Reconstruction of Human Evolutionary Tree Using Polymorphic Autosomal Microsatellites

Qasim Ayub,[1] Atika Mansoor,[1] Muhammad Ismail,[1] Shagufta Khaliq,[1] Aisha Mohyuddin,[1] Abdul Hameed,[1] Kehkashan Mazhar,[1] Sadia Rehman,[1] Saima Siddiqi,[1] Myrto Papaioannou,[2] Alberto Piazza,[3] Luigi Luca Cavalli-Sforza,[4] and S. Qasim Mehdi[1]*

[1]*Biomedical and Genetic Engineering Division, Dr. A.Q. Khan Research Laboratories, Islamabad 44000, Pakistan*
[2]*Unit of Prenatal Diagnosis, Center for Thalassemia, Laiko General Hospital, Athens 11527, Greece*
[3]*Department of Genetics, Biology and Biochemistry, University of Torino, 10126 Torino, Italy*
[4]*Department of Genetics, Stanford University, Stanford, California 94305*

KEY WORDS autosomal short tandem repeats; principal components analysis; human origins

ABSTRACT Allelic frequencies of 182 tri- and tetra-autosomal microsatellites were used to examine phylogenetic relationships among 19 extant human populations. In particular, because the languages of the Basques and Hunza Burusho have been suggested to have an ancient relationship, this study sought to explore the genetic relationship between these two major language isolate populations and to compare them with other human populations. The work presented here shows that the microsatellite allelic diversity and the number of unique alleles were highest in sub-Saharan Africans. Neighbor-joining trees based on genetic distances and principal component analyses separated populations from different continents, and are consistent with an African origin for modern humans. For the first time, with biparentally transmitted markers, the microsatellite tree also shows that the San are the first branch of the human tree before the branch leading to all other Africans. In contrast to an earlier study, these results provided no evidence of a genetic relationship among the two language isolate groups. Genetic relationships, as ascertained by these microsatellites, are dictated primarily by geographic proximity rather than by remote linguistic origin, Mantel test, $R_0 = 0.484$, $g = 3.802$ (critical g value = 1.645; $P = 0.05$). Am J Phys Anthropol 122:259–268, 2003.

Genetic variation at microsatellite loci has been used extensively for the study of human variation, linkage analysis, and DNA fingerprinting (Jeffreys et al., 1985; Tautz, 1989; Weber and May, 1989). For the last 80 years, in association with linguistic, anthropological, and genetic-marker evidence, advances in human genetics have shed light on the process of human evolution and migration (Cavalli-Sforza et al., 1994). DNA-based markers have gained acceptance as the markers of choice for elucidating questions of human evolution and migration. An ideal DNA marker should be both highly polymorphic and selectively neutral. Recent studies focused on microsatellite markers and single-nucleotide polymorphisms (SNPs) for the study of human evolution (Bowcock et al., 1994; Deka et al., 1991; Di Rienzo et al., 1994; Goldstein and Pollock, 1997; Jorde et al., 1997). Microsatellite markers are sequences of 2–6 nucleotide repeats. They have proven both informative and accurate in reconstructing trees that demonstrate evolutionary relationships between human individuals (Bowcock et al., 1994; Deka et al., 1995). In addition to autosomal microsatellite data, Y-chromosome and mitochondrial DNA studies shed light on the patrilinear and maternal relationships, respectively, among modern humans (Cavalli-Sforza, 1998; Hammer et al., 1997; Quintana-Murci et al., 1999).

A comprehensive worldwide analysis of human genes and markers suggests that both linguistics and genetics can provide complementary information with which to synthesize a picture of the human past (Cavalli-Sforza et al., 1994). This finding does not hold in the case of populations that speak a language which cannot be classified into one of the accepted taxa. In language isolate groups, the study of genetic markers may assist in ascertaining the origin of such populations. Two such language isolate groups are the Basque in Europe and the Hunza Burusho in the northern areas of Pakistan (Grimes, 1992; see also Ethnologue, http://www.ethnologue.

Grant sponsor: Government of Pakistan.

*Correspondence to: S. Qasim Mehdi, Biomedical and Genetic Engineering Division, Dr. A.Q. Khan Research Laboratories, G.P.O. Box 2891, Islamabad 44000, Pakistan. E-mail: sqmehdi@comsats.net.pk

Received 8 August 2002; accepted 26 November 2002.

DOI 10.1002/ajpa.10234

com/). According to Starostin, these languages belong to the oldest Eurasian (Dene-Caucasian) family (Greenberg, 2000). The Burusho reside in the harsh environs of the Karakoram Mountains in the Hunza Valley of northern Pakistan, and the Basque along the French-Spanish border.

The present study aimed to determine the genetic relationship between 19 modern human populations. These include populations that were studied earlier with a subset of 29 different dinucletide repeats by Bowcock et al. (1994) and Jin et al. (2000). The additional populations are Basque, Greek, and three diverse populations from the Indo-Pakistan subcontinent, i.e., the Hunza Burusho, Sindhis, and Brahuis. The Sindhis speak an Indo-European language, and the Brahuis a Dravidian language in an area surrounded by Indo-European speakers (Grimes, 1992).

MATERIALS AND METHODS

DNA samples

In addition to the individuals surveyed earlier by Bowcock et al. (1994) and Jin et al. (2000), this study includes Basques and Greeks from Europe and three Pakistani ethnic groups (Brahui, Hunza Burusho, and Sindhi). Informed consent was obtained from all human subjects. The populations included four from Africa, i.e., the Zaire Pygmies, Central African Republic (CAR) Pygmies, Lisongo, and San. Four populations were from Europe, and included North European, North Italian, Greek, and Basque. The Chinese, Japanese, and Cambodian represented East Asian populations. The three Oceania populations included Melanesian, New Guinean, and Australian Aborigine. South American populations were the Mayan Indians from the Yucatan peninsula and the Brazilian Indian tribes.

Two hundred and thirteen tri- and tetranucleotide microsatellite loci were typed in the 19 modern human populations (168 individuals) listed above and in a chimpanzee population (10 samples). Chimpanzee samples were African-born and presumably unrelated. The microsatellite loci were approximately 7 cM apart and were presumed to have no detectable linkage disequilibrium (Adamson et al., 1995; Gastier et al., 1995). These markers were located on human chromosomes 1–6, 9–11, 14, 17, and 18.

