Spatial and Temporal Distribution of the Neutral Polymorphisms in the Last ZFX Intron: Analysis of the Haplotype Structure and Genealogy

Jadwiga Jaruzelska,*,[1] Ewa Zietkiewicz,[†,1] Mark Batzer,[‡] David E. C. Cole,[§] Jean-Paul Moisan, Rosaria Scozzari,[††] Simon Tavaré[‡‡] and Damian Labuda[†,§§]**

**Institute of Human Genetics, Polish Academy of Sciences, 60-479 Poznan, Poland, [†]Centre de Cancérologie Charles-Bruneau, Centre de Recherche de l'Hôpital Sainte-Justine and [§§]Département de Pédiatrie, Université de Montréal, Montreal, Quebec H3T 1C5, Canada, [‡]Department of Pathology, Stanley S. Scott Cancer Center, Louisiana State University Medical Center, New Orleans, Louisiana 70112, [§]Department of Laboratory Medicine and Pathology, Banting Institute, Toronto, Ontario M5G 1L5, Canada, **Laboratoire de Génétique Moléculaire, Plateau Technique de l'Hôtel-Dieu, Centre Hospitalier Régional et Universitaire, 44035 Nantes, Cedes, France, [††]Dipartimento Genetica e Biologia Molecolare, Universita "La Sapienza," 00185 Rome, Italy and [‡‡]Departments of Biological Sciences and Mathematics, University of Southern California, Los Angeles, California 90089-1113*

Manuscript received November 19, 1998
Accepted for publication March 19, 1999

ABSTRACT

With 10 segregating sites (simple nucleotide polymorphisms) in the last intron (1089 bp) of the ZFX gene we have observed 11 haplotypes in 336 chromosomes representing a worldwide array of 15 human populations. Two haplotypes representing 77% of all chromosomes were distributed almost evenly among four continents. Five of the remaining haplotypes were detected in Africa and 4 others were restricted to Eurasia and the Americas. Using the information about the ancestral state of the segregating positions (inferred from human-great ape comparisons), we applied coalescent analysis to estimate the age of the polymorphisms and the resulting haplotypes. The oldest haplotype, with the ancestral alleles at all the sites, was observed at low frequency only in two groups of African origin. Its estimated age of 740 to 1100 kyr corresponded to the time to the most recent common ancestor. The two most frequent worldwide distributed haplotypes were estimated at 550 to 840 and 260 to 400 kyr, respectively, while the age of the continentally restricted polymorphisms was 120 to 180 kyr and smaller. Comparison of spatial and temporal distribution of the ZFX haplotypes suggests that modern humans diverged from the common ancestral stock in the Middle Paleolithic era. Subsequent range expansion prevented substantial gene flow among continents, separating African groups from populations that colonized Eurasia and the New World.

A widespread effort to document the amount and geographic distribution of genetic variation in our species is motivated by our curiosity about the origins and prehistory of human populations and our interest in the genetic basis of different diseases. These issues are closely related because the genetic bases of human conditions are a function of the present-day structure of human populations, which by itself cannot be understood without knowing the mechanisms shaping the present genetic variation. Indeed, a starting point for studies that aim to explain the role of genetic variation in disease risk is a description of the quality, quantity, and organization of genetic variation within and between human populations. With the advent of new techniques of DNA analysis it is possible to investigate this variability directly. The greatest progress has been achieved in the analysis of mitochondrial DNA (see Jorde *et al.* 1998 for a recent review). Nuclear DNA studies have been focused on microsatellite markers (Bowcock *et al.* 1994; Deka *et al.* 1995; Jorde *et al.* 1995, 1997; Di Rienzo *et al.* 1998; Kimmel *et al.* 1998; Reich and Goldstein 1998) and on insertion polymorphisms resulting from *Alu* retropositions (Batzer *et al.* 1994, 1996; Stoneking *et al.* 1997). Studies of nonrecombinogenic portions of the Y chromosome are just now gaining momentum (see Jobling and Tyler-Smith 1995; Hammer and Zegura 1996 for the reviews; Underhill *et al.* 1997; Poloni *et al.* 1997; Scozzari *et al.* 1997; Hammer *et al.* 1998; Malaspina *et al.* 1998).

Genetic systems differ in their capacity to reveal the information pertinent to current structure and likely population history. Due to differences in effective population size, the time depth of the autosomal diversity is expected to be four times greater and the X-chromosomal diversity three times greater than that of the Y-chromosome or mitochondrial genome. How far one can look back in time depends upon mutation rates. The fast mutation rate of microsatellites makes them suitable for tracing recent evolutionary events, while the presence of recurrent parallel mutations makes homo-

Corresponding author: Damian Labuda, Centre de Recherche, Hôpital Sainte-Justine, 3175 Côte-Sainte Catherine, Montreal, Quebec H3T 1C5 Canada. E-mail: labuda@ere.umontreal.ca

[1]These authors have contributed equally to this work.

plasies frequent at a longer time range. Slower mutation rate and virtually negligible probability of a recurrent mutation in human and great ape lineages characterize classical protein markers, simple nucleotide changes underlying restriction fragment length polymorphisms (RFLPs) and *Alu* insertion polymorphisms. Studies of these variants provided abundant data on polymorphisms dispersed throughout the genome. However, they often suffered from ascertainment bias. Polymorphisms were initially characterized in a small number of samples primarily of European origin, which led to genotyping of extant world populations for polymorphisms known from one human group. Neglecting polymorphisms that were endemic in the compared populations resulted in an inadvertent loss of pertinent phylogenetic information. Although the importance of ascertaining the polymorphisms in an unbiased manner in all groups of populations under study is now widely recognized (Bowcock *et al.* 1991; Mountain and Cavalli-Sforza 1994; Rogers and Jorde 1996), only a few non-Y-chromosome nuclear DNA segments have been systematically surveyed for the presence of sequence polymorphisms in a number of distantly related populations. Variable sites were assessed by direct sequencing in an ~3-kb segment of the β-globin locus (Fullerton *et al.* 1994; Harding *et al.* 1997) and a 10-kb portion of the lipoprotein lipase gene (Clark *et al.* 1998) in 349 and 142 chromosomes, respectively. In another survey of 13 populations from four continents, we screened ~20 independent chromosomes from each of these populations (total of 250) by single-strand conformational polymorphism (SSCP) analysis for polymorphisms in an 8-kb segment of the dystrophin locus (Zietkiewicz *et al.* 1997). Most of these studies indicate the effective population size to be in the range of ~10,000 individuals, which is an order of magnitude lower than that concluded from diversity data at the major histocompatibility complex (Takahata 1993). Mitochondrial and microsatellite data suggest recent demographic expansion; this, however, is not the case of simple nuclear polymorphisms. The highest molecular diversity is consistently seen in sub-Saharan Africans, but it is variably interpreted as being due to the older age of this population, the higher population size in Africa, and/or the greater gene flow on this continent. Because the history of a single locus is not sufficient to make conclusions about the history of populations, more variability data from independent nuclear loci are needed.

