose by recurrent mutation. However, because large "jumps" in allele size of STRP repeats are uncommon, it is likely that mutation into the large-sized range would have been a rare and unique event. The fact that we observe a wide range of large-sized alleles (nine alleles 149–173 bp in size) in Papua New Guinean and Micronesian populations, as well as the fact that we observe these alleles on both *Alu*(+) and *Alu*(−) haplotype backgrounds, would indicate that this rare mutational event was ancient and that there has been time for mutation and recombination to produce the haplotype variation observed in modern populations. Alternatively, these chromosomes could have been present at the time of population founding and could have drifted to moderate frequencies, and/or they could have been introduced by recent migration from Africa. We observe, in both highland (48 chromosomes) and lowland (94 chromosomes) Papua New Guineans, haplotypes containing large-sized $(CA)_n$-1 alleles, although the frequency is greater in highlanders (16.8%) than in lowlanders (3.2%). Again, this observation would indicate an ancient origin for these haplotypes. In the PCA based on haplotype-frequency variance (fig. 4), the Papua New Guinean and Micronesian populations cluster closest to the Southeast Asian Chinese Han and to the Taiwanese Ami and Atayal populations, as well as to northeastern-African—that is, the Ethiopian and Somali—populations. The Nasioi appear to be outliers in the PCA plot, but this may be due to the small sample size and/or to high levels of genetic drift in this small, isolated population, as suggested by a number of other genetic studies of this same population (Bowcock et al. 1991; Kidd and Kidd 1996; Tishkoff et al. 1996*a,* 1996*b,* 1998*a,* 1998*b*). Analysis of haplotype variation at the DM locus in these same Papua New Guinean samples indicated the presence, at moderate frequency, of microsatellite haplotypes that are rare in all other non-African populations but that are present in African populations (Tishkoff et al. 1998*a;* S. A. Tishkoff and T. Jenkins, unpublished data). These data are consistent with the hypothesis of an early migration event into Australo-Melanesia, followed by long periods of isolation. However, it is also possible that these haplotypes could have evolved in situ in populations that have been isolated during the past 40,000–65,000 years and are subject to high levels of genetic drift. Additionally, the shared pattern of haplotype diversity and LD—for $(CA)_n$-1/*Alu* haplotypes containing small-sized STRP alleles, for $(CA)_n$-2/*Alu* haplotypes, and for haplotypes at the CD4 loci—suggest that Papua New Guinean populations originated from the same source population as did other non-African populations, although the migration events into different geographic regions may have been distinct and possibly could have occurred at different times. To differentiate among these possibilities, it will be necessary to do, at multiple loci, more-detailed haplotype and sequencing analysis of diverse Papua New Guinean and African populations.

The $(CA)_n$-1/*Alu* haplotype data are also informative for reconstruction of historic migration patterns of remote Oceanic Island populations. According to the "express train" model (Diamond 1988), remote Oceanic populations originated from a wave of migration from Southeast Asia during the past 4,000 years. According to this model, there would have been limited gene flow between Melanesian populations and the ancestors of remote Oceanic populations, and the closest genetic affinity of remote Oceanic Islanders would be expected to be with Southeast Asian populations. This model is supported by a number of mtDNA analyses that identify predominantly Asian mtDNA types in Polynesians and Micronesians, with limited amounts of Melanesian mtDNA types (Hertzberg et al. 1989; Lum et al. 1994, 1998; Redd et al. 1995; Sykes et al. 1995; Lum and Cann 1998). According to the "slow-train" model of origin of remote Oceanic populations, there have been complex patterns of migration and gene flow among Southeast Asian, Melanesian, and Oceanic populations (Lum and Cann 1998; Terrell 1988). This model is supported by data on nuclear and Y-chromosome loci (Serjeantson 1985; O'Shaughnessy et al. 1990; Martinson 1996; Roberts-Thomson et al. 1996; Lum et al. 1998). Lum et al. (1998) have suggested that there may have been male-specific gene flow from coastal Papua New Guinea to Micronesia, which would explain the discrepancy between the mtDNA and nuclear-genetic data. The PLAT $(CA)_n$-1/*Alu* haplotype data also support the slow-train model of origin of remote Oceanic populations. Because mutation from a small-sized repeat to a large-sized repeat probably has been a rare event, it seems unlikely that it would have occurred independently in both Micronesian and Papua New Guinean populations, and this, in turn, suggests that there has been gene flow between these regions. Additionally, the fact that the PCA coordinates of the Ami are identical with those of the Papua New Guineans is intriguing in

light of other studies suggesting that Austronesian-speaking indigenous Taiwanese populations (such as the Ami and Atayal) played a significant role in the Austronesian expansion into Papua New Guinean and other regions of Oceania (Bellwood 1978, 1985, 1995; Melton et al. 1995, 1998; Sykes et al. 1995; Redd and Stoneking 1999). Further studies of autosomal, Y-chromosome, and mtDNA markers will help us to distinguish among these models of Oceanic Island origins.

