J Mol Evol (1991) 33:543–555

Journal of Molecular Evolution

© Springer-Verlag New York Inc. 1991

The Structure of Human Mitochondrial DNA Variation

D. Andrew Merriwether,[1] Andrew G. Clark,[2] Scott W. Ballinger,[3] Theodore G. Schurr,[3] Himla Soodyall,[4] Trefor Jenkins,[4] Stephen T. Sherry,[5] and Douglas C. Wallace[3]

[1] University of Pittsburgh, Department of Human Genetics, School of Public Health, Pittsburgh, PA 15261, USA
[2] Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA 16802, USA
[3] Center for Genetics and Molecular Medicine, Emory University School of Medicine, Rollins Research Center, Atlanta, GA 30322, USA
[4] MRC Human Ecogenetics Research Unit, Department of Human Genetics, School of Pathology, South African Institute for Medical Research and the University of the Witwatersrand, Johannesburg, South Africa
[5] Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA

Summary. Restriction analysis of mitochondrial DNA (mtDNA) of 3065 humans from 62 geographic samples identified 149 haplotypes and 81 polymorphic sites. These data were used to test several aspects of the evolutionary past of the human species. A dendrogram depicting the genetic relatedness of all haplotypes shows that the native African populations have the greatest diversity and, consistent with evidence from a variety of sources, suggests an African origin for our species. The data also indicate that two individuals drawn at random from the entire sample will differ at approximately 0.4% of their mtDNA nucleotide sites, which is somewhat higher than previous estimates. Human mtDNA also exhibits more interpopulation heterogeneity (G_{ST} = 0.351 ± 0.025) than does nuclear DNA (G_{ST} = 0.12). Moreover, the virtual absence of intermediate levels of linkage disequilibrium between pairs of sites is consistent with the absence of genetic recombination and places constraints on the rate of mutation. Tests of the selective neutrality of mtDNA variation, including the Ewens–Watterson and Tajima tests, indicate a departure in the direction consistent with purifying selection, but this departure is more likely due to the rapid growth of the human population and the geographic heterogeneity of the variation. The lack of a good fit to neutrality poses problems for the estimation of times of coalescence from human mtDNA data.

Offprint requests to: D. A. Merriwether

Key words: Mitochondrial DNA — Human evolution — Population genetics — Molecular anthropology

Introduction

A controversy over the time of coalescence of mitochondrial DNA (mtDNA) variation in humans began when Cann et al. (1987) suggested that all human mitochondrial genomes could be traced to a single ancestral haplotype present in Africa approximately 200,000 years ago. This date was determined from the observed level of mtDNA restriction site variation and an estimated rate of sequence divergence of 2–4% per million years (Brown et al. 1979; Brown 1980; Miyata et al. 1982). Although it has long been appreciated that the greatest diversity of mtDNA occurs among native Africans (Johnson et al. 1983), the divergence time of 200,000 years is widely contested by paleoanthropologists studying archaeological data, dental and bone morphology, cultural artifacts, and blood group and other nuclear genetic markers. In addition, ancient prehuman remains dating back to 3 million years ago have been found in Africa (Johanson et al. 1982), indicating that the time of human origin greatly predated the mtDNA coalescence time. However, uncertainties about the demographics of ancient human populations disallow the conclusion that these paleoanthropological findings are incon-

Table 1. Neutrality tests of human mtDNA restriction site variation

Population	Ewens–Watterson test					Tajima test		
	n	h	F	S	C	s	$\hat{k}$	D
Africans								
Wolof (Senegal)	110	14	0.237	5	0.345	13	1.366	−1.187
Peul (Senegal)	47	8	0.232	3	0.319	9	1.428	−0.847
Tukolar (Senegal)	12	5	0.389+	3+	0.583	4	1.273	−0.138
Other Senegalese	17	6	0.273	3	0.353	6	1.515	−0.491
San bushmen (1)	34	9	0.196	3	0.265	11	3.039	0.407
San bushmen (2)	59	10	0.215	3	0.286	24	5.148	0.104
Bantu	40	12	0.192	6	0.325	12	2.113	−0.773
Zulu	30	8	0.222	4	0.300	10	1.761	−0.958
Swazi	41	14	0.183	10++	0.293	28	3.324	−1.677
Xhosa	18	9	0.185	6+	0.333	12	2.111	−1.458
Tsonga	34	7	0.275	4+	0.382	11	1.405	−1.498
Sotho	29	10	0.153	5	0.241	13	2.305	−1.010
Pedi	23	6	0.301	2	0.478	9	1.881	−0.757
Tswana	15	9	0.147	6	0.200	12	2.305	−1.463
Venda	30	11	0.173	5	0.333	9	1.968	−0.416
Lemba	53	13	0.192	7+	0.340	14	1.795	−1.256
Herero	54	9	0.366	6+++	0.500	27	1.926	−2.199−−−
Dama	43	13	0.175	7	0.318	14	2.927	−0.281
Ovambo	22	7	0.343	4+	0.545	8	1.364	−1.243
Nama	46	9	0.318	2	0.532	12	2.296	−0.463
Richtersveld "coloured"	35	9	0.338+	2	0.556+	22	3.664	−1.053
Johannesburg "coloured"	71	13	0.155	6	0.239	17	3.630	0.094
Caucasians								
Finnish	112	10	0.612++	2	0.777++	11	0.881	−1.479
Romans	95	19	0.387+++	15+++	0.600+++	31	1.557	−2.280−−−
Sardinians	511	28	0.412+++	11+++	0.609+++	41	1.060	−2.235−−−
Caucasians (US, Europe)	50	15	0.359+++	12++	0.580+++	13	1.621	−1.323
Israeli Arabs	39	12	0.350+++	9++	0.564++	12	1.371	−1.600−
Israeli Jews	39	9	0.291	6++	0.385	10	1.449	−1.161
Sicilians	90	17	0.327++	11++	0.544++	20	1.763	−1.623−
Calabrians	60	11	0.459++	5	0.667++	8	1.128	−0.903
Albanians	100	13	0.487++	4	0.690++	15	1.098	−1.719−
Apulians (S. Italy)	87	15	0.303+	5	0.529+	14	2.084	−0.697
Bergamo (N. Italy)	103	9	0.604+	4+	0.767+	22	0.795	−2.373−−−
English (S. Africa)	36	7	0.434	2	0.639	12	1.449	−1.569
Jews (S. Africa)	55	6	0.479	4+	0.618	9	0.702	−1.770−
Afrikaners	51	7	0.706+++	5+++	0.837+++	10	0.842	−1.793−
Indians (S. Africa)	50	7	0.777+++	6++	0.880+++	6	0.392	−1.801
Asians								
Japanese	74	10	0.621++	5+	0.784++	21	0.756	−2.499−−−
Ainu	48	3	0.919	2	0.958	15	0.167	−2.936−−−
Koreans	78	13	0.492+++	8+++	0.692+++	22	0.922	−2.402−−−
Aeta	37	3	0.801	0	0.892	13	0.303	−2.858−−−
Vedda	20	4	0.490	2+	0.650	14	0.684	−3.026−−−
Orientals[a]	46	9	0.499++	4	0.696++	8	0.913	−1.382
Tharu	91	13	0.348+	5	0.527+	13	1.262	−1.385
Malaysian Indians	16	6	0.492+	5+	0.688+	5	0.625	−1.929−−
Malaysian Chinese	14	7	0.296+	5+	0.500+	10	1.839	−1.617−
Orang Asli (Malay)	32	7	0.269	3	0.406	6	1.776	0.540
Vietnamese	28	14	0.110	8	0.179	14	2.643	−0.899
Taiwanese	20	9	0.285+	6+	0.500+	15	1.937	−2.000−−−
Kadazan (Borneo)	26	6	0.426	3+	0.615	7	0.855	−1.634−
Rungus (Borneo)	3	2	0.556	1+++	0.667+++	1	0.667	−1.634−
Murud (Borneo)	1	1	–	–	–	–	–	–
Berungei (Borneo)	2	1	–	–	–	–	–	–
Bisaya (Borneo)	2	2	–	–	–	–	–	–
Mainland Malays	14	6	0.265	3	0.429	9	2.165	−0.902