Studying DNA diversity on the X chromosome offers an advantage of straightforward and unequivocal determination of haplotypes in hemizygous (male) samples. The X-chromosome-specific zinc finger protein (ZFX) locus became of particular interest because of the virtual absence of DNA diversity in its Y chromosome homologue, the ZFY gene (Dorit *et al.* 1995; Huang *et al.* 1998). Therefore, we have characterized DNA variation in the last introns of ZFX and ZFY in a worldwide sample of human chromosomes (J. Jaruzelska, E. Zietkiewicz and D. Labuda, unpublished results; see also Huang *et al.* 1998). In the present study, the ZFX polymorphisms were combined into haplotypes. The frequencies of ZFX haplotypes and their geographical stratification in 15 globally distributed populations were analyzed. The age of the underlying mutations and the time to the most recent common ancestor (TMRCA) were estimated from the coalescent haplotype tree. The common distribution of the most frequent haplotypes indicates a diversity that characterized the ancestral population of modern humans. Differential patterns of haplotype variability in Africans and non-Africans reflect more recent events. The data are consistent with the divergence of modern human groups from a common ancestral population ~100 kyr ago, followed by range expansion leading to isolation by distance.

MATERIALS AND METHODS

Human DNA samples (nonnominative, characterized only by their origin) represented 15 human populations from four continents. Europeans included Polish (30 unrelated chromosomes), French-Canadians from the Province of Quebec (23), Italians (19), and French (21); Asians were represented by Siberian Nentsi (25), Japanese (21), and Chinese from mainland China (25); Amerindians by Ojibwa (19), Maya (23), and Karitiana from Brazil (16); Africans by Biaka Pygmies from Central Africa Republic (22), M'Buti Pygmies from Congo (22), Rimaibe (19), and Mossi from Burkina Faso (27), and African-Americans from Michigan (24).

DNA variation within 1089 bp of the last intron of the ZFX gene on Xp21.3 (Shimmin *et al.* 1993) in the above sample of 336 worldwide chromosomes was investigated by SSCP/heteroduplex approach, as described earlier (Zietkiewicz *et al.* 1992a, 1997). Briefly, the analyzed segment was amplified in three overlapping fragments using the following primer pairs: (i) 5′-CGGCAGACTGGCTAAACAA and 5′-ATGCTTA TAACATATTTGAGGG (349 bp, annealing temperature 55°); (ii) 5′-AGGACATGGCTGAAACAT and 5′-GTGACAAAAAT TTCCACTG (306 bp, 55°); (iii) 5′-GCTGTAAGTTAACGT AAGT and 5′-CTGTTCCAGTTTCTTTGCG (666 bp, 50°); amplification of the first two fragments was done in the presence of 2% formamide. Prior to electrophoresis in a 6% polyacrylamide gel (acrylamide to bisacrylamide ratio 50:1 and 30:1), the 666-bp PCR product was digested with *Eco*RI into fragments of 319 and 347 bp. Mutations underlying gel mobility variants were identified by dideoxy sequencing.

The mutation rate in the 1089-bp ZFX segment, estimated from the human-chimpanzee and human-orangutan comparisons, was 2.55 ($\pm$0.44) $\times 10^{-5}$ per generation assuming a generation time of 20 yr. The ancestral state of the segregating sites that were due to nucleotide substitutions was inferred by comparison with the orthologous positions in chimpanzee and orangutan DNA. Given the mutation rate in the order of 10^{-9} per nucleotide position per year, there is a very small chance that a recurrent event has taken place since the divergence of these species from human lineage 5 and 12 mya. In other words, the probability is negligible that a human allele identical with the corresponding site in chimpanzee and orangutan results from a back mutation or that three identical mutations occurred independently after the separation of these lineages. The identity by state of a human allele and the corresponding positions in great apes is in practice tantamount to their identity by descent. Thus, the human allele identical by state with the chimpanzee and orangutan orthologues was considered ancestral, whereas the other allele of the same polymorphic site was considered new.

The derivation of haplotypes from the genotypes was straightforward in hemizygous males (76 chromosomes), in homozygous female samples (112 chromosomes), and those heterozygous at a single position (92 chromosomes). In multiple heterozygotes (female samples having distinct alleles at more than one polymorphic site), the genotype data were easily resolved into haplotypes when two polymorphisms were present within a single SSCP-analyzed fragment (21 individuals). The remaining haplotypes (7 chromosome pairs) were inferred assuming most likely combinations of the unequivocally resolved haplotypes and taking into account their distribution among populations.

A number of summary statistics (haplotype diversity G, mean number of pairwise differences Π, nucleotide diversity π, and different estimates of θ, the scaled mutation parameter) were compared using the Arlequin package v.1.1. (Schneider *et al.* 1997). Arlequin was also used to perform neutrality tests according to Ewens (1972), Watterson (1978), Tajima (1989), and Chakraborty (1990), as well as to assess population genetic structure by analysis of molecular variance (AMOVA; Excoffier *et al.* 1992). For detailed description of the parameters and tests please refer to the Arlequin help manual (http://anthropologie.unige.ch/arlequin/).

The maximum-likelihood estimate of θ using the full information in the sequence data set (a haplotype tree, including the information about the ancestral state) was found using a computational method proposed by Griffiths and Tavaré (1994), implemented in the program *genetree* and available from the mathematical genetics group web page at http://stats.ox.ac.uk/. *genetree* was also used to find the distribution of the TMRCA and the distribution of the ages of the underlying mutations, conditional on the gene tree, using the methods described in Griffiths and Tavaré (1999); alternatively, the average age of the mutations was computed from the frequencies of their new (nonancestral) alleles (Kimura and Ohta 1973; Griffiths and Tavaré 1998; Zietkiewicz *et al.* 1998).

Maximum-likelihood estimates of θ obtained with a constant population size model were compared with those using a model of exponential growth. In the exponential growth model there is a decline of the population size backward in time from a current size $N_0 = \theta/3\mu$ (22,000 for the maximum-likelihood θ estimate of 1.66), such that the population size at time t ago is $N_t = N_0e^{-\beta t}$. Using the *genetree* program with the expansion rate parameter β of 0.1, 0.2, and 0.3 and the generating θ value of 1.66, we obtained no improvement in the log-likelihood compared with the constant population size model (*i.e.*, no evidence for exponential population size expansion was found). The arbitrary β values selected were lower than these suggested in the *genetree* manual (0.5 and above) to keep them close to more realistic range.

RESULTS

Ten single-nucleotide polymorphisms within the 1089-bp sequence of the last ZFX intron were previously ascertained in a sample of 336 worldwide-distributed chromosomes (J. Jaruzelska, E. Zietkiewicz and D. Labuda, unpublished results). The ancestral and the new allele at each site were inferred by comparison with the orthologous great ape sequences (see materials and methods). The new allele frequencies in the world population varied from 0.3 to 99.1% (Table 1). The average heterozygosity per polymorphic site was 8.9%, corresponding to the nucleotide diversity π (or the mean number of pairwise differences per nucleotide position) of 0.082%. The latter values ranged from 0.06% in Europeans to 0.10% in Asians, while African and Amerindian groups were characterized by π of 0.08% (Table 2). The number of polymorphisms in particular populations ranged from two to four (Figure 1B). One polymorphic site (700) was shared by all the groups; at another (1093), the new allele was fixed in all non-African populations. Three sites (514, 632, and 757) had the new allele present only in non-Africans, three others (203, 400, and 491) in Africans, while at two sites (502 and 716) the new allele had a patchy distribution, presumably resulting from admixture (see discussion). Thus, the distribution of the polymorphisms distinguished Africans from other continental groups.