Acknowledgments

We thank Nuria Navarro-Coy for technical assistance. We thank Gabriella Spedini at the University of Rome "La Sapienza" for the Bamileke and Ewondo samples; Sylvana Santachiarra-Benerecetti for the Wolof samples; and Batsheva Bonné-Tamir for samples from the Ethiopian Jewish, Yemenite Jewish, and Druze populations. We also thank the many other individuals, particularly the members of the indigenous populations studied, who have contributed DNA samples. This project was funded by a National Science Foundation (NSF) Sloan postdoctoral fellowship, NSF grant BCS-9905396, and a Burroughs Wellcome Fund Career Award (all to S.A.T.); National Institutes of Health grant GM57672 (to K.K.K.); NSF grant SBR-9632509 (to J.R.K.); a grant from the Sloan Foundation (to K.K.K. and J.R.K.); funding from the South African Medical Research Council and the University of the Witwatersrand (to T.J.); NSF grant 9423118 (to M.S.); and National Science Council, Taiwan, ROC grant NSC 88-2314-B-106-081 (to R.-B. L.). S.A.T. would like to acknowledge the hospitality of the Jenkins Laboratory, the South African Institute of Medical Research, and the University of the Witwatersrand, who hosted her as a Visiting Research Scientist during June–September 1997, during which time the data on the southern-African populations were collected. We thank Jeffrey Long and an anonymous reviewer for comments on the manuscript.

Electronic-Database Information

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

Kidd Lab Home Page, http://info.med.yale.edu/genetics/kkidd (for haplotype frequencies, available via ALFRED)
Microsat (software), http://lotka.stanford.edu/microsat
NCBI GenBank Overview, http://www.ncbi.nlm.nih.gov/Web/Genbank (for human PLAT reference sequence [accession number K03021])
Tishkoff lab Web page, http://www.life.umd.edu/biology/faculty/tishkoff/index.html (for haplotype frequencies)

References

Armour JA, Anttinen T, May CA, Vega EE, Sajantila A, Kidd J R, Kidd KK, Bertranpetit J, Pääbo S, Jeffreys AJ (1996) Minisatellite diversity supports a recent African origin for modern humans. Nat Genet 13:154–160

Barr CL, Kidd KK (1993) Population frequencies of the A1 allele at the dopamine D2 receptor locus. Biol Psychiatry 34(4): 204–209

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, 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

Bellwood PS (1978) Man's conquest of the Pacific: the prehistory of Southeast Asia and Oceania. Oxford University Press, New York

——— (1985) Prehistory of the Indo-Malayan archipelago. Academic Press, London

——— (1989) The colonization of the Pacific: some current hypotheses. In: Hill AVS, Serjeantson SW (eds) The colonization of the Pacific: a genetic trail. Oxford University Press, New York, pp 1–59

——— (1995) The Austronesian prehistory in Southeast Asia: homeland, expansion, and transformation. In: Bellwood P, Fox JJ, Tryon D (eds) The Austronesians: historical and comparative perspectives. Department of Anthropology, Comparative Austronesian Project, Research School of Pacific and Asian Studies, Australian National University, Canberra, pp 96–111

Bertranpetit J, Calafell F (1996) Genetic and geographical variability in cystic fibrosis: evolutionary considerations. Ciba Found Symp 197:97–114

Birdsell JB (1967) Preliminary data on the trihybrid origin of the Australian Aborigines. Archeol Phys Anthropol Oceania 2:100–155

——— (1993) Microevolutionary patterns in Aboriginal Australia: a gradient analysis of clines. Oxford University Press, New York

Blanquer-Maumont A, Crouau-Roy B (1995) Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. J Mol Evol 41(4): 492–497

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 STRPs. Nature 368: 455–457

Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human STRPs: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415