Table 1. Continued

Population	Ewens–Watterson test					Tajima test		
	n	h	F	S	C	s	$\hat{k}$	D
Amerindians								
Warao	30	1	1.000	0	1.000	11	0.000	−3.218
Maya	37	4	0.493	0	0.676	5	1.072	−0.272
Ticuna	31	3	0.505	0	0.645	4	1.544	1.380
Pima	63	6	0.450	4	0.540	6	1.608	0.640
Papago	8	4	0.281	1	0.375	4	2.036	1.394
Hualapai	1	1	—	—	—	—	—	—
Other S.W. Amerindians	2	1	—	—	—	—	—	—
Africans	863	73	0.084+++	36+++	0.140+++	50	2.751	−2.016−
Caucasians	1478	97	0.184+++	48+++	0.298+++	58	1.218	−2.339−−−
Asians	552	74	0.171+++	37+++	0.370+++	61	1.330	−2.396−−−
Amerindians	172	12	0.325	4+++	0.488	18	1.550	−1.431
All pooled	3065	218	0.082+++	100+++	0.201+++	81	2.154	−2.127−−−

Symbols: n, the sample size; θ, the number of distinct haplotypes; F, the homozygosity; S, the number of singletons; C, the frequency of the most common haplotype; s, the number of segregating sites; $\hat{k}$, the estimate of θ derived from infinite alleles sampling theory; D, Tajima's (1989a) test statistic. +, probability of a random value exceeding the observed values is <0.05. ++, $P < 0.01$; +++, $P < 0.001$. Similarly, negative signs indicate significance of a one-tailed test with the observed value lower than expected
[a]This group consisted primarily of Chinese and Taiwanese Han and a few Japanese

sistent with the interpretation of mtDNA data by Cann et al. (1987). Analysis of mtDNA control region sequences from 83 individuals lead Vigilant et al. (1989) to estimate the deepest root of the human mtDNA phylogeny dated to 238,000 years ago. Similarly, Hasegawa and Horai (1991) analyzed the mtDNA sequence data from five different studies, and, after taking into account the variation in substitution rates among sites, they obtained a time for the deepest root of their dendrogram of 280,000 ± 50,000 years ago. Therefore, the consensus coalescence time appears to be about 250,000 years ago.

This paper describes the analysis of the largest set of data of human mtDNA polymorphisms assembled to date to assess the population structure and patterns of divergence of human populations. In studies of mtDNA restriction site variation, one of two sampling strategies has typically been employed. The first involves the analysis of the mtDNAs of a small number of individuals with numerous restriction endonucleases to produce a high resolution map of the mono- and polymorphic sites present in a population (Whittam et al. 1986; Cann et al. 1987; Stoneking et al. 1990). The second restricts attention to just six restriction enzymes, and samples many more individuals at a lower level of resolution (Johnson et al. 1983; Excoffier and Langaney 1989; Excoffier 1990). For questions relating to the impact of demographic history on human evolution, large, geographically structured samples are necessary. Here we focus on studies that took the second approach, which uses larger sample sizes and screens mtDNAs with the six core enzymes to identify polymorphic restriction sites. The entire set of samples consists of 3065 individuals from 62 population samples and is the largest such data set analyzed to date, being over 20-fold greater than the largest high-resolution data set (Cann et al. 1987).

Materials and Methods

Geographic Origins of mtDNA Samples. The strategy in assembling the data used for this analysis was to employ all of the currently available restriction fragment length polymorphism (RFLP) information for the six enzymes *Hpa*I, *Bam*HI, *Hae*II, *Msp*I, *Ava*II, and *Hinc*II. This compilation yielded 105 restriction sites, 81 of which were polymorphic, and identified 218 haplotypes. Because of missing data, not all of these haplotypes were distinct, and the tree-building methods identified the 149 distinct haplotypes. Individuals were sampled from 61 populations and were aggregated into four major ethnic groupings (Asian, Caucasian, African, and Amerindian). Table 1 indicates the geographic origin of all the samples and identifies the sizes and ethnic grouping of the samples. All of the samples were verified to have different maternal grandmothers and, to that extent, represent distinct maternal lineages.

The Asian populations consist of 74 Japanese and 48 Ainu from northern Japan (Harihara et al. 1988), 78 Koreans from South Korea (Harihara et al. 1988; Ballinger et al. 1991), 37 Aeta [Negritos] (Harihara et al. 1988), 20 Vedda from Sri Lanka (Harihara et al. 1988), 91 Tharu from Nepal (Brega et al. 1986b), 46 Asians consisting mainly of Chinese and Taiwanese Han (Johnson et al. 1983), 14 Malays, 14 Chinese Han, and 32 Senoi Aborigines from peninsular Malaysia (Ballinger et al. 1991), 28 Vietnamese from South Vietnam (Ballinger et al. 1991), 20 Taiwanese Han (Ballinger et al. 1991), 26 Kadazan, 3 Rungus, 1 Murud, 2 Brunei, and 2 Bisaya from Sabah State (Borneo), Malaysia (Ballinger et al. 1991), and 16 Indians from the Malay peninsula (unpublished).

The Caucasian populations consist of 112 Finns (Vilkki et al. 1988), 95 Romans (Brega et al. 1986a; Santachiara Benerecetti et al. 1988), 511 Sardinians (Brega et al. 1986a; Santachiara

Benerecetti et al. 1988; Sartoris et al. 1988), 50 Americans and Europeans (Johnson et al. 1983), 39 Israeli Arabs and 39 Israeli Jews (Bonné-Tamir et al. 1986), 103 Northern Italians (Sartoris et al. 1988) from the Bergamo area, 87 Apulians (De Benedictis et al. 1989a,b) from southern Italy, 60 Calabrians (Torroni et al. 1990) from southern Italy, 100 Albanians (Torroni et al., unpublished) living in Calabria, southern Italy, 90 Sicilians (Semino et al. 1989), and 36 English, 55 Jews, 50 Indians, and 51 Afrikaners from South Africa (Soodyall et al., unpublished).

The native African samples consist of 110 Wolof, 47 Peul, 12 Tukolar, and 17 other Senegalese (Scozzari et al. 1988), 34 San bushmen from Botswana (Johnson et al. 1983), and 40 Bantu from Johannesburg, South Africa (Johnson et al. 1983). A more recent sample from South Africa included 123 Nguni speakers (30 Zulu, 41 Swazi, 18 Xhosa, and 34 Tsonga), 67 Sotho/Tswana speakers (29 Sotho, 23 Pedi, and 15 Tswana), 129 Bantu speakers (30 Venda, 53 Lemba, 54 Herero, 22 Ovambo), 148 Khoisan speakers (43 Dama, 59 San, and 46 Nama), and 106 people of mixed race whose mtDNA was mostly of negroid origin (35 Richtersveld "coloured" from the Richtersveld area in the Cape, and 71 Johannesburg "coloured" (Soodyall et al., unpublished).

The Amerindian samples consist of 63 Pima, 8 Tohono O'odham (Papago), 1 Hualapai (Hopi), 1 Navajo, and 1 Pomo from southern Arizona (Wallace et al. 1985), 37 Maya from the Yucatan peninsula of Mexico (Schurr et al. 1990), 31 Ticuna from western Brazil (Schurr et al. 1990), and 30 Warao from Venezuela (Johnson et al. 1983).