Nucleotide diversity parameters reported here could be underestimated because of limitations of the SSCP method. On the other hand, use of various gel conditions, analyzing partially overlapping fragments, as well as screening a large sample of chromosomes largely improve the detection rate of DNA polymorphisms. We estimate the efficiency of the SSCP/heteroduplex approach at well above 80%, especially given our long experience in using this technique (*e.g.*, Zietkiewicz *et al.* 1992a,b, 1997; Akalin *et al.* 1994). This is a conservative estimate in light of the success rate of up to 98% reported by other authors (Hayashi 1991; Ravnik-Glavac *et al.* 1994; Welsh *et al.* 1997). Moreover, Huang *et al.* (1998) sequenced the same ZFX segment in a sample of 29 worldwide chromosomes revealing only the ubiquitous site 700. In this context the probability that a frequent polymorphism was missed in our analysis is very low.

The 10 polymorphic positions described above cosegregated as 11 haplotype variants (Table 1). The phylogenetic relations between these variants are represented by a tree (Figure 1A) that summarizes the underlying history of mutations. The tree is rooted at the ancestral haplotype (H0), in which the alleles at all the polymorphic positions are identical by state with the orthologous positions in nonhuman primates. Two most common haplotypes, H1 and H2 are separated from the ancestral one by one and two mutational steps, respectively, and from each other by a single mutation. The only exception is the single copy of H5/1, which could be regarded as a recombinant of H5 and H1. Given the mutation rate in the order of 10^{-9} per nucleotide per year, the alternative routes of generating H5/1 through reverse/recurrent mutation at position 700 or at position 502 are much less likely.

The worldwide distribution of haplotype frequencies (Table 1), with H1 and H2 representing 77% of chromosomes, fits very well the infinite-allele model according

TABLE 1

Haplotypes generated from 10 polymorphic sites in the last intron of the ZFX gene

	Sequence position[a]										Haplotype	
Haplotype variant	203 T	400 C	491 G	502 A	514 G	632 C	700 A	716 C	757 G	1093 T	Counts (n = 336)	Frequency (%)
H0											3	0.89
H1										c	75	22.32
H2							g			c	183	54.46
H3						g				c	28	8.33
H4							g		t	c	19	5.65
H5				g			g			c	16	4.76
H6							g	t		c	4	1.19
H7	c						g			c	4	1.19
H8		g	a							c	2	0.60
H9					a	g				c	1	0.30
H5/1[b]				g						c	1[b]	0.30
New allele freqency (%)	1.19	0.60	0.60	5.06	0.30	8.63	67.26	1.19	5.65	99.11		

[a] Numbering is that of GenBank sequence HSZFXIN1-X58925; proposed ancestral and new alleles are denoted by upper- and lowercase, respectively.

[b] Excluded from statistical analysis (see the text).

to Ewens' sampling theory (Ewens 1972). The contribution of these two frequent haplotypes was high in all local populations with the exception of the Chinese, in which H1 was not observed. Except for the rare ancestral haplotype H0 (found at low frequency in Mossi and African-Americans), all others were derived from H1 and H2 (Figure 1A) and represented young variants based on coalescent analysis of both the full data and the segregating sites separately (see below). The geographical distribution of these young haplotypes was not uniform (Figure 1B), resembling that of the underlying polymorphisms that differentiated between Sub-Saharan Africans and populations from other continents. Haplotype H9 was restricted to the Chinese population and haplotypes H7 and H8 to Pygmies. The only haplotype resulting from a recombination, H5/1, was found in a single copy in Mossi. A relatively rare haplotype H6 was present in African-Americans in addition to populations in Europe and Siberia, while a single copy of the otherwise African haplotype H5 was found in Maya. The patchy distribution of these two haplotypes presumably represents recent admixture events (see discussion).

In the virtual absence of recombinations, the haplotype diversity within the analyzed ZFX segment results solely from mutation. Leaving aside the H5/1 recombinant, the remaining haplotypes represent sequences that can be treated formally as mitochondrial DNA and/or haplotypes in the nonrecombining portion on the Y chromosome. The lower overall rate of evolution and lower diversity of the ZFX intron, as compared to other genomic segments, suggests that selection may have acted on this sequence (J. Jaruzelska, E. Zietkiewicz and D. Labuda, unpublished results). However, this effect appears too small to influence the population variability of ZFX, which, according to Ewens-Watterson and Tajima tests (not shown and Table 2, respectively), does not differ from that of a neutral locus.

We used a coalescent model to infer the time scale of the origin and evolution of polymorphic variation within the ZFX segment. The distribution of coalescence times and the ages of mutations depend on the mutation rate θ: $\theta = 2N\mu_g$, where N is the number of chromosomes in the population (for the X-linked locus, N corresponds to $1.5N_e$, assuming equal contribution of both sexes to N_e), and μ_g is the mutation rate per DNA segment per generation. The N_e estimated from θ_Π, using $\mu_g = 2.55 \times 10^{-5}$ per DNA segment per generation (J. Jaruzelska, E. Zietkiewicz and D. Labuda, unpublished results) is ~12,000 (*i.e.*, ~18,000 chromosomes). The corresponding N_e estimate (of ~20,000 individuals) based upon θ_S is almost twice as large (Table 2). The maximum-likelihood estimate of θ conditional on the ZFX gene tree (θ_{ML}) is 1.66 ($\pm$0.43), and the corresponding N_e of ~22,000 is close to that based on θ_S. In what follows, we compare the distribution of mutational ages using different estimates of θ (and hence N_e). In principle it is possible to allow for the effect of variability in θ when inferring the ages of mutations and coalescence times (*cf.*, Tavaré *et al.* 1997), but here we treat the values of θ as known.

The mutation tree shown in Figure 2 summarizes the results obtained by applying the approach of Griffiths and Tavaré (1999) to infer mutational ages. This tree is rooted because the ancestral state at each site is known. The distribution of haplotypes among the continental groups is indicated below the tree. Using the