Genotyping by multiplex PCR

Each sample was PCR-amplified in a multiplex reaction consisting of 5–7 primer pairs (Mappairs Multiplex kit, Research Genetics), which were labelled either with TET, HEX, or FAM. The multiplex PCR assay was performed in a 5-μl.final volume. Briefly, the reaction consisted of the following: 1 × PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.1 mg/ml gelatin, and 3 mM $MgCl_2$), 200 μM dNTPs, 0.25 U of AmpliTaq (Perkin Elmer), 0.083 μg TaqStart antibody (Clontech), and 0.05 μM of each primer. AmpliTaq DNA polymerase was preincubated with the TaqStart antibody for 5–7 min, after which this mixture was added to the master mix and dispensed into tubes containing 50 ng genomic DNA. PCR was performed by a "touchdown" protocol, as described previously (Jin et al., 1997). After amplification, 0.3 μl of each sample was mixed with 0.3 μl of either TAMRA 350 or TAMRA 500 internal lane standard and 2.4 μl of a loading dye consisting of formamide and Dextran blue. The samples were then denatured at 98°C for 2 min, and 1 μl of this mix was loaded on a 12-cm-long, 4% denaturing polyacrylamide gel. The gel was electrophoresed for 1.5 hr on an ABI 377 DNA sequencer, according to the manufacturer's instructions. Data were collected using Collection Software version 1.1, and fragment sizes were estimated based on known internal lane size standards, using the software GeneScan (version 2.1). The genotypes were called using Genotyper (version 2.0) software.

Data analyses

Allelic frequencies were used to construct phylogenetic trees by using Dispan software (http://www.bio.psu.edu/People/Faculty/Nei/Lab/software.html) containing the programs GNKDST for calculating genetic distances and Treeview for drawing phylogenetic trees by neighbor-joining and UPGMA methods (Nei, 1973; Nei et al., 1983; Sokal and Michener, 1958). Principal components analysis was carried out by using code R, a freely available, open source version of the S-Plus statistical package (Venables and Ripley, 1999). For graphic representation, the first and second principal components were plotted using Sigma Plot (version 1) software for Windows. The Mantel test (Mantel, 1967) was used to test correlation of genetic distances with geographic distance, as demonstrated by Eller (1999).

RESULTS

Of the 213 loci amplified, genotypes from 182 loci (28 tri- and 154 tetranucleotide repeats) could be scored unambiguously. Twenty-four loci (6 tri- and 18 tetranucleotide repeats) failed to amplify in the chimpanzees. Two successive attempts were made to amplify these microsatellite markers, after which they were excluded. In addition, 7 loci (1 tri- and 6 tetranucleotide repeats) did not amplify in most of the human populations from unfavorable multiplex PCR cycling conditions. These loci could be amplified in singleplex PCR reactions, but the amount of DNA available for some human populations was not enough to complete the analysis. Subsequent analysis was carried out on loci for which genotypes could be scored confidently in all 19 human populations and chimpanzees.

In total, 2,069 alleles were found in the human and chimpanzee populations. For most loci, the length of the repeat unit was longer in humans as compared to the chimpanzees. The latter were relatively less heterozygous (average heterozygosity ±

SE = 0.6378 ± 0.018) for all loci examined. Low heterozygosity values were also observed for dinucleotide repeat loci in chimpanzees (Deka et al., 1995), and it was suggested that microsatellites evolve more rapidly in humans (Rubinsztein et al., 1995). In total, 12 monomorphic loci were found in the chimpanzee samples. In the human populations, the Hunza Burusho, Cambodian, and Melanesian had one monomorphic locus each, while the New Guineans had two. Anomalous alleles, due to simple insertions or deletions, were not observed within the human samples studied. In humans, the average number of alleles was 10 (Table 1), with a range of between 4 (D6S1006) and 22 (D9S921).

Marker heterozygosities and G_{ST}

The characteristics of the 182 microsatellite loci examined in this study, their G_{ST}, as a measure of allelic variability and heterozygosity (H_T) for each locus, are given in Table 1. The modal size of the repeat unit was 9 for both the trinucleotide and tetranucleotide repeats, with a range of between 4–16 and 6–22, respectively. The heterozygosity values ranged between 0.63–0.92 (D6S1006 and D9S302, respectively). There was no significant difference in the average heterozygosities of the 28 tri- and 154 tetranucleotide repeats (independent *t*-test, $P = 0.5443$).

The average heterozygosity value for each population group is given in Table 2. Among the humans, the African, European, and Pakistani populations were more heterozygous (0.72–0.74), followed closely by the East Asian (0.70–0.72). The Oceania and South American populations were the least heterozygous. Among all human populations, the Brazilian Indians showed the lowest heterozygosity (0.63). The highest number of unique alleles was found in Africans, with the Zaire Pygmies having a total of 40 unique alleles. Despite their low heterozygosity, a total of 27 unique alleles was found in the Brazilian Indian population. The lowest number of unique alleles was found in the North European (1%) and Greek (0.7%) samples.

Phylogenetic analysis

Two different methods were used for genetic data analysis: 1) genetic distance analysis between two populations, using the proportion of shared alleles (D_{AS}) averaged over all loci (Bowcock et al., 1994), and 2) principal components analysis of the allelic frequency data (Cavalli-Sforza et al., 1994). The matrix of genetic distances between populations, with their standard error (calculated by bootstrapping), is presented in Table 3. Using D_{AS} values, neighbor-joining trees were generated, including (Fig. 1) and excluding (Fig. 2) the chimpanzee samples. Neighbor-joining is a heuristic method that employs genetic distances to produce a minimum evolution tree (Saitou and Nei, 1987). Neighbors are pairs of operational taxonomic units (OTUs) connected through a single interior node, and tree validation is carried out by bootstrap resampling.

The neighbor-joining trees based on the allelic frequencies resolved the individual populations into their regional and continental groups (Figs. 1, 2). The human populations were grouped into six distinct geographical clusters: Africa (Lisongo and Pygmies), Europe (Basque, Greek, Northern European, and Italian), Pakistan (Brahui, Hunza Burusho, and Sindhi), East Asia (Cambodian, Chinese, and Japanese), Oceania (Australian Aborigine, Melanesian, and New Guinean), and South America (Brazilian and Mayan Indian). Despite the range of bootstrap values, except for the San, no individual population is separated from its expected geographic cluster. The first split in the phylogenetic tree separates Africans from non-Africans (Fig. 2), with strong statistical support (100% of 10,000 bootstrap resamplings). The separation is also seen in the UPGMA tree, based on standard genetic distances (data not shown) and principal components analysis (Fig. 3). A maximum likelihood tree, based on the allelic frequencies of 182 microsatellite repeats, is also topologically identical (data not shown).

The major branch immediately after the separation of the San in the phylogenetic trees (Figs. 1, 2) separates the sub-Saharan Africans from the rest of the world. Subsequent branching of the tree separates East Asian, Oceania, and South American populations from the European and Pakistani populations. There is a genetic affinity among populations on either side of the Pacific divide, i.e., the Japanese, Chinese, Cambodians, and the Mayan and Brazilian Indians, as well as the Oceanic populations. This connection straddles the Australasian and American continents. Subsequent branching of the tree separates the South American Indian tribes from East Asia and Oceania (99% of 10,000 bootstrap resamplings). Further branching distinguishes each individual population into groups that relate to their geographic location, although with less statistical support than the earlier separations. In some cases, such as that of the Brazilian and Mayan Indians, Basque and Greek, and Chinese and Japanese, the bootstrap values are high, (100, 99, and 93, respectively, in Fig. 2). However, in other comparisons such as that between the CAR Pygmy and Lisongo, the bootstrap value is lower (Fig. 2). These two populations have had some genetic exchange, as a fraction of the Lisongo men have married Pygmy women.