Molecular Methods. The procedures for collecting blood samples, isolating the mtDNAs, and performing Southern blot hybridizations are described in the original papers. The complete nucleotide sequence of a human mtDNA (Anderson et al. 1981) made constructing unambiguous restriction maps possible. The resolution of the restriction maps was moderate, with 105 restriction sites screened and a total of 81 polymorphic sites observed among the samples. All these samples were surveyed with at least four of the six enzymes. The Harihara et al. (1986) samples were surveyed without *Hae*II, the Johnson et al. (1983) and Bonné-Tamir et al. (1986) samples without *Hinc*II, and the Johnson et al. (1983) Warao Indian samples without *Hae*II or *Hinc*II, the Sartoris et al. (1988) samples without *Hpa*I or *Hinc*II. The remaining samples were surveyed with all six enzymes. Those sites corresponding to enzymes that were not screened were coded as missing data.

Descriptive Statistics. Heterozygosity per necleotide site was estimated from the restriction site data with the methods of Engels (1981) and Nei and Miller (1990). The Engels method provides an estimate of the heterozygosity, $\hat{H}$ [Eq. (11), p. 6330] and the variance of $\hat{H}$, $V(\hat{H})$ [Eq. (21), p. 6333] were calculated in a manner that makes no assumptions about the evolutionary past of the sample. The degree of population subdivision was quantified with the statistic h following the methods of Weir and Cockerham (1984) for a single haploid locus with many alleles, and with the statistic G_{ST} following the method of Takahata and Palumbi (1985). The statistical association among pairs of sites, or linkage disequilibrium, was calculated and reported as D', which is scaled to the minimum and maximum value that D can attain (Lewontin 1988). Because 3240 pairwise D' values were generated, these results are only reported as a histogram. Nucleotide divergences were estimated from the number of shared and different restriction sites using the method of Nei and Miller (1990) and are expressed as the number of substitutions per nucleotide site. In this method, the estimators take into account the possibility of multiple substitutions and reverse mutations. Nucleotide divergences were estimated for all pairs of types scored for 74 or more of the polymorphic sites. This distance matrix was used to generate dendrograms with the unweighted pair-group method using arithmetic averages (UPGMA) cluster algorithm and with the neighbor-joining algorithm of Saitou and Nei (1987). Confidence in the tree topology was assessed by bootstrapping over the restriction sites. A tree whose terminal branches are the 62 population samples was calculated by estimating the average nucleotide divergence between individuals randomly drawn from all possible pairs of populations. In a similar way, a tree with the four major groups (Asian, Caucasian, African, and Amerindian) was constructed.

Neutrality Tests. The Ewens–Watterson test was used to test the fit of observed data to the infinite alleles model (Ewens 1972; Watterson 1977). Sampling theory of the infinite alleles model shows that the expected allele frequency distribution is fully specified by the sample size (n) and the number of observed alleles (k). Given n and k, the test generates realizations of a sampling from this expected distribution. One-tailed tests were done by scoring the fraction of the simulations whose test statistic was greater than the observed number. In addition to the traditional test of homozygosity or gene identity (F), the significance of departures from expected number of singleton alleles (S) and from the frequency of the most common allele (C) was tested. The advantages of this approach were demonstrated in Clark (1987).

The neutrality test of Tajima (1989a) was also performed for each population and for various groupings of the data. The Tajima test is based on the concordance between estimates of $\theta = 4Nu$ obtained from the heterozygosity calculations and from the number of segregating sites. The former estimate depends on haplotype frequencies and is derived from the infinite alleles model. The latter estimate does not use population frequency data and is based only on the sample size and the number of segregating sites. The estimate is derived from the infinite sites model of neutral molecular evolution. The test can be applied either to nucleotide sequence data or to RFLP data. Rejection of the null hypothesis (that the data acceptably fit the predictions of the neutral theory) can be attributed to the same causes as rejection of the Ewens–Watterson test, including population heterogeneity, lack of equilibrium, or natural selection. In addition, the Tajima test assumes that the mutation rate of all sites is equal.

Results

Measurements of Diversity

The algorithm of Nei and Miller (1990) gave an estimate of the average number of nucleotide substitutions per site as 0.0043 ± 0.0010. The matrix of pairwise nucleotide divergences was estimated both by the method of Nei and Miller (1990) and Nei and Tajima (1983), and the values were found to be nearly indistinguishable. The most diverse pair of haplotypes differed at an estimated 3.1% of the nucleotide sites. Engels (1981) measure of heterozygosity per site yielded $\hat{H} = 0.0040 \pm 0.0006$. Engel's measure of mean heterozygosity corrects for the ascertainment bias that results from restriction site polymorphism data and it makes no assumptions about the processes by which the variability arose or is maintained. These estimates imply that the probability that a particular nucleotide site will differ between two randomly chosen humans is about

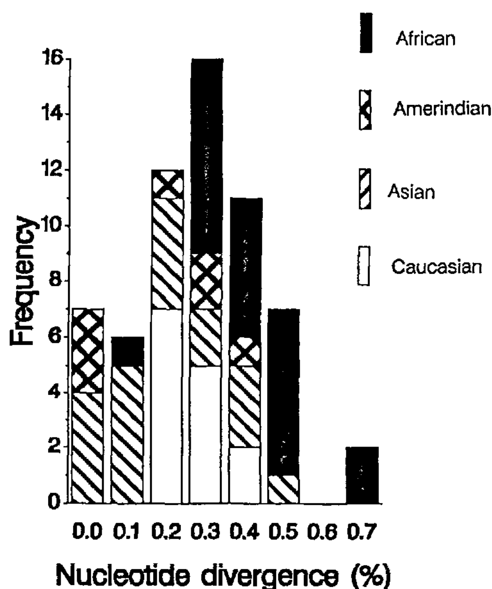

Fig. 1. Histogram of nucleotide diversity (π) within 59 of the 62 populations sampled (the two San samples were pooled and the admixed Richtersveld and "coloured" populations were excluded). Note that the Africans have the most diverse populations, and the Amerindians and Asians have the least diverse populations.

0.004, which is not significantly different from the estimate of 0.0032 of Cann et al. (1987). Human mtDNA is thus about 10 times as diverse as untranslated and nondegenerate sites in human nuclear genes, and about three times as diverse as fourfold degenerate sites in human nuclear genes (Li and Sadler 1991). In addition, the probability of drawing two individuals having identical mtDNA restriction patterns (with respect to the enzymes examined here) was 0.08. Moreover, the probability of identity varied widely among the four major groups, from 0.083 for Africans to 0.324 for Amerindians. Figure 1 shows that the different populations exhibit different levels of polymorphism at the nucleotide level, and that the major groups also differ in mean nucleotide diversity (e.g., African populations have the highest diversity).

Population Subdivision

The Weir and Cockerham (1984) measure of the degree of population subdivision was $\hat{\theta} = 0.354 \pm 0.025$. Similarly, the algorithm of Takahata and Palumbi (1985) yielded an estimate of $G_{ST} = 0.46 \pm 0.083$. These estimates do not differ significantly from that of Stoneking et al. (1990), who estimated $G_{ST} = 0.31$ from their sample of 241 individuals and 182 haplotypes.

Linkage Disequilibrium

Figure 2 shows the frequency histogram from the 3240 values of D'; 1597 values were less than -0.95, 1172 values were greater than 0.95, and 319 values fell between -0.05 and 0.05. Only 152 values were intermediate, suggesting that no or virtually no recombination has occurred in the sample. This result confirms a similar observation made in a study of Amerindian mtDNA variation (Schurr et al. 1990).