TABLE 2

Interpretation of molecular diversity at the ZFX locus

	World	African	Non-African	Asian	Amerindian	European
Data						
Number of chromosomes (n):	335	113	222	71	58	93
Number of polymorphic sites (S):	10	7	7	5	5	4
Number of haplotypes (k):	10[a]	7[a]	7	6	5	5
Population statistics						
Haplotype diversity, G (SD)	0.64 (0.02)	0.63 (0.04)	0.63 (0.03)	0.63 (0.04)	0.64 (0.04)	0.57 (0.04)
Mean no. pairwise differences, Π (SD)	0.89 (0.62)	0.84 (0.60)	0.89 (0.62)	1.12 (0.74)	0.82 (0.60)	0.67 (0.52)
Nucleotide diversity (%), $\pi = \Pi/L$ (SD)	0.082 (0.045)	0.077 (0.043)	0.082 (0.049)	0.103 (0.061)	0.075 (0.049)	0.061 (0.041)
Expected θ (based on Π), $\theta_\Pi = \Pi$ (SD)	0.89 (0.69)	0.84 (0.67)	0.89 (0.69)	1.12 (0.82)	0.82 (0.66)	0.67 (0.57)
Expected θ (based on S), $\theta_S = S/\Sigma^{n-1}(1/i)$ (SD)	1.56 (0.57)	1.32 (0.57)	1.17 (0.45)	1.03 (0.52)	1.08 (0.47)	0.78 (0.43)
Tajima's estimator, D	−0.95	−0.83	−0.51	+0.18	−0.57	−0.30
Expected θ (maximum likelihood, θ_{ML} (SD)	1.66 (0.43)	1.36 (0.62)	1.06 (0.49)	1.08 (0.49)	0.91 (0.47)	0.81 (0.46)
Expected number haplotypes, E(k)	8.0	6.4	7.3	5.7	5.7	5.1
Effective population size[b]						
$N_{e\Pi}$ (SD)	11,600 (9,000)	11,000 (8,800)	11,600 (9,000)	14,600 (10,700)	10,700 (8,800)	8,800 (7,500)
N_{eS} (SD)	20,400 (7,500)	17,300 (7,500)	13,100 (5,900)	13,500 (6,800)	11,200 (6,100)	10,200 (5,600)
N_{eML} (SD)	21,700 (5,600)	17,800 (8,100)	13,900 (6,400)	14,100 (6,400)	11,900 (6,100)	10,600 (6,000)
Coalescent time[c], T_{tree} (SD)	2.52 (0.90)	2.70 (1.02)	2.57 (1.05)	2.62 (1.05)	2.61 (1.14)	2.79 (1.22)

[a] Excluding variant H5/1.
[b] Mutation rate, 2.55×10^{-5} per 1089-bp segment per 20-yr generation.
[c] In N_e generation units, using μ_{ML} (2×10^6 simulation runs).

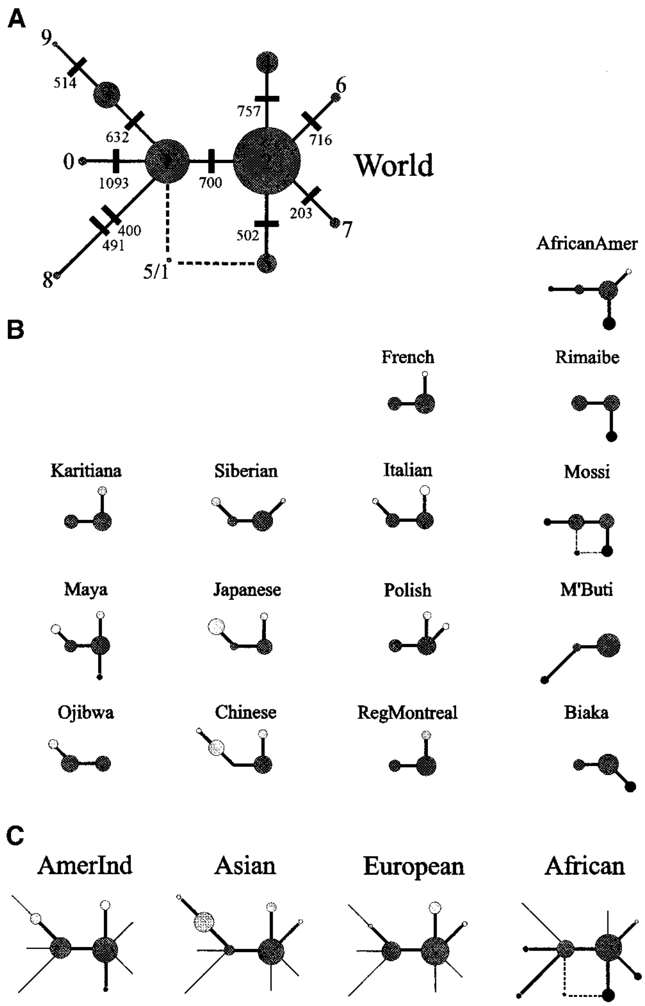

Figure 1.—Maximum parsimony network of ZFX haplotypes. Circle sizes in the tree are proportional to the haplotype frequencies in the world (A), individual populations (B), or continents (C). In A, small numbers next to the lines connecting individual haplotypes indicate the underlying mutation sites. Note that most of the haplotypes can be derived from each other by a single change; the order of two mutations separating haplotypes H1 and H8 is unknown; H5/1 is assumed to result from the recombination between H1 and H5 (broken line).

maximum-likelihood estimate of θ (θ_{ML}), assuming neutrality, random mating, and a constant population size, the expected height of the tree (TMRCA) is estimated at 2.5, and the standard deviation of the height of the tree is 0.9, time being measured in units of N_e generations (Table 2). The corresponding expected TMRCA value using the estimate θ_{Π} of θ is 3.1 N_e generations. Of note, we found no significant evidence of exponential population expansion (see materials and methods). Conversion of the age estimates above into years (Figure 2) depends on the effective population size N_e.

In Figure 2 we report the expected ages of the mutations obtained using different values of θ. The values shown alongside the tree indicate the expected ages in N_e generation units, obtained using the θ_{ML} estimate and the θ_{Π} estimate (the corresponding values in kyr, using 20 yr per generation, are between parentheses). The expected TMRCA values of 2.5 and 3.1 N_e generation units correspond to 1100 and 740 kyr, respectively. TMRCA also represents the expected age of the ancestral haplotype H0. Two globally distributed haplotypes H1 and H2, which arose as the result of mutation at sites 1093 and 700, are 550 to 840 kyr and 200 to 400 kyr old, respectively. The estimated ages of the remaining haplotypes, *i.e.*, those with the restricted or patchy geographical distribution, vary from 5–9 to 120–180 kyr (Figure 2). The average age of each polymorphism estimated from the full data is very similar to the average age computed from just the frequencies of its new (nonancestral) alleles (Kimura and Ohta 1973; Griffiths and Tavaré 1998; Zietkiewicz *et al.* 1998). Figure 3 compares the average ages of polymorphisms (in N_e generation units) and associated 5 and 95% percentiles obtained from the new allele frequencies with those obtained from the full data using the estimates θ_{ML} and θ_{Π} of θ. The expected age of the recombinant H5/1 is the same as that of haplotype H9.

Table 2 summarizes the population parameters estimated for the world, continents, and all non-Africans together. The estimates of population parameters based on π are generally (except for Asians) lower than these based on *S*, but this difference is not statistically significant. This holds, whether we consider the whole sample or its subsets as representative of the world population. Thus, these population groups look quantitatively similar in spite of the qualitative differences in distribution of young haplotypes. The respective parameters calculated for each of the 15 populations analyzed (not shown) are within the range indicated by the standard deviation values for the larger, continental groups. It is noteworthy that the coalescent time (Table 2) for the world and for the continental samples is almost identical, suggesting again that these groups spread recently from a common ancestral stock and share most of their earlier genetic history (see also Di Rienzo *et al.* 1998).