Principal components analysis

A graph of the eigenvalues of the principal components analysis, based on the allelic frequencies of 182 microsatellite repeats, with and without the chimpanzee samples, demonstrates that inclusion of the chimpanzee samples mainly highlights the differences between chimpanzees and humans (data not shown). As expected, the percentage of variances explained by the first six principal components de-

TABLE 1. Marker designation, type of repeat unit, G_{ST}, and heterozygosity (H_T) values for 182 microsatellite markers used in this study[1]

Marker	Type[2]		G_{ST}	H_T	Marker	Type[2]		G_{ST}	H_T
D1S1612	GGAA	(9)	0.2428	0.8262	D4S2394	ATA	(12)	0.2027	0.8263
D1S552	GGAT	(8)	0.1471	0.6831	D4S1644	GATA	(9)	0.0910	0.7493
D1S1622	ATA	(9)	0.2425	0.7916	D4S1625	GATA	(8)	0.0784	0.7575
D1S2130	GATA	(8)	0.1518	0.7246	D4S1629	GATA	(6)	0.1898	0.7195
D1S2134	GATA	(10)	0.1708	0.8005	D4S2368	GATA	(8)	0.0948	0.7409
D1S1669	GATA	(15)	0.1556	0.8442	D4S2431	GGAA	(9)	0.1214	0.7880
D1S1665	GATA	(12)	0.1778	0.7211	D4S2417	GATA	(7)	0.1499	0.6537
D1S1728	GATA	(8)	0.1291	0.6746	D4S1652	GATA	(6)	0.1513	0.6800
D1S551	GATA	(8)	0.1552	0.7386	D5S807	GATA	(12)	0.1321	0.7675
D1S1588	ATA	(9)	0.1187	0.6705	D5S817	GATA	(7)	0.1246	0.7085
D1S1675	GGAA	(7)	0.1686	0.7067	D5S1470	GATA	(8)	0.0982	0.8243
D1S1595	GATA	(9)	0.1677	0.6774	D5S2494	GATA	(10)	0.1536	0.7277
D1S1679	GGAA	(9)	0.2298	0.8365	D5S2500	GATA	(9)	0.1526	0.8055
D1S1677	GGAA	(7)	0.2215	0.7461	D5S1501	GATA	(10)	0.2023	0.6558
D1S1589	ATA	(11)	0.2051	0.7991	D5S1719	GATA	(10)	0.1269	0.8010
D1S518	GATA	(10)	0.1240	0.8546	D5S1462	GATA	(10)	0.1411	0.8010
D1S1660	GATA	(9)	0.1355	0.7999	D5S1453	ATA	(14)	0.2232	0.7756
D1S1678	GGAA	(8)	0.1631	0.7586	D5S2501	GATA	(8)	0.1003	0.7630
D1S1663	GATA	(9)	0.1107	0.7076	D5S1505	GATA	(13)	0.1173	0.8452
D1S2141	GATA	(10)	0.1253	0.8278	D5S816	GATA	(9)	0.1746	0.8318
D1S549	GATA	(9)	0.1532	0.7899	D5S1480	ATA	(11)	0.1712	0.8280
D1S1656	GATA	(11)	0.1026	0.8671	D5S820	GATA	(9)	0.1308	0.7833
D1S3462	ATA	(9)	0.3050	0.8103	D5S1471	GATA	(11)	0.1116	0.8006
D1S547	GATA	(9)	0.1095	0.7778	D5S1456	GATA	(11)	0.1085	0.7939
D2S1780	GATA	(10)	0.1323	0.7563	D6S1006	ATC	(4)	0.3074	0.6227
D2S423	GAAT	(9)	0.1734	0.7404	D6S1281	GATA	(11)	0.0978	0.7640
D2S1400	GGAA	(12)	0.2282	0.7318	D6S1019	GTAT	(11)	0.1385	0.7842
D2S405	GATA	(8)	0.2010	0.7242	D6S1017	GGAT	(9)	0.1666	0.7623
D2S1788	GATA	(14)	0.1209	0.8973	D6S1280	GATA	(11)	0.1127	0.7645
D2S1356	ATA	(10)	0.1805	0.8475	D6S1960	GATA	(7)	0.1317	0.7467
D2S441	GATA	(6)	0.2233	0.7828	D6S1053	GATA	(9)	0.0819	0.8032
D2S1394	GATA	(7)	0.1465	0.7432	D6S1031	ATA	(13)	0.2134	0.8353
D2S1777	GATA	(11)	0.1815	0.6061	D6S1270	GATA	(10)	0.1040	0.7065
D2S1790	GATA	(12)	0.1199	0.7913	D6S1056	GATA	(10)	0.1333	0.8342
D2S436	GATA	(9)	0.1076	0.6977	D6S1021	ATA	(9)	0.1820	0.7147
D2S1328	GATA	(10)	0.1592	0.7964	D6S474	GATA	(10)	0.1284	0.7674
D2S442	GATA	(12)	0.0984	0.7749	D6S1009	GATA	(14)	0.1676	0.7781
D2S1326	GATA	(11)	0.1271	0.8578	D6S1003	ATA	(12)	0.2097	0.7599
D2S1399	GGAA	(15)	0.2383	0.7996	D6S1277	GATA	(9)	0.1245	0.7691
D2S1353	ATA	(10)	0.1046	0.8142	D6S503	GGAA	(6)	0.1440	0.7202
D2S1776	GATA	(7)	0.1351	0.7330	D6S1027	ATA	(9)	0.1970	0.8009
D2S2944	GATA	(9)	0.0885	0.7989	D9S2169	GATA	(7)	0.1326	0.7322
D2S434	GATA	(8)	0.0952	0.7656	D9S921	GATA	(22)	0.1798	0.8812
D2S1363	GATA	(10)	0.1789	0.8150	D9S925	GATA	(13)	0.1541	0.8625
D2S427	GATA	(7)	0.2633	0.7634	D9S1118	GATA	(16)	0.2342	0.8731
D3S2387	GATA	(15)	0.0906	0.8800	D9S922	GATA	(7)	0.1610	0.8138
D3S3050	GATA	(8)	0.0905	0.7500	D9S910	ATA	(8)	0.1966	0.7346
D3S2403	GGAA	(11)	0.1310	0.7216	D9S938	GGAA	(13)	0.2541	0.8175
D3S3038	GATA	(13)	0.1169	0.7858	D9S930	GATA	(9)	0.0799	0.8006
D3S2432	GATA	(12)	0.1311	0.8132	D9S302	GATA	(19)	0.0939	0.9176
D3S1768	GATA	(8)	0.1589	0.8028	D9S934	GATA	(9)	0.1096	0.7818
D3S1766	GATA	(10)	0.0817	0.7239	D9S915	ATA	(16)	0.1486	0.7081
D3S2406	GGAT	(18)	0.1181	0.9086	D10S1435	GATA	(11)	0.0601	0.7314
D3S2465	GGAA	(14)	0.1762	0.7534	D10S1412	ATA	(7)	0.1081	0.6599
D3S2459	GATA	(9)	0.1175	0.8093	D10S2325	GAAT	(13)	0.1082	0.8718
D3S3045	GATA	(10)	0.1870	0.8337	D10S674	GATA	(9)	0.1476	0.8173
D3S2460	GATA	(12)	0.1176	0.7389	D10S1423	GATA	(8)	0.0891	0.7754
D3S1764	GATA	(12)	0.2665	0.8014	D10S1426	GATA	(9)	0.0630	0.7168
D3S1744	GATA	(9)	0.1043	0.7970	D10S1220	ATA	(9)	0.1781	0.6084
D3S1763	GATA	(9)	0.1333	0.7898	D10S1225	ATA	(7)	0.1472	0.7521
D3S3053	GATA	(8)	0.1548	0.7618	D10S1432	GATA	(13)	0.1073	0.7345
D3S2436	GATA	(12)	0.0876	0.7476	D10S2327	GGAT	(9)	0.1653	0.7571
D3S2398	GATA	(8)	0.1202	0.7629	D10S677	GGAA	(14)	0.1288	0.8418
D3S2418	ATA	(11)	0.1913	0.7556	D10S1239	GATA	(11)	0.1447	0.7635
D4S2366	GATA	(8)	0.1454	0.8041	D10S1230	ATA	(10)	0.1460	0.7470
D4S2639	GATA	(10)	0.1342	0.8215	D10S1213	GGAA	(16)	0.1547	0.8070
D4S2397	ATA	(12)	0.1514	0.8051	D11S1984	GGAA	(12)	0.1244	0.8581
D4S2408	GATA	(7)	0.1945	0.7333	D11S2362	ATA	(11)	0.1618	0.8379
D4S1627	GATA	(8)	0.0706	0.8018	D11S1999	GATA	(9)	0.1568	0.8268
D4S3248	GATA	(13)	0.1752	0.8209	D11S1981	GATA	(13)	0.1354	0.8326
D4S2367	GATA	(10)	0.1707	0.7694	ATA34E08	ATA	(10)	0.1697	0.7193
D4S2361	ATA	(12)	0.2237	0.7418	D11S1392	GATA	(8)	0.1266	0.7591
D4S1647	GATA	(7)	0.1224	0.7974	D11S2371	GATA	(8)	0.0814	0.7116
D4S2623	GATA	(13)	0.1313	0.8564	D11S1986	GGAA	(19)	0.2986	0.8524