Nucleotide Diversity and Dendrograms

From the matrix of pairwise estimates of nucleotide divergence (d_{ij}), a UPGMA tree having 149 haplotypes was constructed (Fig. 3). The tree has several rather deep branches, but its most noticeable feature is the very deep branch that separates the topmost haplotype from the others. A similar deep branch was found by Johnson et al. (1983) and by Cann et al. (1987), who found seven African lineages in a clade that departed markedly from the rest of their sample. Figure 4 presents the same dendrogram as in Fig. 3 except the four panels are highlighted to

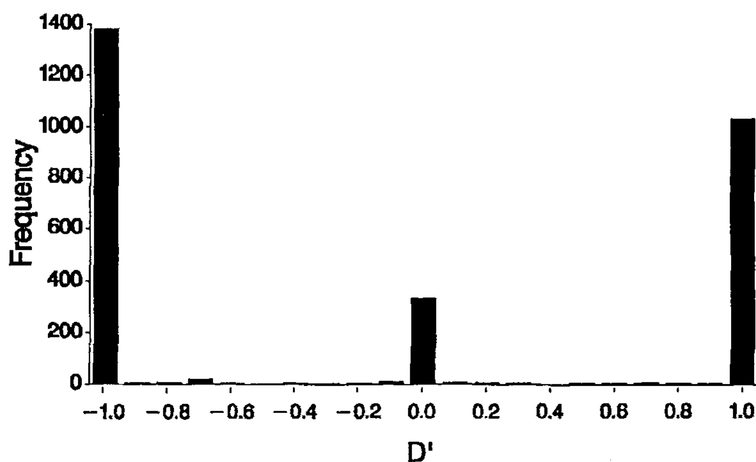

Fig. 2. Histogram of linkage disequilibria between all pairs of polymorphic sites

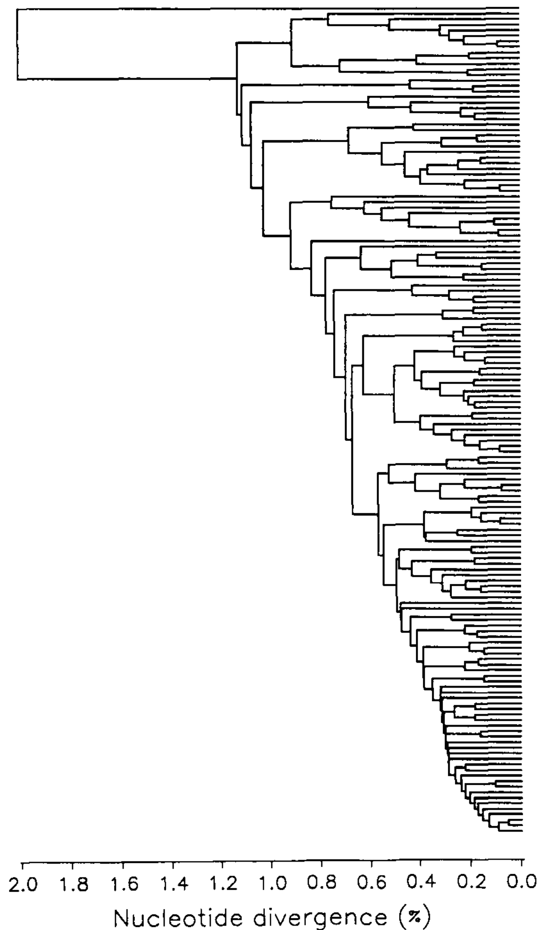

Fig. 3. UPGMA dendrogram of the 149 haplotypes identified by analysis of mtDNA with six restriction enzymes. The dendrogram was constructed from the distance matrix that was calculated following the algorithm of Nei and Miller (1990). The type designations of the ends of each branch of the tree are, starting at the top: 5-5, 5, 14; 73-2; 61-2; 19; 20; 66-2; 18-2, 18; 18, 18-5; 90-2; 57-2; 36; 119-2; 15, 15-2; 142-2; 125-2: 51-2; 78; 2; 128-2; 89-2; 58-2; 11, 11-2; 120-2; 4, 4-9; 129-2; 124-2; 40, 40-2; 33; 137-2; 134-2; 115-2; 32; 121-2; 3-5; 3, 116-2; 29; 82-1; 106-1; 111-1; 9-1, 9-1@, 9; 8-1*, 110-1; 8-1@, 8-1; 8, 96-2; 81-2; 42-2, 42; 50-2; 49-2; 54-2; 48-2; 132-2; 41; 46-6; 25; 93-6; 1-6, 1-6@, 1-9; 92-3; 69-3; 12; 138-2; 133-2; 130-2; 63-2; 139-2; 123-2; 135-2; 118-2; 113-2; 68-2; 30; 10, 10-2; 71-2; 136-2; 140-2; 7, 7-2, 31; 131-2; 34, 34-2; 65-2; 35; 43; 2-7; 2-2; 2, 2-10; 75-2; 53-2; 24, 24-2; 37-2*, 37; 22-2, 22, 22-2*; 103-2, 21-2*; 21-2; 21-5; 21, 21-9; 80-2; 67-2; 122-2, 47-2*; 47-2, 47-2@, 105-2; 141-2; 44; 23, 23-15; 127-2; 85-2; 86-2; 13'-2; 16-2, 16; 28-2, 28; 144-2; 102-2; 6-2, 6-2@; 6, 6-6; 100-16; 101-17; 98-2; 72-2, 97-2; 117-2; 60-2; 38, 38-2; 79-2; 59-2; 56-2, 56-2@; 64-2; 39-2, 39-2@, 39; 27-2, 27; 77-14; 1-14; 143-2; 109-2; 95-2; 93-2; 83-2; 84-2; 62-2; 112-2; 147-2; 126-2; 70-2; 52-2; 74-2; 76-2; 17; 1-10, 1-10@; 16-2, 26-2@; 45, 1-5*; 1-5, 1-5@; 1-7; 13-2, 13; 1-13; 91-15; 1-18; 99-2, 1-2*; 1-2, 1-2@, 104-2; 1, 147-2. The labels for the ends of the branches are separated by semicolons. Note that many branches are represented by more than one type, which could appear equivalent due to missing data.

show the haplotypes present in each of the four major groupings. These dendrograms show that the Africans span the deepest branch of the tree, and are the most diverse group. Caucasians and Asians are nearly as diverse, but neither of these two groups had the haplotype in the deepest branch of the tree. The figures clearly show that the Amerindians have the lowest diversity. The dendrograms not only allow one to infer differences in diversity among the groups, they also provide evidence that the time to the most recent common ancestor is the greatest for the Africans and the least for the Amerindians.

Figure 5 shows the UPGMA tree drawn from the matrix of average distances among the 62 population samples, and Fig. 6 presents the neighbor-joining tree depicting the relationships among the four major groups. The latter tree is consistent with the interpretation of the highlighted trees of Fig. 4, with a branching order of Africans, then Caucasians, followed by Asians and Amerindians. Bootstrapping this tree over sites and individuals showed that all three nodes were significant at the 95% level.

Neutrality Tests

Results of the Ewens–Watterson test of fit of the data to the infinite alleles model are shown in Table 1. In most populations and population groupings, the observed values for F, the sum of squared haplotype frequencies (or homozygosity), was higher than expected by the infinite alleles model, based on the sample size and the observed number of haplotypes. The most common allele frequencies (C) were greater than the expected values, and the number of singleton alleles (S) generally exceeded the expected number of singletons. In no case was there a significant deficit of any of these statistics. All these observations can be interpreted as a deficit of heterozygosity or an excess of homozygosity and signify a clear departure from the expectation of the infinite alleles model.