To assess the level of population structure, we estimated the F_{ST} parameter, describing the contribution of variance among populations to total variance. The F_{ST} values were computed using an AMOVA test (Excoffier *et al.* 1992) based on the pairwise distances between haplotypes. The significance of the components of variance was assessed by comparing observed levels with the distribution of 16,000 values obtained by randomization. F_{ST} calculated for the world population, considered to be composed of 15 subpopulations, was 0.081, indicating that only 8.1% of the total variance resulted from the differences among populations and ~92% was due to the variance within populations. When the world was divided into four continental populations or only into Africans and non-Africans, F_{ST} values were 0.067 and 0.048, respectively. The F_{ST} value for Africans was 0.064 and for the cluster combining the remaining groups,

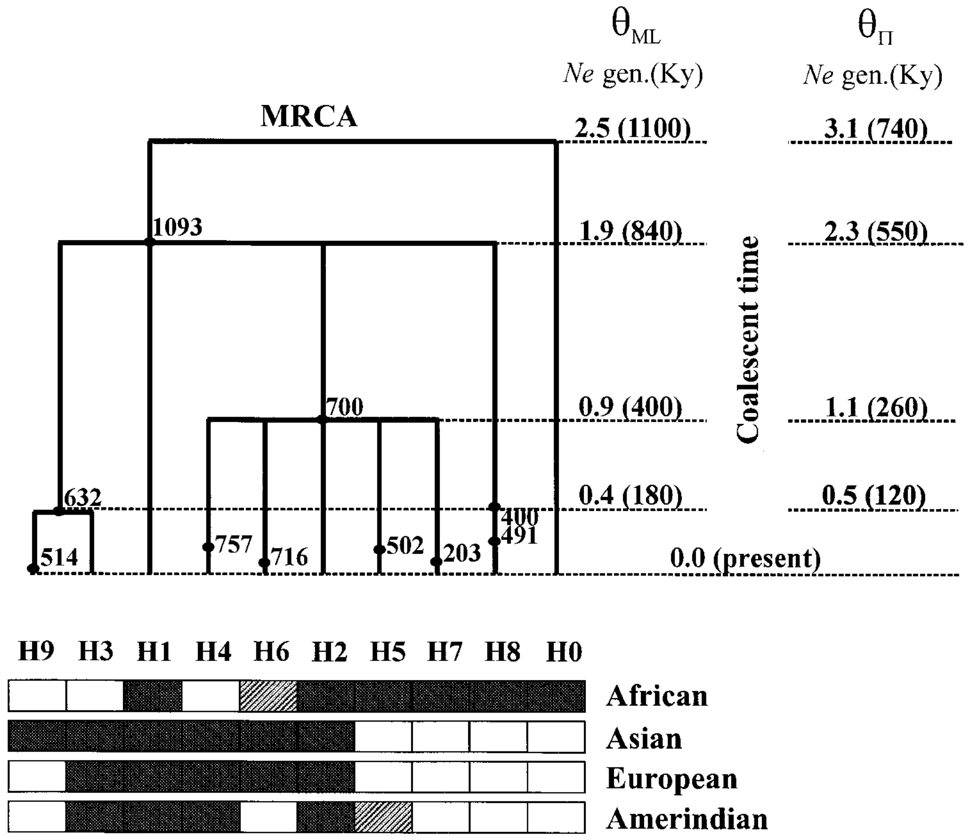

Figure 2.—Gene tree of the ZFX polymorphisms in the sample of 335 worldwide chromosomes. Mutations are indicated on the branches at the distances proportional to their estimated average age (2×10^7 simulation runs), using θ_{ML} (maximum-likelihood estimate of θ) or θ_{Π}. The order of mutations at positions 400 and 491 cannot be resolved. The values at the right correspond to the coalescent times in N_e generations or, between parentheses, in years (with a generation time of 20 yr and N_e conditional on θ_{ML} or θ_{Π}, *i.e.*, N_e = 22,000 or 12,000, respectively). The corresponding haplotypes are indicated below the tree; the presence of a specific haplotype in the continental population is indicated by a filled box whereas a hatched box indicates the presumed admixture (see text).

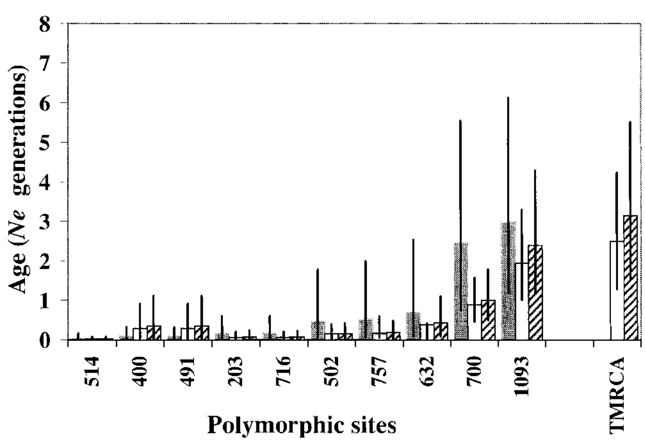

Figure 3.—Comparison of the different estimates of the average ages of polymorphisms and of the time to the most recent common ancestor (TMRCA). Polymorphisms are ordered according to increasing frequency of the new allele. Age estimates obtained from the new allele frequencies (Kimura and Ohta 1973) are represented by the filled bars and coalescent estimates from the full data conditional on θ_{ML} and θ_{Π} by the empty and hatched bars, respectively. The range between the 5th and 95th percentiles is indicated on the bars. Percentiles for estimates based on the full data were obtained from the *genetree* program; those for the ages estimated from the new allele frequencies were calculated as in Griffiths and Tavaré (1998).

0.062. The results for Europe, Asia, and the Americas were nonsignificant, indicating the absence of population structure in these groups, which is consistent with the conclusions based on other summary statistics above.

DISCUSSION

The geographical distribution of the DNA diversity in the 1.1-kb ZFX segment is not uniform. Two out of 11 haplotypes (H1 and H2) are ubiquitous and represent 77% of the worldwide sample of 336 chromosomes. Similar prevalence of common haplotypes was also found at other genomic loci. At the 2.67-kb β-globin segment (Harding *et al.* 1997), 5 common haplotypes contributed ∼77% to the worldwide diversity in a sample of 349 chromosomes, and at the 8-kb-long *dys44* segment, 6 frequent haplotypes represented ∼76% of the worldwide sample of 973 chromosomes (our unpublished data). The remaining 9 haplotypes at the ZFX locus have patchy or continentally restricted distribution. There is good reason to believe that the occurrence of haplotype H6 in African-Americans is due to European admixture in this population (Chakraborty *et al.* 1992), although this will have to be confirmed by studies of a greater number of local populations and/or chromosomes. Similarly, haplotype H5 in Maya may be the result of an African-American admixture. The observation of Seielstad *et al.* (1994), who found evidence of African admixture in the same Mayan samples, reinforces our conclusion; African-American admixture was also observed among other Amerindian populations by A. Ruiz-Linares (personal communication). We can therefore tentatively assign 5 of the haplotypes (H0, H5, H7, H8, and H5/1) to populations from Sub-Saharan Africa. Haplotypes H3, H4, H6, and H9 appear specific to the group of non-African populations. Quantitatively, the haplotype diversity (*G* values in Table 2) is similar in Africa and the remaining continents. Summary statistics based on the mean number of pairwise differences (θ_{Π}) indicate the greatest diversity in Asia. However, similar to what was observed at the β-globin locus (Harding *et al.* 1997), the estimates based on the number of segregating sites as well as maximum-likelihood estimates from the whole data set show that the diversity is greatest in Africa (θ_S and θ_{MLE}).