Calafell F, Shuster A, Speed WC, Kidd JR, Black FL, Kidd KK (1999) Genealogy reconstruction from short tandem repeat genotypes in an Amazonian population. Am J Phys Anthropol 108(2): 137–146

Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK (1998) Short tandem repeat polymorphism evolution in humans. Eur J Hum Genet 6:38–49

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

Castiglione CM, Deinard AS, Speed WC, Sirugo G, Rosenbaum HC, Zhang Y, Grandy DK, Grigorenko EL, Bonne-Tamir B, Pakstis AJ, Kidd JR, Kidd KK (1995) Evolution of haplotypes at the DRD2 locus. Am J Hum Genet 57:1445

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

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide STRP loci. Proc Natl Acad Sci USA 94:1041–1046

Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengård J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF (1998) Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am J Hum Genet 63:595–612

Clark DJ (1959) The prehistory of southern Africa. Penguin Books, Middlesex, England

Degen SJ, Rajput B, Reich E (1986) The human tissue plasminogen activator gene. J Biol Chem 261:6972–6985

Deinard A, Kidd K (1999) Evolution of a HOXB6 intergenic region within the great apes and humans. J Hum Evol 36: 687–703

Destro-Bisol G, Presciuttini S, d'Aloja E, Dobosz M, Spedini G, Pascali VL (1994) Genetic variation at the ApoB 3′ HVR, D2S44, and D7S21 loci in the Ewondo ethnic group of Cameroon. Am J Hum Genet 55:168–174

Diamond J (1988) Express train to Polynesia. Nature 336: 307–308

Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK, Barch DH (1998) Heterogeneity of STRP mutations within and between loci, and implications for human demographic histories. Genetics 148:1269–1284

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Res 6(3): 211–217

Goldman D, Brown GL, Albaugh B, Robin R, Goodson S, Trunzo M, Akhtar L, Lucas-Derse L, Long J, Linnoila M, Dean M (1993) DRD2 dopamine receptor genotype, linkage disequilibrium, and alcoholism in American Indians and other populations. Alcohol Clin Exp Res 17:199–204

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on STRPs and the origin of modern humans. Proc Natl Acad Sci USA 92: 6723–6727

Greenberg JH (1963) The languages of Africa. Indiana University Press, Bloomington

Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. J Mol Evol 44: 336–340

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Guthrie M (1962) Some developments in the prehistory of the Bantu origins. J Afr Hist 3:273–282

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378:376–378

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15: 427–441

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of human Y chromosome variation. Genetics 145:787–805

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Harpending H, Batzer M, Gurven M, Jorde L, Rogers A, Sherry S (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Harpending HC, Relethford J, Sherry ST (1996) Methods and models for understanding human diversity. In: Boyce AJ, Mascie-Taylor CGN (eds) Molecular biology and human diversity. Cambridge University Press, Cambridge, pp 288–299

Harpending HC, Sherry ST, Rogers AR, Stoneking M (1993) Genetic structure of ancient human populations. Curr Anthropol 34:483–496

Harris EE, Hey J (1999) X chromosome evidence for ancient human histories. Proc Natl Acad Sci USA 96:3320–3324

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 2:204–211

Hawley ME, Kidd KK (1995) HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. J Hered 86:409–411

Hertzberg M, Mickleson KN, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510

Howells WW (1989) Skull shapes and the map: craniometric analyses in the dispersion of modern Homo. Harvard University Press, Cambridge, MA

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Biol Evol 9:138–151

Johnson BJ, Miller GH, Fogel ML, Magee JW, Gagan MK, Chivas AR (1999) 65,000 years of vegetation change in central Australia and the Australian summer monsoon. Science 284:1150–1152

Jones R (1995) Tasmanian archaeology: establishing the sequences. Annu Rev Anthropol 24:423–446

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) STRP diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, et al (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet 64:817–831

Kidd JR, Black FL, Weiss KM, Balazs I, Kidd KK (1991) Studies of three Amerindian populations using nuclear DNA polymorphisms. Hum Biol 63:775–794

Kidd JR, Pakstis AJ, Zhao H, Lu RB, Okonofua FE, Odunsi A, Grigorenko E, Tamir BB, 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, Kidd JR (1996) A nuclear perspective on human evolution. In: Boyce AJ, Mascie-Taylor CGN (eds) Molecular biology and human diversity. Cambridge University Press, Cambridge, pp 242–264

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

Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB (1998) Signatures of population expansion in STRP repeat data. Genetics 148:1921–1930