Results of the test of Tajima (1989a) are also reported in Table 1. In most cases the D statistic is negative, indicating that the estimate of $\theta = 4Nu$ based on the allele frequency distribution is less than the estimate of θ based on the number of segregating sites. This finding is consistent with the excess homozygosity seen in the Ewens–Watterson test results. Figure 7 shows that θ estimated from the number of segregating sites ranges as high as 10, while the estimate of θ estimated from the heterozygosity never exceeds 4.

Discussion

The dendrograms of Figs. 3–6 clearly reveal that the Africans possess the greatest mtDNA diversity. This observation is consistent with paleontological data and suggests that the African population represents the oldest group. The diversity in the oldest parental

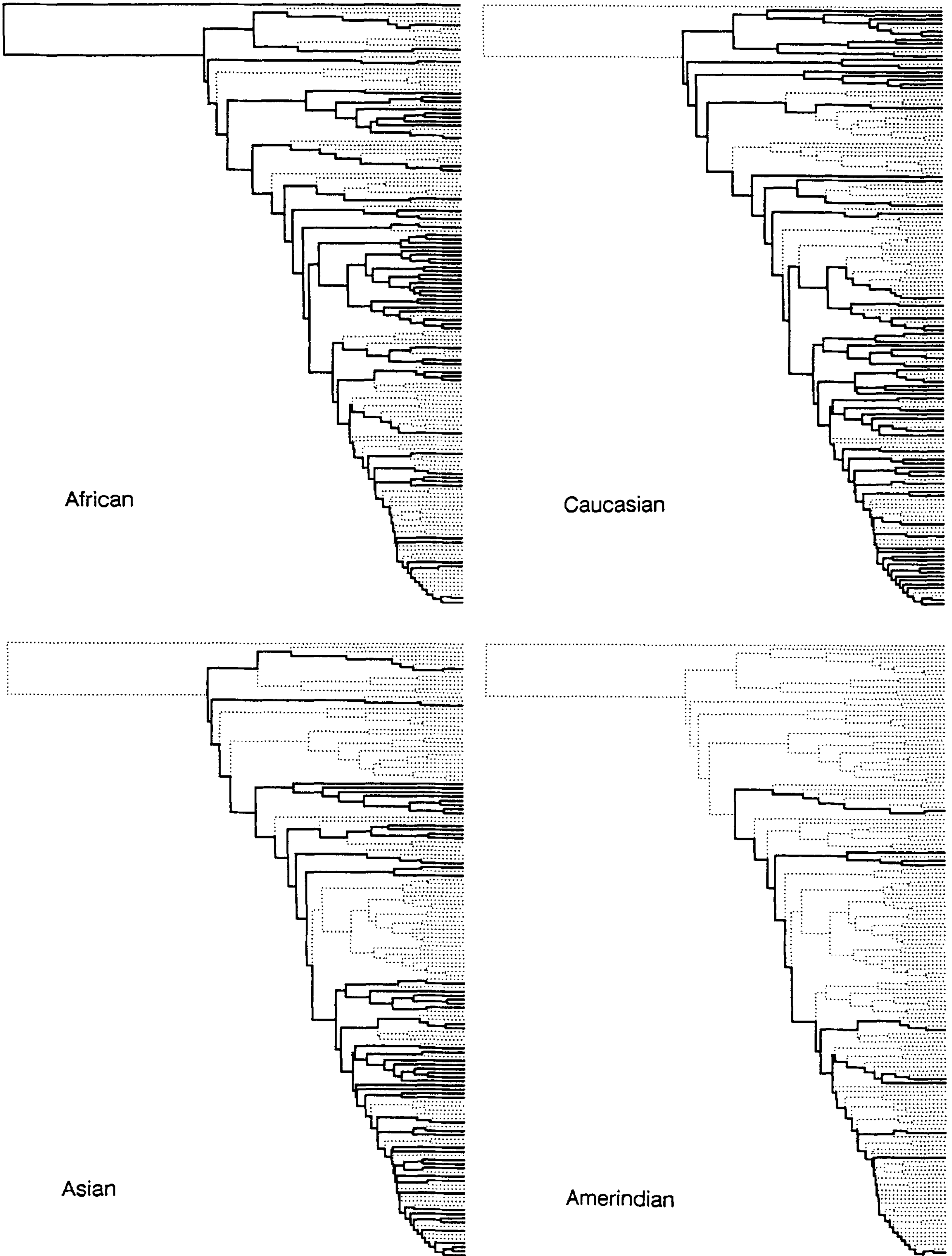

Fig. 4. UPGMA dendrogram of all haplotypes, as in Fig. 3, which the paths connecting haplotypes present in each of the four major groups that are highlighted

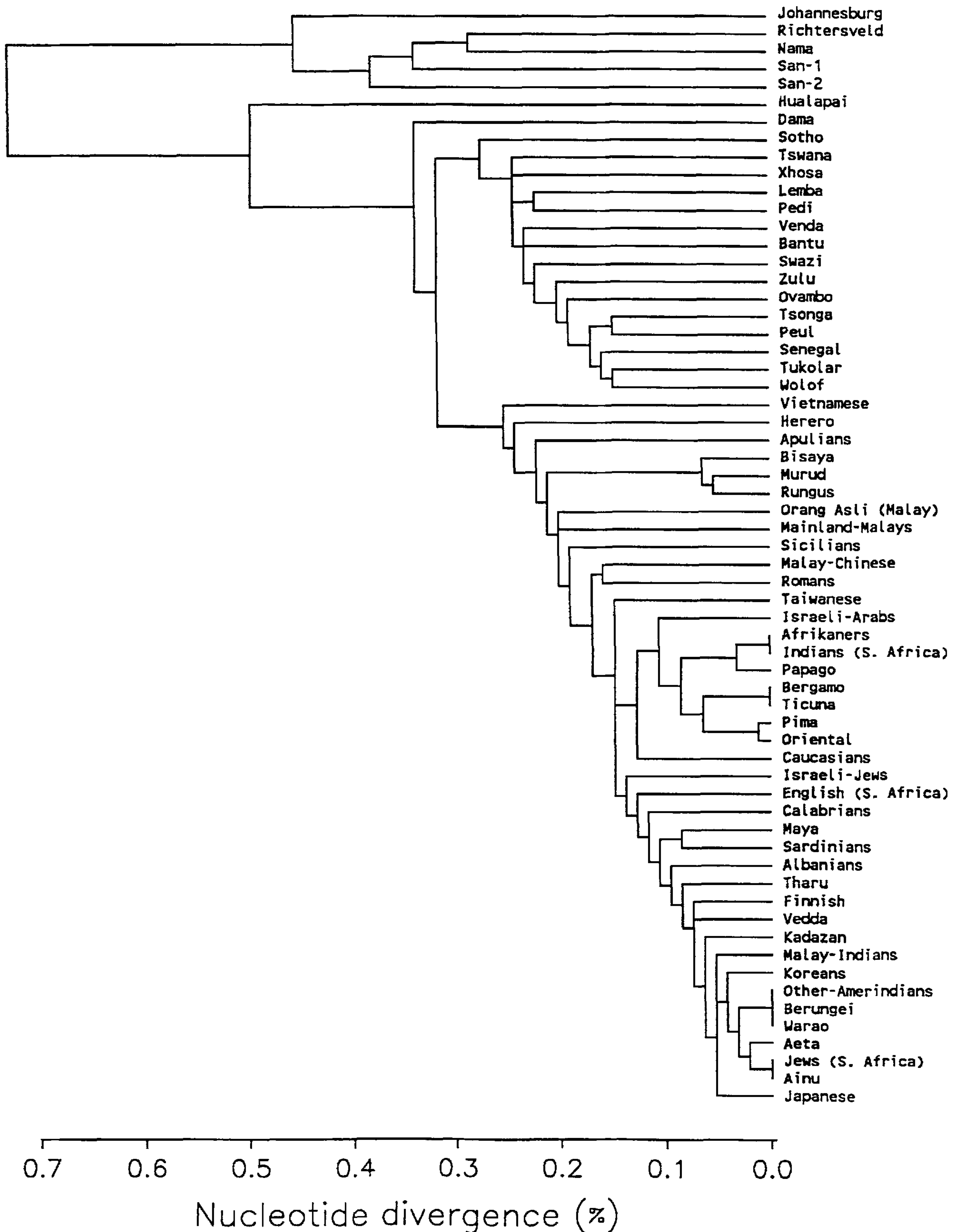

Fig. 5. UPGMA dendrograms based on the average distance between randomly chosen individuals from pairs of populations