All continentally restricted haplotypes except for the ancestral H0 are estimated to have arisen between 120 and 180 kya and the present, on the basis of a coalescent analysis of the full data, and between 160 kya and the present using the mutation ages estimated from the new allele frequency (Figures 2 and 3). In contrast, variants H1 and H2 estimated at 550 to 840 and 200 to 400 kyr, respectively, are found at high proportions in all populations except from Chinese (Figure 2). The absence of H1 from Chinese, where it is replaced by the much younger haplotype H3, may reflect founder effect associated with peopling of Asia and its subregions (*e.g.*, Wang *et al.* 1991). The presence of H1 in two other Asian populations, the paucity of H3 in Europeans, as well as the tree structure in Figure 1A (indicating that H1 "missing" in Chinese is an intermediary form between H0 and both H2 and H3) preclude the interpretation that the two most common haplotypes in the Chinese reflect ancestral haplotype frequencies. Rather, it is the presence of H1 and H2 and their frequency profile that are similar across the continents (Figure 1C), together with the oldest variant H0 that is still detectable in Africa, which can be regarded as the ancestral configuration.

Considering the continents separately, one obtains the same coalescent time (Table 2) as for the pooled sample representing the world. A similar observation was made with microsatellite data (Di Rienzo *et al.* 1998), which are consistent with the shared roots of the contributing populations throughout the substantial period of their earlier genetic history. Worldwide similarity of the frequency profiles of the most common haplotypes could reflect extensive gene flow between the continents. An AMOVA test performed to provide estimates of variance components and of the *F*-statistics representing the correlation of haplotypes at different levels of hierarchical groupings revealed low variation among constituting populations or groups of populations as compared to the variation within populations. This is similar to the observation at the dystrophin locus (Zietkiewicz *et al.* 1997) and could again indicate a high degree of gene flow connecting the populations as well as continents. However, if such intercontinental genetic exchange existed during the last 100 to 200 kyr, why would a number of distinct young haplotypes persist in isolation, either in Africa or outside, in Eurasia, and the Americas? Likewise, why would the oldest H0 occur at detectable frequency only among Sub-Saharan Africans? We argue that if substantial genetic exchange occurred, it was at the time of ancestral population divergence, preceding the expansion of modern groups. Thereafter, gene flow was restricted by geographic isolation and physical distance between populations colonizing new continents.

With reduced genetic drift, population growth could be conducive to maintaining the frequency profile of polymorphisms characteristic of the ancestral stock from which expanding populations diverged. The excess of the number of segregating sites over the nucleotide diversity π (*i.e.*, the negative value of Tajima's *D* parameter) and the excess of the observed over expected number of haplotypes, E(*k*), according to Ewens (1972) sampling (see World, Table 2) may suggest that a weak trace of the recent demographic growth is reflected in our sample. However, selection could also be evoked to explain these observations (J. Jaruzelska, E. Zietkiewicz and D. Labuda, unpublished results). Moreover, other non-Y-chromosome nuclear loci studied in a similar way provide no evidence for population expansion. Hey (1997) indicated that most nuclear genes show slightly positive skew in frequency spectrum, and studies of β-globin, dystrophin, and LPL loci (Clark *et al.* 1998; Harding *et al.* 1997; Zietkiewicz *et al.* 1997, 1998) reported positive *D* associated with the non-Y-chromosome nuclear diversity. On the other hand, these observations do not allow firm conclusions. The positive *D* values reported for β-globin, dystrophin, and LPL, as well as the negative ones observed here for ZFX were not significantly different from zero. This is further illustrated by the fact that the *D* values calculated here for individual populations were both positive and negative (not shown). The coalescent estimates from β-globin data (Harding *et al.* 1997) did not perform better under a model of expanding population than under a constant population size model, which was also the case for ZFX (see materials and methods). Thus, if there was any population size expansion, it was not detected by the exponential model or it was probably too recent to be detectable in the observed patterns of ZFX and/or β-globin diversity. On the other hand, a model that assumes expansion from the Middle/Upper Paleolithic up to the present is suggested by paleontological studies (Lahr and Foley 1998). A recent population growth is also consistent with genetic studies of mitochondrial DNA or nuclear microsatellites (Di Rienzo and Wilson 1991; Slatkin and Hudson 1991;

Rogers and Harpending 1992; Jorde *et al.* 1997; Di Rienzo *et al.* 1998; Kimmel *et al.* 1998). These genetic systems, because of the higher mutation rates, are more suitable for detecting expansion at relatively short evolutionary time scales in the range of 100 kyr. In conclusion, most nuclear loci characterized for simple nucleotide polymorphisms imply a different past in comparison to mitochondrial or microsatellite studies (Jorde *et al.* 1995; Hey 1997).

Interestingly, the answer to these discrepancies may lie in the analysis of extended haplotypes, where nuclear diversity generated by rare mutational events would be enhanced by recombinational events, giving insight into more recent population histories, and corresponding to the time frame of mitochondrial and microsatellite mutations. It should be emphasized that the virtual absence of recombinants in ZFX does not necessarily indicate lack of recombination in this region. Assuming the genomic average of 44 crossovers per female meiosis per generation (Broman *et al.* 1998) and taking into account that only two-thirds of the X chromosome population participates in female meioses, an average genomic crossover rate is 10^{-8} per base pair per generation, *i.e.*, $\sim 10^{-5}$ per 1089 bp in the ZFX segment analyzed. Yet, informative chromosomal pairs that would lead to detectable recombinants represent only 3% of all ZFX haplotype combinations (in other words, 97% of recombinations between "noninformative" ZFX segments remain undetected). Thus, an apparent recombination rate in ZFX is 3×10^{-7}, a hundred times less than the mutation rate of 2.6×10^{-5} in the same segment.

The ZFX data truncation (a single chromosome representing the recombined haplotype 5/1 was excluded) introduced for the sake of the coalescent analysis was minimal. In some loci, however, the contribution of crossovers to haplotype diversity may be non-negligible. Such an effect is very pronounced in the recently studied LPL locus (Clark *et al.* 1998; Nickerson *et al.* 1998), as well as in the *dys44* locus (our unpublished data). In this context the good fit between the mutational age estimates based on the full ZFX data and those using just the frequencies of the new allele (Figure 3) is noteworthy. Similar concordance was shown for the β-globin data of Harding *et al.* (1997; see Figure 4 in Zietkiewicz *et al.* 1998). Although the latter age estimate has a larger variance, it can be calculated for each mutation in the haplotype without knowing the underlying tree.

The genetic diversity measures, pointing to the population size of $\sim$10,000 to 20,000, presumably reflect the size of the ancestral population or speciation bottleneck rather than the harmonic mean of the effective sizes of the world population from Middle Pleistocene to the present. Speciation bottleneck and/or divergence and range expansion from a relatively small ancestral population would have a greater impact on the number of segregating sites than on the nucleotide diversity. If such an event had taken place relatively recently, it would have left no time (depending also upon the size of the expanding populations) for the system to fully recover and reach the equilibrium. As a result, we may observe either the shortage of segregating sites, as in non-African populations at the *dys44* locus (Zietkiewicz *et al.* 1998) or a number of new sites that contribute to θ_S but very little to nucleotide diversity, as in Africans at *dys44* (Zietkiewicz *et al.* 1998) and in this study. The net effect on the world diversity reflected by π and S may be that the data fit well and can be thus analyzed in terms of a constant population size model.