(*Continued*)

TABLE 1. (Continued)

Marker	Type[2]		G_{ST}	H_T
D11S1998	GATA	(7)	0.1069	0.7299
D11S4464	GATA	(9)	0.1052	0.7749
D11S2359	ATA	(16)	0.1799	0.8256
D14S742	GATA	(6)	0.1181	0.7303
D14S1280	GATA	(7)	0.1245	0.6986
D14S297	GATA	(9)	0.1112	0.6385
D14S306	GATA	(11)	0.1289	0.8054
D14S587	GGAA	(12)	0.2062	0.8113
D14S592	ATA	(15)	0.2025	0.8596
D14S588	GGAA	(9)	0.2413	0.7957
D14S606	GATA	(8)	0.1067	0.6980
D14S610	GATA	(8)	0.1213	0.6859
D14S617	GGAA	(11)	0.2041	0.7656
D14S749	GATA	(8)	0.1155	0.7039
D16S2622	GATA	(6)	0.1749	0.6136
D16S2619	GATA	(8)	0.1272	0.6919
D16S753	GGAA	(8)	0.2214	0.7949
D16S771	GGAA	(8)	0.1335	0.6826
D16S3253	GATA	(12)	0.2221	0.8045
D16S2624	GATA	(7)	0.1452	0.7400
D16S539	GATA	(7)	0.1242	0.7825
D17S1298	GAAT	(8)	0.2290	0.6354
D17S974	GATA	(9)	0.1432	0.7650
D17S1294	GGAA	(10)	0.3448	0.8307
D17S1299	GATA	(7)	0.0964	0.6893
D17S1290	GATA	(15)	0.1510	0.8640
D17S2059	GATA	(9)	0.1529	0.6547
D17S1301	GATA	(14)	0.1583	0.7572
D18S976	GATA	(9)	0.1286	0.8189
D18S877	GATA	(8)	0.1155	0.7545
D18S535	GATA	(11)	0.1592	0.8203
D18S851	GATA	(10)	0.1018	0.7400
D18S858	ATA	(9)	0.1453	0.7453
D18S1270	GATA	(11)	0.1288	0.8227

[1] Markers are arranged according to their location on chromosomes.
[2] Numbers of alleles found in human populations are given in parentheses.

creases gradually. In the analysis that included the chimpanzee sample (Fig. 3A), the majority of variation (29%) was accounted for between the first two components, and it gave results similar to the principal component analysis excluding the chimpanzees (Fig. 3B) that accounted for 21% of the variation. The principal components analysis excluding the chimpanzee samples was carried out to give a better perspective of the clustering structure among human populations (Fig. 3B–D). Both analyses separated the African populations from the non-African populations, and separated the San as an outlier of the African cluster (Fig. 3A,B). A plot of the first and third (Fig. 3C) and second and third (Fig. 3D) principal components separated the extant human populations into well-defined clusters, and corroborates the results obtained from the phylogenetic analysis. In these analyses, the Pakistani populations (Brahui, Hunza Burusho, and Sindhi) cluster together with the populations from Europe. The separation between the East Asians, South American, and Oceanic populations is not well-defined (Fig. 3C,D). This is the only example based on autosomal markers that shows a clear separation between the San and other Africans, which is even more pronounced in the plot of the second and third principal components (Fig. 3D).

Mantel test

To check if genetic distances were correlated with geographic distances, the Mantel test was applied (Mantel, 1967). The Mantel Nonparametric Test Calculator, version 2.0 for Windows, developed by Adam Liedloff (http://www.sci.qut.edu.au/nrs/mantel.htm), calculates the values of Z, the Mantel coefficient, and R_0, the correlation coefficient for two distance matrices, one for genetic distances and the other for geographical distances between populations. The geographic distances were measured by using the ENCARTA World Atlas (Microsoft), and are the best estimates of the shortest path from one location to another, avoiding major geographic barriers. The critical values from which to compare the standard normal variate (g) are provided for the most common level of significance ($P = 0.05$). The matrix of genetic distances was significantly correlated with geographic distances $Z = 85607$, $R_0 = 0.484$, and $g = 3.802$ (critical g value for 0.05 level of significance = 1.645).