population is expected to be greater under the following simple scenario. If a series of migrations out of a parental population occurred, then the new daughter populations would each have less variability than the source population, because of a combination of founder effects and population bottlenecks. If the original parental population remains large in size, it should possess more variation than

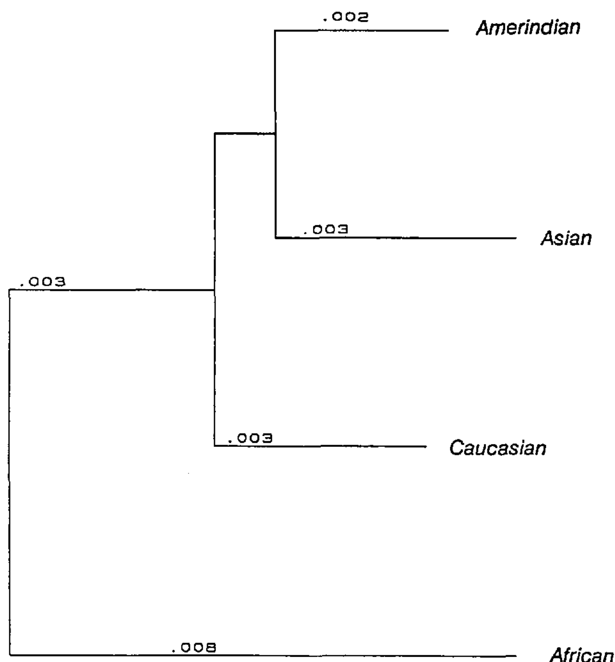

Fig. 6. Neighbor-joining dendrogram constructed from the matrix of average distances among people chosen from the four majors groups

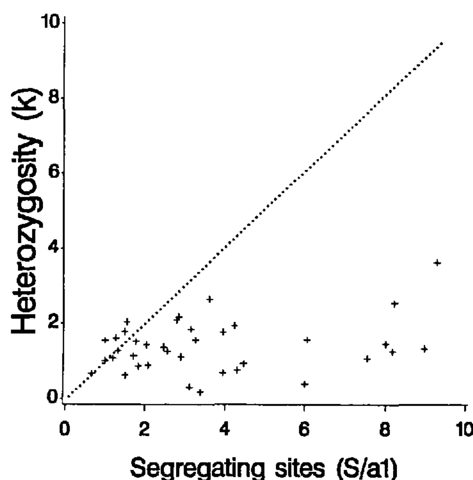

Fig. 7. Results of the test of Tajima (1989a) performed separately on each of the 62 population samples. $\hat{k}$ is the heterozygosity, or probability that two genes drawn at random from the population differ. $\hat{k}$ is an estimator of $h = 4Nu$ under the infinite alleles model. S is the number of segregating sites, and $a_1 = \Sigma (1/i)$. The quantity S/a_1 is an estimate of $\theta = 4Nu$ following the infinite sites model. If the variation were strictly neutral, an equilibrium population would yield estimates with $\hat{k} \approx S/a_1$, as indicated by the dotted line. Note that there were several populations with a large excess of segregating sites in proportion to their heterozygosity. This departure from neutrality is consistent with the findings of the Ewens–Watterson test.

any of its daughter populations. This scenario emphasizes why attention must be paid to demographic patterns of the ancient populations. Human populations have undergone tremendous demographic changes in the last 200,000 years, and the rate of accumulation of molecular variation and the maintenance of that variation depends on this history. Uncertainties in this history, coupled with the large stochasticity in sampling of phylogenies, force us to be cautious in interpreting the results of Figs. 3–6. Numerical simulations of Avise et al. (1984) show that coalescence to a single mtDNA type is likely in a population of 15,000 people in the time spans suggested by the molecular data.

Several other species, including eels, catfish, and redwing blackbirds, also exhibit a deficit of mtDNA variation compared to the expectation based on the current breeding population size. Avise et al. (1988) determined estimates of long-term effective population sizes that fit the observed distribution of pairwise differences and found the estimates to be much smaller than the current observed population sizes. The problem with the human data is that the Ewens–Watterson and Tajima tests show that there is no single value of population size that allows the data to fit the neutral model. We are convinced that the reason is the long-term growth of the human population, which is further illustrated by the nongeometric distribution of pairwise differences (Slatkin and Hudson 1991). Attempts to fit demographic models to the human data are in progress.

The suggestion of an African origin is consistent with the data from a variety of sources, including other studies of mtDNA (Johnson et al. 1983; Cann et al. 1987). More recent data using either high resolution restriction mapping (Stoneking et al. 1990) or sequencing of the control region (Horai and Hayasaka 1990; Vigilant et al. 1990) also support the African origin. One dissenting opinion was expressed by Excoffier and Langaney (1989), who claimed that the type I haplotype (the bottom haplotype of Fig. 3) was the oldest because it was the most common. They further argued that Caucasians were the closest to the ancestral mtDNA type. The dendrogram of Fig. 3 does not disallow the possibility that type I is the oldest haplotype, but the data support an African origin for the reasons stated above.

Horai and Hayasaka (1990) examined DNA sequence variation in the control region of 95 mtDNA isolates and also found that Africans formed a major branch. Contrary to our results, however, they found Orientals to branch next, with the Japanese forming two major branches, as in Horai and Matsunaga (1986). Our finding that the Caucasians branched first after the Africans also departs from the findings of Cann et al. (1987). The discrepancy may illustrate a problem caused by heterogeneity of demographic factors in different parts of the phylogeny. Livshits and Nei (1990) noted that isolated populations with a small effective population size (such as the San bushmen) generally have longer branches than do

groups with larger effective population sizes, but the problem of bottlenecks in parts of a phylogeny requires further study.

Measures of diversity of mtDNA have varied among studies in part because of differences in analytical methods and in part because of the differing evolutionary rates of the coding and noncoding D-loop regions of the mitochondrial genome being examined. The high resolution restriction mapping of Stoneking et al. (1990) yielded very similar estimates of nucleotide divergence to ours, with the deepest root of their tree extending to a nucleotide divergence of about 0.6%. They also obtained a measure of population subdivision that was remarkably close to ours ($G_{ST} = 0.31$). In contrast to the G_{ST} value for mtDNA, the value for nuclear genes was estimated to be 0.12 (Nei and Livshits 1990). The high G_{ST} value for mtDNA underscores the utility of mtDNA for inferring human population structure and past migrations. In addition, it probably reflects the more rapid evolutionary rate of mtDNA, and may also indicate sex differences in migration (Takahata and Palumbi 1985).