Our interpretation of the ZFX data is supported by similar results obtained for the β-globin (Harding *et al.* 1997) and *dys44* (Zietkiewicz *et al.* 1998) loci. Taken together with mitochondrial, microsatellite, and paleontological evidence, these data suggest the divergence of modern human groups from a common ancestral population $\sim$100 kyr ago (taking into account that population trees are usually shorter than these of the underlying genes) followed by range expansion leading to the isolation by distance. What is the geographic origin of the ancestral population? Greater genetic diversity in Africans has frequently been taken as evidence of this population's antiquity (Nei *et al.* 1975; Batzer *et al.* 1994; Horai *et al.* 1995; Tishkoff *et al.* 1996). The fact that genetic diversity outside of Africa can be interpreted as being a subset of the African set has been seen as even more compelling evidence for the out-of-Africa origin (Stoneking *et al.* 1997; Jorde *et al.* 1998). However, in the *dys44* study, the excess of the African diversity was shown to have accumulated only following the divergence of modern humans (Zietkiewicz *et al.* 1998). Greater diversity in one group may be due to its better representation in the worldwide sample analyzed (Bowcock *et al.* 1991; Mountain and Cavalli-Sforza 1994; Rogers and Jorde 1996) or to the way we group or define the populations. It may also reflect greater long-term effective population size following expansion (Relethford 1995; Relethford and Harpending 1995; Rogers and Jorde 1995) or particular history of founder effects, extinctions, and expansions of local populations examined (Takahata 1994).

Assuming that the ancestral population diverged and underwent range expansion in the Middle/Upper Paleolithic, old haplotypes could have been carried away from the place of origin and preferentially preserved in populations that colonized new areas conducive to relatively uninterrupted growth and demographic prosperity. Another model, which can be described as a single-species model (see Foley 1998), favors extensive gene flow among distant hominid populations over the whole Paleolithic and thus the contribution of dispersed ancient populations to modern human origins (reviewed in Templeton 1997). The argument becomes circular if assuming gene flow we subsequently argue that the presence of a haplotype in a geographically limited sample points to its ancestral place of origin there. Thus, a demonstration of an archaic lineage on one or another continent does not prove anything under any of the above scenarios. With migration and range expansion playing an important role in all the underlying models, the geographic location of the modern human origin cannot be inferred from the genetic evidence alone. Paleontological findings and paleoclimatic data (Tattersall 1995; Lahr and Foley 1998) place the origin of the ancestral population in Africa or the Middle East, between the Klasies River and Qafzeh. Our data neither contradict nor prove this scenario. We believe, however, that integrating data from a variety of genetic systems and interpreting them in the context of the outside evidence will be more illuminating in this respect. The need for more extensive sampling of human populations from different geographic areas and of different loci from the nuclear genome appears more than evident, and this study is one of the contributions to fulfill this task.

Thanks to Dominik Gehl for his assistance in computing and to Raffaela Ballarano for typing the manuscript. This work was supported by National Science Foundation grants DMS 95-04393 to S.T. and SBR-9610147 to M.B., by the Canadian Genome Analysis and Technology Program, and by the Medical Research Council of Canada grant to D.L.

LITERATURE CITED

Akalin, N., E. Zietkiewicz, W. Makalowski and D. Labuda, 1994 Are CpG sites mutation hot spots in the dystrophin gene? Hum. Mol. Genet. **3:** 1425–1426.

Batzer, M. A., M. Stoneking, M. Alegria-Hartman, H. Bazan, D. H. Kass *et al.*, 1994 African origin of human-specific polymorphic Alu insertions. Proc. Natl. Acad. Sci. USA **91:** 12288–12292.

Batzer, M. A., S. S. Arcot, J. W. Phinney, M. Alegria-Hartman, D. H. Kass *et al.*, 1996 Genetic variation of recent Alu insertions in human populations. J. Mol. Evol. **42:** 22–29.

Bowcock, A. M., J. Kidd, J. L. Mountain, J. M. Hebert, L. Carotenuto *et al.*, 1991 Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc. Natl. Acad. Sci. USA **88:** 839–843.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd *et al.*, 1994 High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455–457.

Broman, K. W., J. C. Murray, V. C. Sheffield, R. L. White and J. L. Weber, 1998 Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am. J. Hum. Genet. **63:** 861–869.

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

Chakraborty, R., M. I. Kamboh, M. Nwankwo and R. E. Ferrell, 1992 Caucasian genes in American Blacks: new data. Am. J. Hum. Genet. **50:** 145–155.

Clark, A., K. M. Weiss, D. A. Nickerson, S. L. Taylor, A. Buchanan *et al.*, 1998 Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am. J. Hum. Genet. **63:** 595–612.

Deka, R., L. Jin, M. D. Shriver, L. M. Yu, S. Decroo *et al.*, 1995 Population genetics of dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms in world populations. Am. J. Hum. Genet. **56:** 461–474.

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

Di Rienzo, A., P. Donnelly, C. Toomajian, B. Sisk, A. Hill *et al.*, 1998 Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics **148:** 1269–1284.

Dorit, R. L., H. Akashi and W. Gilbert, 1995 Absence of polymorphism at the ZFY locus on the human Y chromosome. Science **268:** 1183–1185.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3:** 87–112.

Excoffier, L., P. Smouse and J. Quattro, 1992 Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondria DNA restriction data. Genetics **131:** 479–491.

Foley, R., 1998 Genes, evolution and diversity: yet another look at the problem of modern human origins. Evol. Anthropol. **6:** 191–193.

Fullerton, S. M., R. M. Harding, A. J. Boyce and J. B. Clegg, 1994 Molecular and population genetic analysis of allelic sequence diversity at the human β-globin locus. Proc. Natl. Acad. Sci. USA **91:** 1805–1809.

Griffiths, R. C., and S. Tavaré, 1994 Ancestral inference in population genetics. Stat. Sci. **9:** 307–319.

Griffiths, R. C., and S. Tavaré, 1998 The age of a mutation in a general coalescent tree. Stochastic Models **14:** 273–295.

Griffiths, R. C., and S. Tavaré, 1999 The ages of mutations in gene trees. Ann. Appl. Prob. (in press).

Hammer, M. F., and S. L. Zegura, 1996 The role of the Y chromosome in human evolutionary studies. Evol. Anthropol. **5:** 116–134.

Hammer, M. F., T. Karafet, A. Rasanayagam, E. T. Wood, T. K. Altheide *et al.*, 1998 Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol. Biol. Evol. **15(4):** 427–441.

Harding, R. M., S. M. Fullerton, R. C. Griffiths, J. Bond, M. J. Cox *et al.*, 1997 Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. **60:** 772–789.

Hayashi, K., 1991 PCR-SSCP: A simple and sensitive method for detection of mutations in the genomic DNA. PCR Methods Appl. **1:** 34–38.