Kofler A, Braun A, Jenkins T, Serjeantson SW, Cleve H (1995) Characterization of mutants of the vitamin-D-binding protein/group specific component: GC aborigine (1A1) from Australian aborigines and South African blacks, and 2A9 from south Germany. Vox Sang 68:50–54

Kolman CJ, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321–1334

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Lahr MM, Foley R (1995) Multiple dispersals and modern human origins. Evol Anthropol 3:48–50

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49: 49–67

——— (1995) The detection of linkage disequilibrium in molecular sequence data. Genetics 140:377–388

Lichter JB, Barr CL, Kennedy JL, Van Tol HH, Kidd KK, Livak KJ (1993) A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. Hum Mol Genet 2:767–773

Lum JK, Cann RL (1998) mtDNA and language support a common origin of Micronesians and Polynesians in Island Southeast Asia. Am J Phys Anthropol 105:109–119

Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. Am J Hum Genet 63:613–624

Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Martinson JJ (1996) Molecular perspectives on the colonization of the Pacific. In: Boyce AJ, Macie-Taylor CGN (eds) Molecular biology and human diversity. Cambridge University Press, Cambridge, pp 171–195

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

Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, Martinson J, Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet 57:403–414

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

Miller GH, Magee JW, Johnson BJ, Fogel ML, Spooner NA, McCulloch MT, Ayliffe LK (1999) Pleistocene extinction of *Genyornis newtoni:* human impact on Australian megafauna. Science 283:205–208

Nei M (1995) Genetic support for the out-of-Africa theory of human evolution. Proc Natl Acad Sci USA 92:6720–6722

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

Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, et al (1998) Polymorphic Alu insertions and the Asian origin of Native American populations. Hum Biol 70:23–39

Nurse GT, Weiner JS, Jenkins T (1985) The peoples of southern Africa and their affinities. Claredon Press, Oxford

O'Connell JF, Allen J (1998) When did humans first arrive in greater Australia and why is it important to know? Evol Anthropol 6:132–146

O'Shaughnessy DF, Hill AV, Bowden DK, Weatherall DJ, Clegg JB (1990) Globin genes in Micronesia: origins and affinities of Pacific Island peoples. Am J Hum Genet 46:144–155

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using STRP data from bear (*Ursidae*) populations. Genetics 147:1943–1957

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet 62:420–434

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

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

Peterson RJ, Goldman D, Long JC (1999) Effects of worldwide population subdivision on ALDH2 linkage disequilibrium. Genome Res 9:844–852

Pritchard JK, Seilestad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16: 1791–1798

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441

Rannalla B, Slatkin M (1998) Likelihood analysis of disequilibrium mapping, and related problems. Am J Hum Genet 62:459–473

Redd AJ, Stoneking M (1999) Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet 65:808–828

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M (1995) Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci USA 95:8119–8123

Relethford JH (1995) Genetics and modern human origins. Evol Anthropol 4:53–63

Relethford JH, Harpending HC (1994) Craniometric variation, genetic theory, and modern human origins. Am J Phys Anthropol 95:249–270

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

Ridker PM, Baker MT, Hennekens CH, Stampfer MJ, Vaughan DE (1997) Alu-repeat polymorphism in the gene coding for tissue-type plasminogen activator (t-PA) and risks of myocardial infarction among middle-aged men. Arterioscler Thromb Vasc Biol 17:1687–1690

Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, Breakefield X, Bressman S (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat Genet 9:152–159

Ritte U, Neufeld E, Broit M, Shavit D, Motro U (1993) The differences among Jewish communities—maternal and paternal contributions. J Mol Evol 37:435–440

Roberts RG, Jones R (1994) Luminescence dating of sediments: new light on the human colonization of Australia. Aust Aboriginal Stud 2:2–17

Roberts-Thomson JM, Martinson JJ, Norwich JT, Harding RM, Clegg JB, Boettcher B (1996) An ancient common origin of Aboriginal Australians and New Guinean highlanders is supported by β-globin haplotype analysis. Am J Hum Genet 58:1017–1024

Rogers AR, Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569

Rogers AR, Jorde LB (1995) Genetic evidence on modern human origins. Hum Biol 67:1–36

Rosenbaum HC, Deinard A (1998) Caution before claim: an overview of microsatellite analysis. In: Desalle R, Schierwater B (eds) Molecular approaches to ecology and evolution. Birkhauser, Basel, pp 87–106