The Ewens–Watterson test reveals an excess of the most common allele and an excess of singletons in the pooled samples as well as in most of the individual populations. This result can be viewed as an excess of homozygosity or a deficit of variation (heterozygosity). The Tajima test produced a qualitatively similar result, showing an excess of segregating sites over what would be expected based on the level of diversity. There are four possible explanations for the rejection of neutrality, including natural selection, population growth or failure to attain equilibrium, population heterogeneity, and inhomogeneity in the neutral substitution rate. Purifying selection is expected to be manifested in the excess of synonymous over nonsynonymous nucleotide substitutions, and this will be examined in detail in a subsequent paper. The opportunity for purifying selection is clear, given the deleterious consequences of certain mitochondrial mutations (Wallace et al. 1988), but the data at present do not provide a strong case for selective differences among mtDNA types. Stoneking (1990) objected to Excoffier's (1990) application of the Ewens–Watterson test to low-resolution mtDNA data, claiming that with sufficiently high resolution data, neutrality cannot be rejected. Although Excoffier (1990) overstated the case for selection, the evidence that human mtDNA variation is not in mutation–drift equilibrium is incontrovertible. Stoneking's failure to reject neutrality with high resolution data reflects the lack of power of the Ewens–Watterson test when the allele frequency distribution becomes too even (as more sites are examined, more haplotypes appear unique). High resolution data provide a particularly useful test of the distribution of counts of mismatches (or distribution of branch lengths), and these analyses clearly demonstrate the impact of the expansion of the human population on extant patterns of mtDNA variation (Di Rienzo and Wilson 1991).

Population heterogeneity is apparently another cause for departure from neutrality, as the pooled population exhibits greater departure than does the single population (Whittman et al. 1986; Chakraborty 1990). A G_{ST} of 0.35 indicates that over one-third of the variation is interdemic, although we have no measure of the degree of heterogeneity at a more localized geographic level. Rapid population growth is almost certainly a major cause of departure from neutrality in these data. We should see a large number of singletons because we expect to see a large number of new unique mutations in the rapidly expanding populations. These unique mutations are typically population specific and remain rare (often appearing in only one individual). This trend is not overly surprising considering the rapid expansion of most human populations during recorded history (Weiss 1989).

Every sample of genes must exhibit the phenomenon of coalescence to a common ancestor, but the estimation of this time of coalescence requires that extant genetic variation be interpreted in terms of population genetic models. The implicit assumptions of these models are that the genetic variation is neutral, and that the population is in mutation–drift equilibrium. Our data clearly demonstrate that mtDNA variation violates these assumptions. The observed departures from equilibrium may significantly influence the accuracy of coalescence time estimates. One possible cause for finding a common ancestor that occurred more recently than expected is rapid population growth. Nei and Li (1975) and Nei et al. (1975) discuss the genetic consequences of rapid population growth, especially in a population that has just passed through a bottleneck. When the group first passes through the bottleneck, it loses many alleles. During the subsequent population growth, the alleles that are present will undergo rapid drift in this small population as an allele is much more easily lost if it is possessed by only a few individuals. Any new mutations will be introduced at a relatively high frequency (as compared to larger populations such as the parent population) because the population size is small, thus increasing the probability that the mutation will remain in the population. The increased protection from loss of a new variant in a rapidly growing population results in a faster recovery of the number of segregating sites than of the recovery of heterozygosity. Tajima (1989b) showed that the number of segregating sites is more influenced by the current population size than is the average number of nucleotide differences,

but that a bottleneck in the past will have a greater effect on the average number of nucleotide differences than on the number of segregating sites. This finding is also consistent with the observed pattern of human mtDNA variation. Because it is clear that mtDNA diversity is not in equilibrium among human populations, estimates of population age based on coalescence time need to be interpreted with caution.

The Amerindians possess the shallowest root in the dendrogram indicating a recent divergence from Asian populations. Archaeological evidence suggests that the founding of the New World took place in the last 10–35,000+ years, and although the shallow Amerindian root in the mtDNA tree is consistent with this evidence, the current data are insufficient to sharpen this estimate. It has been shown (Wallace et al. 1985; Schurr et al. 1990; Torroni et al. 1991) that the Amerindian mtDNAs represent a subset of Asian mtDNAs with the Asian haplotypes present at drastically altered frequencies. This finding is indicative of a founder effect where a small group split off from the parent group and founded a new population. The subset of alleles present in this founder population constitute all of the alleles in the new population. The subset of alleles present in this founder population constitute all of the alleles in the new population. Any rare alleles entering in the founder group will initially be present at much higher frequency than had existed in the parent population. Genetic drift will fix or eliminate these alleles in the new population, resulting in some rare alleles being present in much higher frequency than in the parent population and some common alleles from that parent population being lost in the newly founded population.

It is clear that a relatively small number of mtDNA lineages founded populations in the New World and that migrations through the Americas occurred fairly rapidly (Schurr et al. 1990). Indeed, sites at the southern tip of South America (Pell's Cave) were inhabited at least 11,000 years ago (Shutler 1983). If we take 30,000 years ago as an entry point, ancestral Amerindians had reached and inhabited all parts of the New World in just 19,000 years. We know that at the time of European contact the Amerindians in the New World numbered in the millions, as some Maya and Aztec cities comprised up to 100,000 people. This extraordinary rate of population growth forces a conservative interpretation of any tests that are based on equilibrium population genetic theory.

One particular estimate that has received considerable attention is the estimate of the time of coalescence of mtDNA variation to a single common ancestor (Cann et al. 1987). The estimate of 180,000–360,000 years for this coalescence was recently supported by estimates of 238,000 years and 280,000 years based on sequence variation in the control region (Vigilant et al. 1989; Hasegawa and Horai 1991). These estimates were based on the maximum divergence seen in the sample, and an estimate of the rate of nucleotide divergence in the mtDNA molecular clock. The clock rate was estimated from variation within humans and, as Nei and Livshits (1990) point out, the stochastic error in estimating molecular clock rates is much less when based on interspecific divergence. Using the human–chimpanzee split, Nei and Livshits (1990) arrive at a somewhat greater time for the coalescence of the data of Cann et al. (1987). Perhaps a more serious problem is the rapid growth of the human population. As Krüger and Vogel (1989) demonstrate, the expected number of maternal lineages that remain after a period of time is greatly increased when a population is growing. According to this argument, the estimate of ~200,000 years may be an underestimate. The marked departure from the prediction of the neutral theory makes us skeptical of the accuracy of applying neutral clock theory for estimating the time of coalescence of a sample. If we are able to estimate the rate of growth of the population of prehistoric man, better estimates of the coalescence time might be made.

Conclusions

We have presented further evidence supporting an African origin of human mtDNAs, as first observed by Johnson et al. (1983) and Cann et al. (1987) based upon the deep African root in their mtDNA dendrograms. We also observed the same pattern reported by Cann et al. (1987) where Asian and Caucasian mtDNAs appear to have diverged soon after the founding of the African mtDNA population, as is evidenced by the wide distribution of both Asian and Caucasian mtDNAs throughout the dendrogram. In addition, we find a similar pattern of allele distribution to that reported by Whittam et al. (1986) and Excoffier (1990), which shows a deficit of diversity relative to the number of haplotypes. The departure from neutrality indicated by the Ewens–Watterson and Tajima tests may identify a rapidly growing nonequilibrium population. These data suggest that multiple mitochondrial founder events have occurred in human history, including the founding of the New World (Wallace et al. 1985; Schurr et al. 1990; Torroni et al. 1991), and the founding of Australia and many island groups worldwide. The result is that the human population is not at equilibrium with respect to the gain of variation by mutation and its stochastic loss by drift. The departure from the pattern expected under strict

neutrality and the rapid rate of growth of the human population introduces complications to the problem of estimating coalescence times from these data that have yet to be resolved.