Hey, J., 1997 Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14:** 166–172.

Horai, S., K. Hayasaka, R. Kondo, K. Tsugane and N. Takahata, 1995 Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92:** 532–536.

Huang, W., Y.-X. Fu, B. H. Chang, X. Gu, L. B. Jorde *et al.*, 1998 Sequence variation in ZFX introns in human populations. Mol. Biol. Evol. **15(2):** 138–142.

Jobling, M. A., and C. Tyler-Smith, 1995 Fathers and sons: the Y chromosome and human evolution. Trends. Genet. **11:** 449–456.

Jorde, L. B., M. J. Bamshad, W. S. Watkins, R. Zenger, A. E. Fraley *et al.*, 1995 Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am. J. Hum. Genet. **57:** 523–538.

Jorde, L. B., A. R. Rogers, M. Bamshad, W. S. Watkins, P. Krakowiak *et al.*, 1997 Microsatellite diversity and the demographic history of modern humans. Proc. Natl. Acad. Sci. USA **94:** 3100–3103.

Jorde, L. B., M. Bamshad and A. R. Rogers, 1998 Using mitochondrial and nuclear DNA markers to reconstruct human evolution. Bioessays **20:** 126–136.

Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins *et al.*, 1998 Signatures of population expansion in microsatellite repeat data. Genetics **148:** 1921–1930.

Kimura, M., and T. Ohta, 1973 The age of a neutral mutant persisting in a finite population. Genetics **75:** 199–212.

Lahr, M. M., and R. Foley, 1998 Towards a theory of modern human origins: geography, demography and diversity in recent human evolution. Am. J. Phys. Anthropol. **27**(Suppl.): 137–176.

Malaspina, P., F. Cruciani, B. M. Ciminelli, L. Terranato, P. Santolamazza *et al.*, 1998 Network analyses of Y-chromosomal types in Europe, Northern Africa, and Western Asia reveal specific patterns of geographic distribution. Am. J. Hum. Genet. **63:** 847–860.

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

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

Nickerson, D. A., S. L. Taylor, K. M. Weiss, A. G. Clark, T. G. Hutchinson *et al.*, 1998 Genome resequencing and variation analysis in a 9.7 kb region of the human lipoprotein lipase gene. Nat. Genet. **19:** 233–240.

Poloni, E. S., O. Semino, G. Passarino, A. S. Santachiara-Benerecetti, I. Dupanloup *et al.*, 1997 Human genetic affinities for Y-chromosome P49a,f/*Taq*I haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. **61:** 1015–1035.

Ravnik-Glavac, M., D. Glavac and M. Dean, 1994 Sensitivity of single-strand conformation polymorphism and heteroduplex method for mutation detection in the cystic fibrosis gene. Hum. Mol. Genet. **3:** 801–807.

Reich, D. E., and D. B. Goldstein, 1998 Genetic evidence for a Paleolithic human population expansion in Africa. Proc. Natl. Acad. Sci. USA **95:** 8119–8123.

Relethford, J. H., 1995 Genetics and modern human origins. Evol. Anthropol. **4:** 53–63.

Relethford, J. H., and H. C. Harpending, 1995 Ancient differences in population size can mimic a recent African origin of modern humans. Curr. Anthropol. **36:** 667–673.

Rogers, A. R., and H. Harpending, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9:** 552–569.

Rogers, A. R., and L. B. Jorde, 1995 Genetic evidence on modern human origins. Hum. Biol. **67:** 1–36.

Rogers, A. R., and L. B. Jorde, 1996 Ascertainment bias in estimates of average heterozygosity. Am. J. Hum. Genet. **58:** 1033–1041.

Schneider, S., J.-M. Kueffer, D. Roessli and L. Excoffier, 1997 Arlequin ver. 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Scozzari, R., F. Cruciani, P. Malaspina, P. Santolamazza, B. M. Ciminelli *et al.*, 1997 Differential structuring of human populations for homologous X and Y microsatellite loci. Am. J. Hum. Genet. **61:** 719–733.

Seielstad, M. T., J. M. Hebert, A. A. Lin, P. A. Underhill, M. Ibrahim *et al.*, 1994 Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum. Mol. Genet. **3(12):** 2159–2161.

Shimmin, L. C., B. H. Chang and W. H. Li, 1993 Male-driven evolution of DNA sequences. Nature **362:** 745–777.

Slatkin, M., and R. R. Hudson, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129:** 555–562.

Stoneking, M., J. J. Fontius, S. L. Clifford, H. Soodyall, S. S. Arcot *et al.*, 1997 Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. **7:** 1061–1071.

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

Takahata, N., 1993 Allelic genealogy and human evolution. Mol. Biol. Evol. **10:** 2–22.

Takahata, N., 1994 Repeated failures that led to the eventual success in human evolution. Mol. Biol. Evol. **11:** 803–805.

Tattersall, I., 1995 *The Fossil Trail: How We Know What We Think We Know About Human Evolution*. American Museum of Natural History. Oxford University Press, New York.

Tavaré, S., D. J. Balding, R. C. Griffiths and P. Donnelly, 1997 Inferring coalescence times for molecular sequence data. Genetics **145:** 505–518.

Templeton, A. R., 1997 Out of Africa? What do genes tell us? Curr. Opin. Genet. Dev. **7:** 841–847.

Tishkoff, S. A., E. Dietzsch, W. Speed, A. J. Pakstis, J. R. Kidd *et al.*, 1996 Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science **271:** 1380–1387.

Underhill, P. A., L. Jin, A. A. Lin, S. Q. Mehdi, T. Jenkins *et al.*, 1997 Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res. **7:** 996–1005.

Wang, T., Y. Okano, R. C. Eisensmith, M. L. Harvey, W. H. Y. Lo *et al.*, 1991 Founder effect of a prevalent phenylketonuria mutation in the Oriental population. Proc. Natl. Acad. Sci. USA **88:** 2146–2150.

Watterson, G. A., 1978 The homozygosity test of neutrality. Genetics **88:** 405–417.

Welsh, J. A., K. Castren and K. H. Vahakangas, 1997 Single-strand conformation polymorphism analysis to detect p53 mutations: characterisation and development of controls. Clin. Chem. **43:** 2251–2255.

Zietkiewicz, E., D. Sinnett, C. Richer, G. Mitchell, M. Vanasse *et al.*, 1992a Single-strand conformational polymorphisms (SSCP): detection of useful polymorphisms at the dystrophin locus. Hum. Genet. **89:** 453–456.

Zietkiewicz, E., L. R. Simard, S. B. Melançon, M. Vanasse and D. Labuda, 1992b Carrier status diagnosis in Duchenne muscular dystrophy with "conformational" DNA polymorphism. Lancet **339:** 134.

Zietkiewicz, E., V. Yotova, M. Jarnik, M. Korab-Laskowska, K. K. Kidd *et al.*, 1997 Nuclear DNA diversity in worldwide distributed human populations. Gene **205:** 161–171.

Zietkiewicz, E., V. Yotova, M. Jarnik, M. Korab-Laskowska, K. K. Kidd *et al.*, 1998 Genetic structure of the ancestral population of modern humans. J. Mol. Evol. **47:** 146–155.

Communicating editor: A. G. Clark