DISCUSSION

This study sought to find the genetic relationship between two major language isolate populations, the Basques in Europe and the Hunza Burusho in Pakistan, and to compare them with other extant human populations. Population relationships were examined using genetic distance measures based on allele frequencies of 182 tri- and tetranucleotide microsatellite repeats. The use of such a large number of microsatellite markers overcomes the potential limitations of the small sample size (n = 168) that may not be especially representative. Despite these caveats, the results suggest that genetic relationships are dictated primarily by geographic proximity rather than linguistics (Mantel test, $R_0 = 0.484$, $g = 3.802$). The fact that there is gene flow across linguistic boundaries is supported by the observation that the language isolate populations (Basque and Hunza Burusho) are closer to their geographic neighbors. Similarly, the Dravidian language-speaking Brahui population of Pakistan is genetically similar to the other Pakistani populations analyzed (99% of 10,000 bootstrap resamplings in Fig. 2).

The first suggestion of a greater genetic similarity between Basques and the Hunza Burusho came from a preliminary survey by Sun (1997) of a smaller number of world populations using 35 dinucleotide microsatellites, almost exclusively of the CA type. The present analysis, with a larger number of populations and a much larger number of other microsatellites, provided no evidence of genetic affinity between the Basque and Hunza Burusho populations. The isolation of the Hunza Burusho and Basque, in the Karakoram and Pyrenees Mountains

TABLE 2. Total number of alleles, number of unique alleles, and average heterozygosities (± SE) for 19 worldwide human populations

Population	Total alleles	Unique alleles[1]	Average heterozygosities (± SE)
Africa			
San	313	4 (1.3%)	0.719780 ± 0.033382
CAR Pygmy	935	24 (2.6%)	0.745903 ± 0.008161
Lisongo	928	23 (2.5%)	0.743866 ± 0.008240
Zaire Pygmy	934	40 (4.3%)	0.743122 ± 0.008057
Europe			
Basque	921	13 (1.4%)	0.728433 ± 0.007909
Greek	944	7 (0.7%)	0.745549 ± 0.006445
North European	975	10 (1.0%)	0.741964 ± 0.007075
North Italian	904	21 (2.3%)	0.740306 ± 0.007929
Pakistan			
Brahui	766	9 (1.2%)	0.736307 ± 0.009642
Burusho	768	9 (1.2%)	0.735597 ± 0.010897
Sindhi	770	9 (1.2%)	0.743139 ± 0.008948
East Asia			
Cambodian	945	25 (2.6%)	0.723923 ± 0.009010
Chinese	903	11 (1.2%)	0.718882 ± 0.009373
Japanese	874	25 (2.9%)	0.704175 ± 0.010442
Oceania			
Australian	931	23 (2.5%)	0.712815 ± 0.009934
Melanesian	812	17 (2.1%)	0.656073 ± 0.012457
New Guinean	846	20 (2.4%)	0.671247 ± 0.012411
South America			
Brazilian Indian	831	27 (3.2%)	0.631065 ± 0.012375
Mayan Indian	861	13 (1.5%)	0.690266 ± 0.010406

[1] Values in parentheses indicate percentage of unique alleles in each population.

of northern Pakistan and Europe, respectively, may have helped to preserve their languages, Burushaski (Burusho) and Euskera (Basque), but the two groups are genetically more closely related to their geographic neighbors than to each other. However, it might be objected that the genetic relationship of Basques and Hunza may be ancient, judging from linguistic commonalities, and because of their high mutation rate, microsatellites are not best suited for the study of old relationships, but are generally more useful for the analysis of more recent population separations. One might hope instead to find ancient relationships with markers of low mutation rates, e.g., Y-chromosome single-nucleotide polymorphisms. In Underhill et al. (2000), 38 Hunza and 45 Basques were analyzed with 19 other world populations (total of individuals examined, 1,062) for 167 polymorphisms, defining 116 haplotypes. Four Hunza Burusho and 27 Basques shared two common haplotypes, indicating that some ancient genetic relationship may have existed between these two populations.

The present microsatellite data show that the four European and three Pakistani populations are phylogenetically close, forming two rather similar subclusters within a cluster (Figs. 1, 2). The average distance between the four European populations is 0.066, and that between the three Pakistani is 0.068, while the average of the 12 distances between Pakistani and European populations is 0.127. They may well have shared a common language superfamily, Dene-Caucasian, the dominant early Eurasian family prior to the expansion of the now prevalent Eurasiatic superfamily. The latter includes the Indo-European family and may have spread to Eurasia between 10–20 kya, replacing the languages spoken earlier almost everywhere (Greenberg, 2000). Only peoples living in fairly segregated areas and forming highly cohesive societies like the Basque and the Hunza Burusho may have resisted replacement of their language with those spoken by latecomers near their habitat. It is not surprising that this happened only in a few areas, in which language survival was favored by external and internal circumstances. In the long time that has elapsed since these events, the differences between the populations that retained their original languages and their neighbors, which have acquired new languages, have been inevitably greatly diluted by genetic exchange among neighbors.

The major clusters identified in this study largely agree with the currently accepted view of human migration and evolution as deduced from archeological and genetic evidence. Previous studies typed between 3–200 different microsatellite loci (primarily CA repeats) across worldwide populations, to generate sufficient resolving power to distinguish among various world populations (Barbujani et al., 1997; Bowcock et al., 1994; Cooper et al., 1999; Mountain and Cavalli-Sforza, 1997; Pérez-Lezaun et al., 1997; Reich and Goldstein, 1998). In an earlier study on these populations, Bowcock et al. (1994) succeeded primarily in distinguishing between larger population groups (e.g., Africans and non-Africans). The resolving power of microsatellites in differentiating between more closely related population groups was not conclusive. In this study, within the closely related Pakistani populations, as well as the European populations, the bootstrapping values decrease significantly in comparison to values that

TABLE 3. D_{AS} genetic distances for pairwise comparisons between 19 world populations and chimpanzees are given below diagonal, and standard errors for these genetic distances, based on 10,000 bootstraps, are given above diagonal