Acknowledgments. We thank Thomas Whittam for providing the program for bootstrapping neighbor-joining trees. This work was supported by grants from the National Institutes of Health and the National Science Foundation to A.G.C. and D.C.W.

References

Anderson S, Bankier AT, Barrel BG, DeBrulin MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Avise JC, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J Mol Evol 20:99–105

Avise JC, Ball RM, Arnold J (1988) Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. Mol Biol Evol 5:331–344

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge J, Hassan K, Chen KH, Wallace DC (1991) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics (in press)

Bonné-Tamir B, Johnson MJ, Natali A, Wallace DC, Cavalli-Sforza LL (1986) Human mitochondrial DNA types in two Israeli populations—a comparative study at the DNA level. Am J Hum Genet 38:341–351

Brega A, Scozzari R, Maccioni L, Iodice C, Wallace DC, Bianco I, Cao A, Santachiara Benerecetti AS (1986a) Mitochondrial DNA polymorphisms in Italy I. Population data from Sardinia and Rome. Ann Hum Genet 50:327–338

Brega A, Gardella R, Semino O, Morpurgo G, Astaldi Ricotti GB, Wallace DC, Santachiara Benerecetti AS (1986b) Genetic studies of the Tharu populations of Nepal: restriction endonuclease polymorphisms of mitochondrial DNA. Am J Hum Genet 39:502–512

Brown W (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci USA 77:3605–3609

Brown W, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967–1971

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

Chakraborty R (1990) Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. Am J Hum Genet 47:87–94

Clark AG (1987) Neutrality tests of highly polymorphic RFLPs. Am J Hum Genet 41:948–956

De Benedictis G, Rose G, Caccio S, Picardi P, Quagliariello C (1989a) Mitochondrial DNA polymorphism in Calabria (southern Italy). Gene Geography 3:33–40

De Benedictis G, Rose G, Passarino G, Quagliariello C (1989b) Restriction fragment length polymorphism of human mitochondrial DNA in a sample population from Apulia (southern Italy). Ann Hum Genet 53:311–318

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

Engels WR (1981) Estimating genetic divergence and genetic variability with restriction endonucleases. Proc Natl Acad Sci USA 78:6329–6333

Excoffier L (1990) Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. J Mol Evol 30:125–139

Excoffier L, Langaney A (1989) Origin and differentiation of human mitochondrial DNA. Am J Hum Genet 44:73–85

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Pop Biol 3:87–112

Harihara S, Hirai M, Omoto K (1986) Mitochondrial DNA polymorphism in Japanese living in Hokkaido. Jpn J Hum Genet 31:73–83

Harihara S, Saitou N, Hirai M, Gojobori T, Park KS, Misawa S, Ellepola SB, Ishida T, Omoto K (1988) Mitochondrial DNA polymorphisms among five Asian populations. Am J Hum Genet 43:134–143

Hasegawa M, Horai S (1991) Time of the deepest root for polymorphism in human mitochondrial DNA. J Mol Evol 32:37–42

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46:828–842

Horai S, Matsunaga E (1986) Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. Hum Genet 72:105–117

Horai S, Gojobori T, Matsunaga E (1986) Distinct clustering of mitochondrial DNA types among Japanese, Caucasians, and Negroes. Jpn J Genet 61:271–275

Johanson D, Taieb M, Coppens Y (1982) Pliocene hominoids from the Hadar Formation, Ethiopia (1973-1977): stratigraphic, chronologic, and paleoenvironmental contexts, with notes on hominoid morphology and systematics. Am J Phys Anthropol 57:373–402

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

Krüger J, Vogel L (1989) The problem of our common mitochondrial mother. Hum Genet 82:308–312

Lewontin RC (1988) On measures of gametic disequilibrium. Genetics 120:849–852

Li W-H, Sadler LA (1991) Low nucleotide diversity in man. Genetics (in press)

Livshits G, Nei M (1990) Relationships between intrapopulational and interpopulational genetic diversity in man. Ann Hum Biol 17:501–513

Miyata T, Hayashida H, Kikuno R, Hasegawa M, Koyabashi M, Koike K (1982) Molecular clock of silent substitutions: at least a six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J Mol Evol 19:28–35

Nei M, Li W-H (1975) The transient distribution of allele frequencies under mutation pressure. Genet Res 28:205–214

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Nei M, Livshits G (1990) Evolutionary relationships of Europeans, Asians, and Africans at the molecular level. In: Takahata N, Crow J (eds) Population biology of genes and molecules. Baifukan, Tokyo, pp 251–266

Nei M, Miller JC (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. Genetics 125:873–879

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. Genetics 105:207–217

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29: 1–10

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

Santachiara Benerecetti AS, Scozzari R, Semino O, Torroni A, Brega A, Wallace DC (1988) Mitochondrial DNA polymorphisms in Italy. II. Molecular analysis of new and rare morphs from Sardinia and Rome. Ann Hum Genet 52: 39–56

Sartoris S, Varetto O, Migone N, Cappello N, Piazza A, Ferrara GB, Ceppellini R (1988) Mitochondrial DNA polymorphism in four Sardinian villages. Ann Hum Genet 52:327–340

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies suggesting a limited number of founders. Am J Hum Genet 46:613–623

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

Semino O, Torroni A, Scozzari R, Brega A, DeBenedictis G, Santachiara Benerecetti AS (1989) Mitochondrial DNA polymorphisms in Italy. III. Population data from Sicily: a possible quantitation of maternal African ancestry. Ann Hum Genet 53:193–202

Shutler R (ed) (1983) Early man in the New World. Sage Publications, Beverly Hills, CA

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics (in press)

Spuhler JN (1988) Evolution of mitochondrial DNA in monkeys, apes, and humans. Year Book Phys Anthropol 31:15–48

Stewart FM (1977) Computer algorithm for obtaining a random set of allele frequencies for a locus in an equilibrium population. Genetics 86:482–483

Stoneking M (1990) Departure of human mitochondrial DNA variation from neutral expectations: an alternative explanation. J Mol Evol 31:343–346

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

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460

Tajima F (1989a) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tajima F (1989b) The effect of change in population size on DNA polymorphism. Genetics 123:597–601

Takahata N, Palumbi SR (1985) Extranuclear differentiation and gene flow in the finite island model. Genetics 109:441–457

Torroni A, Semino O, Rose G, De Benedictis G, Brancati C, Santachiara Benerecetti AS (1990) Mitochondrial DNA polymorphisms in the Albanian population of Calabria (southern Italy). Int J Anthropol 5:97–104

Torroni A, Schurr TG, Yang CC, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowlder WC, Lawrence DN, Weiss KM, Wallace DC (1991) Amerindian mitochondrial DNA analysis reveals pre-Clovis colonization of the New World. Genetics (in press)

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

Vilkki J, Savontaus M, Nikoskelainen EK (1988) Human mitochondrial DNA in Finland. Hum Genet 80:317–321

Wallace DC, Garrison K, Knowler WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. Am J Phys Anthropol 68:149–155

Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AMS, Elsas LJ, Nikoskelainen EK (1988) Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. Science 242:1427–1430

Watterson GA (1977) The homozygosity test of neutrality. Genetics 88:405–417

Weir BS (1990) Genetic data analysis. Sinauer, Sunderland MA

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

Weiss KM (1989) On the number of members of the genus *Homo* who have ever lived, and some evolutionary implications. Hum Biol 56:637–650

Whittam TS, Clark AG, Stoneking M, Cann RL, Wilson AC (1986) Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc Natl Acad Sci USA 83:9611–9615

Received May 5, 1991/Revised June 24, 1991