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142: 1357–1362

Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. Genetics 140:1413–1409

Sadler LA, Blanton SH, Daiger SP (1991) Dinucleotide repeat polymorphism at the human tissue plasminogen activator gene (PLAT). Nucleic Acids Res 19:6058

Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. Am J Phys Anthropol 108:1–39

Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A (1999) Combined use of biallelic and STRP Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet 65:829–846

Scozzari R, Torroni A, Semino O, Sirugo G, Brega A, Santachiara-Benerecetti AS (1988) Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. Am J Hum Genet 43:534–544

Serre JL, Simon-Bouy B, Mornet E, Jaume-Roig B, Balassopoulou A, Schwartz M, Taillandier A, Boue J, Boue A (1990) Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in populations genetics. Hum Genet 84:449–554

Serjeantson SW (1985) Migration and admixture in the Pacific: insights provided by human leucocyte antigens. In: Kirk R, Szathmary E (eds) Out of Asia: peopling the Americas and the Pacific. Journal of Pacific History, Canberra, pp 133–145

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

Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol 66: 761–775

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty RA (1995) Novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol Biol Evol 12: 914–920

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983–993

Shriver MD, Jin L, Ferrell RE, Deka R (1997) STRP data support an early population expansion in Africa. Genome Res 7:586–591

Sokal RR, Rohlf J (1995) Biometry, 3d ed. WH Freedman, New York

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608

Spedini G, Destro-Bisol G, Mondovi S, Kaptué L, Taglioli L, Paoli G (1999) The peopling of sub-Saharan Africa: the case study of Cameroon. Am J Phys Anthropol 110:143–162

Spurdle A, Jenkins T (1992) Y chromosome probe p49a detects complex *Pvu*II haplotypes and many new *Taq*I haplotypes in southern African populations. Am J Hum Genet 50:107–125

Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM,

Wallace DC (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. Am J Hum Genet 63:1473–1491

Steeds R, Adams M, Smith P, Channer K, Samani NJ (1998) Distribution of tissue plasminogen activator insertion/deletion polymorphism in myocardial infarction and control subjects. Thromb Haemost 79:980–984

Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, et al (1998) Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 62:1507–1515

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, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Stringer CB (1993) New issues in modern human origins. In: Rasmussen T (ed) The origin and evolution of humans and humanness. Jones & Bartlett, Boston, pp 75–94

Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. Science 239:1263–1268

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mi tochondrial lineage analysis. Am J Hum Genet 57:1463–1475

Tautz D, Schlötterer C (1994) Simple sequences. Curr Opin Genet Dev 4:832–837

Terrell J (1988) History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. Antiquity 62:642–657

Thomas W, Drayna D (1992) A polymorphic dinucleotide repeat in intron 1 of the human tissue plasminogen activator gene. Hum Mol Genet 1:138

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Cheung K, Kidd JR, Bonne-Tamir B, Santachiara-Benerecetti A-S, Moral P, Watson E, Krings M, Pääbo S, Risch N, Jenkins T, Kidd KK (1996*a*) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380–1387

Tishkoff SA, Goldman A, Calafell F, Speed WC, Dienard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK (1998*a*) 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

Tishkoff SA, Kidd KK, Clark AG (1998*b*) Inferences of modern human origins from variation in CD4 haplotypes. In: Uyenoyama MK, von Haesler A (eds) Proceedings of the trinational workshop on molecular evolution. Duke University Press, Raleigh-Durham, NC, pp 181–198

Tishkoff SA, Pakstis AJ, Ruano G, Kidd KK (2000) The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. Am J Hum Genet 67:518–522

Tishkoff SA, Ruano G, Kidd JR, Kidd KK (1996*b*) Distribution and frequency of a polymorphic Alu insertion at the PLAT locus in humans. Hum Genet 97:759–764

Valdes AM, Slatkin M, Freimer MB (1993) Allelic frequencies at STRP loci: the stepwise mutation model revisited. Genetics 133:737–749

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354

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

Underhill PA, Jin L, Lin AA, Medhi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Weir BS, Cockerham CC (1984) Estimating F statistics for the analysis of population structure. Evolution 38:1358–1370

Wilson AC, Sarich VMA (1969) Molecular time scale for human evolution. Proc Natl Acad Sci USA 63:1088–1093

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159

Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, et al (1997) Nuclear DNA diversity in worldwide distributed human populations. Gene 205:161–171

Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D (1998) Genetic structure of the ancestral population of modern humans. J Mol Evol 47:146–155