Population	PT	OM	P	F	PG	NE	ITA	GRK	BAS	BRU	SDH	BSK	CAM	CH	J	M	NG	AUS	BI	MI
Chimpanzee (PT)		0.015	0.016	0.018	0.017	0.018	0.017	0.017	0.017	0.017	0.017	0.016	0.017	0.018	0.018	0.018	0.017	0.017	0.017	0.016
San (OM)	0.830		0.015	0.016	0.015	0.015	0.015	0.016	0.016	0.017	0.016	0.015	0.015	0.016	0.016	0.018	0.017	0.017	0.018	0.017
CAR-Pygmy (P)	0.736	0.664		0.013	0.013	0.013	0.011	0.012	0.013	0.013	0.013	0.012	0.013	0.013	0.013	0.013	0.013	0.013	0.014	0.013
Lisongo (F)	0.706	0.648	0.412		0.013	0.013	0.013	0.013	0.013	0.014	0.014	0.014	0.013	0.014	0.015	0.014	0.014	0.014	0.015	0.014
Zaire Pygmy (PG)	0.726	0.668	0.454	0.442		0.013	0.013	0.013	0.014	0.014	0.014	0.013	0.014	0.013	0.014	0.015	0.014	0.014	0.015	0.014
North European (NE)	0.717	0.654	0.432	0.427	0.473		0.011	0.010	0.010	0.012	0.014	0.013	0.012	0.012	0.012	0.013	0.014	0.012	0.013	0.011
North Italian (ITA)	0.729	0.667	0.462	0.462	0.492	0.329		0.011	0.011	0.013	0.013	0.013	0.012	0.013	0.013	0.014	0.014	0.013	0.014	0.013
Greek (GRK)	0.724	0.660	0.437	0.424	0.472	0.306	0.332		0.008	0.012	0.013	0.012	0.012	0.012	0.011	0.014	0.013	0.013	0.013	0.012
Basque (BAS)	0.735	0.664	0.450	0.447	0.475	0.315	0.344	0.279		0.012	0.014	0.013	0.012	0.012	0.012	0.013	0.014	0.013	0.013	0.012
Brahui (BRU)	0.746	0.671	0.483	0.479	0.477	0.390	0.405	0.376	0.393		0.012	0.013	0.013	0.013	0.014	0.014	0.013	0.013	0.015	0.014
Sindhi (SDH)	0.737	0.689	0.489	0.480	0.507	0.411	0.419	0.391	0.422	0.404		0.012	0.013	0.013	0.014	0.014	0.014	0.014	0.015	0.013
Burusho (BSK)	0.745	0.678	0.493	0.485	0.508	0.406	0.405	0.378	0.400	0.390	0.392		0.013	0.012	0.013	0.014	0.014	0.013	0.014	0.012
Cambodian (CAM)	0.729	0.663	0.458	0.443	0.486	0.374	0.393	0.366	0.390	0.422	0.409	0.414		0.011	0.012	0.012	0.012	0.012	0.013	0.012
Chinese (CH)	0.719	0.659	0.457	0.453	0.469	0.391	0.407	0.370	0.392	0.422	0.416	0.423	0.317		0.011	0.013	0.013	0.012	0.012	0.013
Japanese (J)	0.718	0.670	0.466	0.452	0.499	0.407	0.410	0.392	0.419	0.440	0.439	0.441	0.345	0.321		0.013	0.012	0.012	0.013	0.012
Melanesian (M)	0.733	0.684	0.499	0.482	0.492	0.424	0.456	0.424	0.439	0.456	0.462	0.455	0.371	0.387	0.412		0.013	0.013	0.015	0.013
New Guinean (NG)	0.749	0.694	0.508	0.490	0.518	0.429	0.459	0.424	0.437	0.463	0.460	0.448	0.390	0.393	0.412	0.369		0.012	0.014	0.014
Australian (AUS)	0.730	0.676	0.464	0.459	0.473	0.391	0.412	0.390	0.399	0.413	0.435	0.431	0.359	0.372	0.394	0.364	0.367		0.013	0.013
Brazilian Indian (BI)	0.751	0.682	0.510	0.511	0.528	0.437	0.465	0.440	0.453	0.475	0.474	0.452	0.417	0.397	0.407	0.444	0.448	0.432		0.012
Mayan Indian (MI)	0.744	0.673	0.462	0.464	0.496	0.375	0.408	0.384	0.404	0.445	0.435	0.430	0.366	0.367	0.375	0.403	0.410	0.408	0.339	

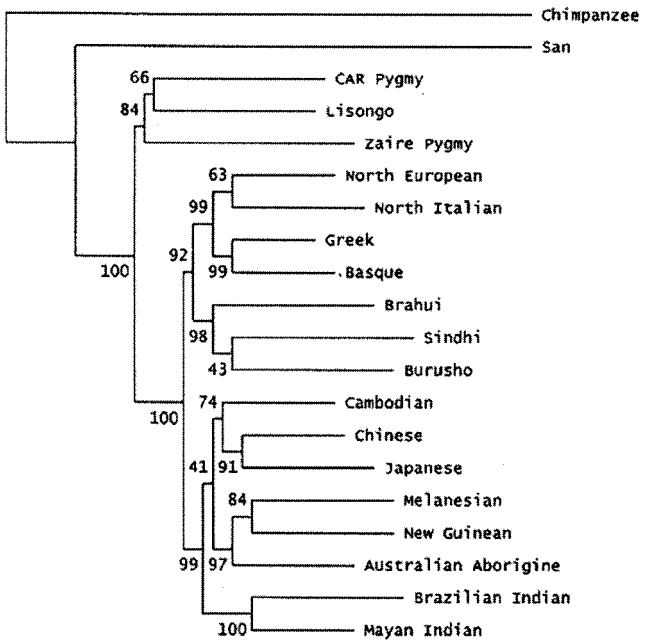

Fig. 1. Neighbor-joining tree representing relationship between 19 modern human populations and chimpanzees. Tree is based on D_{AS} genetic distance for 182 tri- and tetranucleotide microsatellite repeats. Bootstrap resampling values are provided at each fork, and represent number of times each group to right of that fork clusters among 10,000 trees generated from D_{AS} data.

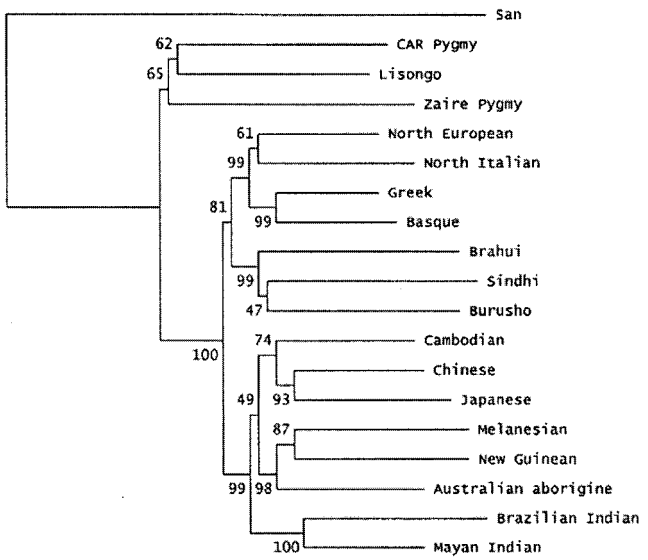

Fig. 2. Neighbor-joining tree, representing relationship between 19 modern human populations. Tree is based on D_{AS} genetic distance for 182 tri- and tetranucleotide microsatellite repeats. Bootstrap resampling values are provided at each fork.

define larger clusters. In general, it is likely that the lower bootstrap values are due to recent admixtures. Cavalli-Sforza et al. (1994) gave estimates of the time necessary for reducing the proportion of original genes in a population subject to known rates of gene flow from the outside, and exemplified distortions in a tree due to admixture. Cavalli-Sforza and Minch (unpublished observations) showed that

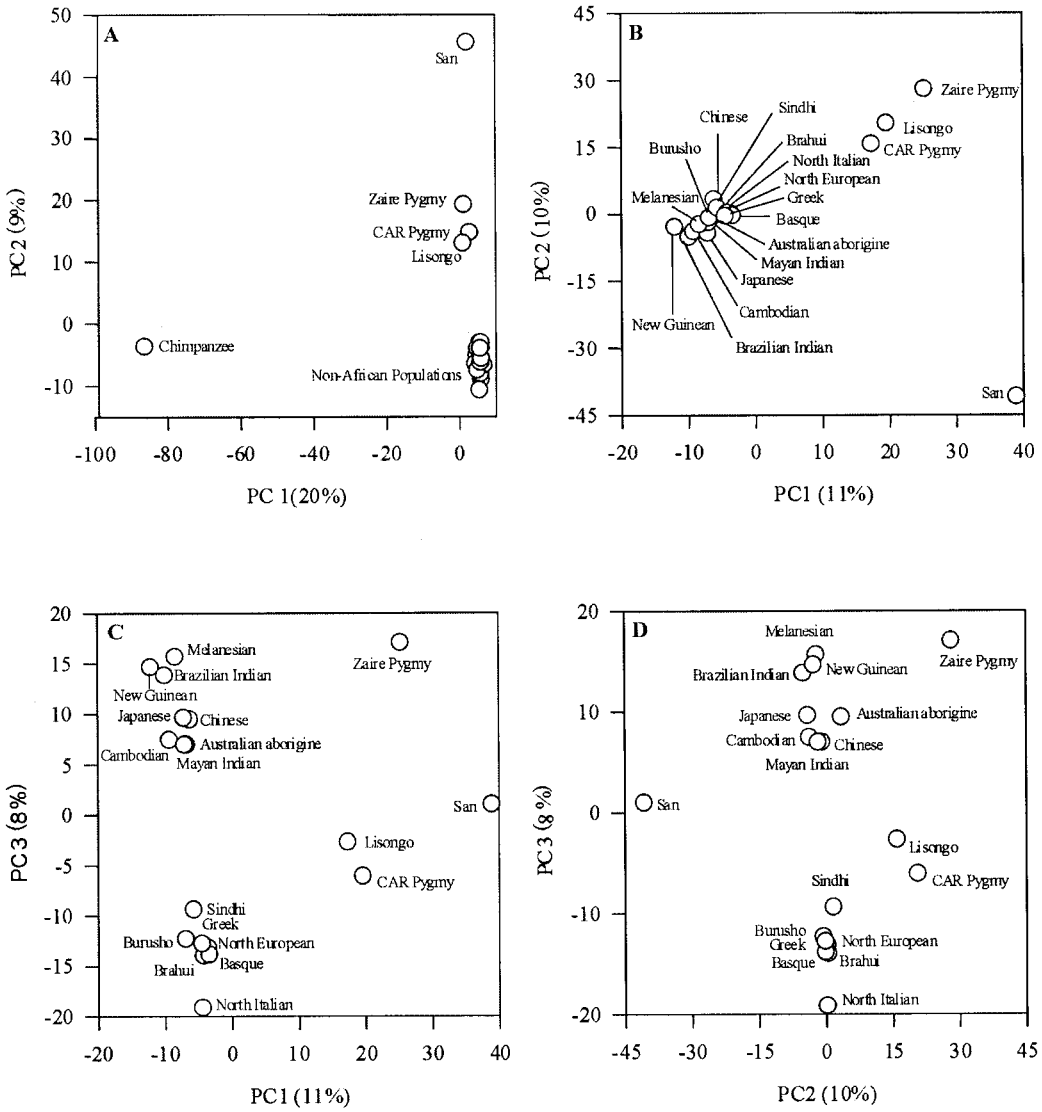

Fig. 3. Principal components (PC) analyses based on allele frequencies of 182 microsatellite markers in 19 worldwide human populations. **A:** Two-dimensional plot of PC1 vs. PC2, including chimpanzee samples. Bidimensional plots of PC1 vs. PC2 (**B**), PC1 vs. PC3 (**C**), and PC2 vs. PC3 (**D**), excluding chimpanzee samples.

neighbor-joining trees are especially sensitive to distortion due to admixture, which causes shortening of the branches of mixed populations, and also a movement, towards the center of the tree, of the location at which the mixed population branches off. Bowcock et al. (1991) illustrated an example of how the location of Europe on a neighbor-joining tree can be distorted by admixture.

The data are consistent with the "Out of Africa" theory, which proposes an African origin of modern humans, and which has also been supported by data from classical markers such as restriction fragment length polymorphisms, mitochondrial DNA sequence data, SNPs, and primarily dinucleotide repeats (Bowcock et al., 1994; Goldstein et al., 1995). The PC analyses corroborate the phylogenetic tree. The African and non-African populations separate into well-defined clusters. The African cluster is separated from all other world populations (Fig. 3). Based on autosomal markers, this is the only example that shows a clear separation between the San and other African populations. It is interesting that the San are the first branch, before the branch collecting the great majority of other Africans, and is in excellent agreement with the Y-chromosome tree (Underhill et al., 2000). This microsatellite tree is the first one offering a very strong confirmation (with a bootstrap value of 100% from 10,000 replications) of this finding, at least for biparentally transmitted markers (Fig. 2). The same result was already seen with mitochondrial DNA in studies by Johnson et al. (1983). Unfortunately, we have no populations from East Africa, some of which show that they also have, like the San, a substantial presence of haplogroups I and II, the two earliest branches of the Y chromosome (Underhill et al.,

TABLE 4. Population sample size (n), total number of alleles, unique alleles, and heterozygosities for grouped human populations

Population group	n	Total alleles	Unique alleles[1]	Heterozygosities
Africa	28	1,333	125 (9.4%)	0.7382
Europe	39	1,302	59 (4.5%)	0.7391
Pakistan	15	1,092	29 (2.7%)	0.7383
East Asia	30	1,237	70 (5.7%)	0.7157
Oceania	30	1,218	69 (5.7%)	0.6800
South America	26	1,061	41 (3.9%)	0.6607

[1] Values in parentheses indicate percentage of unique alleles in each region.

2000). This, together with the mtDNA information, is the strongest evidence in favor of "Out of Africa" and the major role played by the African populations presently living in the Rift Valley in the expansion of modern humans.

The 182 microsatellites show a significantly higher heterozygosity and number of alleles in Africa as compared with other continents, which is also consistent with an African origin of modern humans (Bowcock et al., 1994; Jorde et al., 2000; Reich and Goldstein, 1998). As shown in Table 4, the Africans also have a disproportionately higher number of total and unique alleles. Of the total of 1,826 alleles detected in human populations, 125 unique alleles were found in sub-Saharan Africa. The relatively higher heterozygosity in European and Pakistani populations could be due to the bias introduced by the initial selection of these microsatellite markers in Caucasoid populations (Mountain and Cavalli-Sforza, 1994; Rogers and Jorde, 1996). The Pacific Islanders (Oceania) and East Asian populations have the second highest percentage (5.7%) of unique alleles, whereas the South American Indians and Pakistanis have the lowest percentage of unique alleles. The lower average heterozygosity in Oceania and South America is probably indicative of their small effective population size (Deka et al., 1995). The split between the Africans and non-Africans is also seen with a smaller data set consisting of 85 of the 182 microsatellite loci, with alleles unique to the African population. A phylogenetic tree based on this data set (data not shown) separates the Africans from the non-Africans in 89% of 10,000 bootstrap resamplings, but does not provide strong statistical support for the separation of the other world populations. This suggests that the subsequent clustering seen in non-African populations is probably not due to continent- or population-specific alleles.

The study shows that these sets of microsatellite loci are efficient markers for high-resolution differentiation among various modern human populations. A positive correlation exists between genetic and linguistic classification on the continental level, but the lines blur at a local level, as in the case of Basques and Greeks in Europe, and the Brahuis, Hunza Burusho, and Sindhis within Pakistan. Preliminary surveys (Sun, 1997; Underhill et al., 2000) suggesting a genetic relationship between the Basque and Hunza Burusho were not supported by the present analysis. Further resolution of these population samples would probably involve a combination of SNPs and microsatellite data, and construction of compound haplotypes for phylogenetic analysis (Wilson and Balding, 1998; Xiong and Jin, 1999; Goddard et al., 2000). The genotyping of more microsatellite loci would probably not alter the genetic relationships observed in this study.

ACKNOWLEDGMENTS

We thank the donors for the samples used in this study. This work was supported by a core grant from the Government of Pakistan to S.Q.M. We are grateful to R. Qamar for much valuable help, and to the Aga Khan Medical Foundation Rural Health Program and the Department of Health, Government of Baluchistan, Pakistan, for their assistance in the collection of Burusho and Brahui samples, respectively. Greek samples were provided by M.P. DNA samples from chimpanzees and from populations outside Pakistan were from the collection of L.L.C.-S.

LITERATURE CITED

Adamson D, Albertsen H, Ballard L, Bradley P, Carlson M, Cartwright P, Council C, Elsner T, Fuhrman D, Gerken S, Harris L, Holik PR, Kimball A, Knell J, Lawrence E, Lu J, Marks A, Matsunami N, Melis R, Milner B, Moore M, Nelson L, Odelberg S, Peters G, Plaetke R, Riley R, Robertson M, Sargent R, Staker G, Tingey A, Ward K, Zhao X, White R. 1995. A collection of ordered tetranucleotide-repeat markers from the human genome. The Utah Marker Development Group. Am J Hum Genet 57:619–628.

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

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci USA 88:839–843.

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

Cavalli-Sforza LL. 1998. The DNA revolution in population genetics. Trends Genet 14:60–65.

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

Cooper G, Amos W, Bellamy R, Siddiqui MR, Frodsham A, Hill AV, Rubinsztein DC. 1999. An empirical exploration of the $(\delta\mu)^2$ genetic distance for 213 human microsatellite markers. Am J Hum Genet 65:1125–1133.

Deka R, Chakraborty R, Ferrell RE. 1991. A population genetic study of six VNTR loci in three ethnically defined populations. Genomics 11:83–92.

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

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.

Eller E. 1999. Population substructure and isolation by distance in three continental regions. Am J Phys Anthropol 108:147–159.

Gastier JM, Pulido JC, Sunden S, Brody T, Buetow KH, Murray JC, Weber JL, Hudson TJ, Sheffield VC, Duyk GM. 1995. Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. Hum Mol Genet 4:1829–1836.

Goddard KAB, Hopkins PJ, Hall JM, Witte JS. 2000. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. Am J Hum Genet 66:216–234.

Goldstein DB, Pollock DD. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J Heredity 88:335–342.

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

Greenberg JH. 2000. Indo-European and its closest relatives. The Eurasiatic language family. Stanford: Stanford University Press.

Grimes B. 1992. Ethnologue. Languages of the world: twelfth edition. Dallas, Texas: Summer Institute of Linguistics, Inc.

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.

Jeffreys AJ, Wilson V, Thein LS. 1985. Individual-specific "fingerprints" of human DNA. Nature 316:76–79.

Jin L, Underhill PA, Buoncristiani M, Robertson JM. 1997. Defining microsatellite alleles by genotyping global indigenous human populations and non-human primates. J Forensic Sci 42:496–499.

Jin L, Baskett ML, Cavalli-Sforza LL, Zhivotovsky LA, Feldman MW, Rosenberg NA. 2000. Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. Ann Hum Genet 64:117–134.

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL. 1983. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol 19:255–271.

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

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

Mantel NA. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–218.

Mountain JL, Cavalli-Sforza LL. 1994. Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. Proc Natl Acad Sci USA 91:6515–6519.

Mountain JL, Cavalli-Sforza LL. 1997. Multilocus genotypes, a tree of individuals, and human evolutionary history. Am J Hum Genet 61:705–718.

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170.

Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J. 1997. Microsatellite variation and the differentiation of modern humans. Hum Genet 99:1–7.

Quintana-Murci L, Semino O, Bandelt H-J, 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.

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

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

Rubinsztein DC, Leggo J, Amos W. 1995. Microsatellites evolve more rapidly in humans than in chimpanzees. Genomics 30: 610–612.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Sokal RR, Michener CD. 1958. A statistical method for evaluating systematic relationship. Univ Kans Sci Bull 38:1409–1438.

Sun B. 1997. The use of polymorphic microsatellites to determine genetic relationships between language isolates and neighboring populations. M.Sc. thesis. Stanford: Stanford University Press.

Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17: 6463–6471.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad M, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. 2000. Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361.

Venables WN, Ripley BD. 1999. Statistics and computing: modern applied statistics with S-Plus, third edition. Berlin: Springer-Verlag.

Weber JL, May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396.

Wilson IJ, Balding DJ. 1998. Genealogical inference from microsatellite data. Genetics 150:499–510.

Xiong M, Jin L. 1999. Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. Am J Hum Genet 64:629–640.