Archaic Lineages in the History of Modern Humans

Damian Labuda,*,† Ewa Ziętkiewicz† and Vania Yotova†

**Department of Pediatrics, University of Montreal and †Research Center, Sainte-Justine Hospital, Charles-Bruneau Cancer Center, Montreal, Quebec H3T 1C5, Canada*

Manuscript received January 28, 2000
Accepted for publication June 26, 2000

ABSTRACT

An important question in the ongoing debate on the origin of *Homo sapiens* is whether modern human populations issued from a single lineage or whether several, independently evolving lineages contributed to their genetic makeup. We analyzed haplotypes composed of 35 polymorphisms from a segment of the dystrophin gene. We find that the bulk of a worldwide sample of 868 chromosomes represents haplotypes shared by different continental groups. The remaining chromosomes carry haplotypes specific for the continents or for local populations. The haplotypes specific for non-Africans can be derived from the most frequent ones through simple recombination or a mutation. In contrast, chromosomes specific for sub-Saharan Africans represent a distinct group, as shown by principal component analysis, maximum likelihood tree, structural comparison, and summary statistics. We propose that African chromosomes descend from at least two lineages that have been evolving separately for a period of time. One of them underwent range expansion colonizing different continents, including Africa, where it mixed with another, local lineage represented today by a large fraction of African-specific haplotypes. Genetic admixture involving archaic lineages appears therefore to have occurred within Africa rather than outside this continent, explaining greater diversity of sub-Saharan populations observed in a variety of genetic systems.

IN the ongoing debate on the origin of *Homo sapiens*, genetic studies usually support the recently-out-of-Africa model, according to which modern humans emerged in Africa 200,000 to 100,000 years ago and then dispersed throughout the Old World, replacing preexisting archaic hominids with little or no admixture (Cann *et al.* 1987; Stringer and Andrews 1988; Harpending *et al.* 1993; Cavalli-Sforza *et al.* 1994; Chen *et al.* 1995; Horai *et al.* 1995; Nei 1995; Batzer *et al.* 1996; Tishkoff *et al.* 1996; Krings *et al.* 1997). The greater genetic variability in sub-Saharan Africa, seen in the molecular data (Cann *et al.* 1987; Bowcock *et al.* 1991, 1994; Jorde *et al.* 1997; Clark *et al.* 1998; Hammer *et al.* 1998; Zietkiewicz *et al.* 1998) as well as in craniometic measurements (Relethford and Harpending 1994), was used as an argument for (i) the older age of the population in Africa (Cann *et al.* 1987) and (ii) a larger long-term effective population size on this continent (Relethford and Harpending *et al.* 1995; Rogers and Jorde 1995; Zietkiewicz *et al.* 1998). The underlying assumption in these interpretations was a continuity of genetic pools among populations of modern humans, possibly constrained by the recent out-of-Africa demographic bottleneck (Tishkoff *et al.* 1996). Our analysis of DNA haplotypes in an 8-kb segment from the dystrophin gene (hereafter referred to as *dys44*) provides evidence for a discontinuity in the genetic makeup of the present-day humans. In particular, (i) it provides evidence for an expansion of one lineage throughout different continents and (ii) it suggests that a greater genetic diversity in sub-Saharan Africans as compared to other continental populations could be due to enrichment of their genetic pool through admixture with independently evolving local lineage(s).

Corresponding author: Damian Labuda, Charles-Bruneau Cancer Center, Research Center, Sainte-Justine Hospital, 3175 Côte-Ste Catherine, Montreal, Quebec H3T-1C5, Canada.
E-mail: damian.labuda@umontreal.ca

MATERIALS AND METHODS

Populations: The reported haplotypes represent the population sample consisting of 262 sub-Saharan African chromosomes (58 West African, 56 M'Buti, and 81 Biaka Pygmy, and 67 African American); 187 European (25 Polish, 26 Italian, 106 French-Canadian, and 30 of mixed European origin); 195 Asian (65 Japanese, 80 Chinese, 22 Siberian, 28 Mongolian); 159 from the Americas (76 Maya and 83 Karitiana), and 65 from Oceania (24 Coastal and 41 Highland New Guinea Papuan chromosomes). DNA samples were kindly provided by M. Batzer, M. Jamba, J. Jaruzelska, K. Kidd, D. Modiano and L. Osipova or isolated from peripheral blood collected with prior informed consent from volunteers in Montreal. All samples, except for Mongolians and additional Europeans, were previously analyzed for *dys44* polymorphisms, as reported by Zietkiewicz *et al.* (1997, 1998).

Nucleotide diversity data and derivation of haplotypes: The genomic segment *dys44* comprises 7622 bp of mostly intronic sequence surrounding exon 44 of the human dystrophin gene on the Xp21. Thirty-five simple nucleotide polymorphisms, including two three-nucleotide deletions and one eight-nucleotide duplication as well as one three-allelic site due to two substitutions, were previously ascertained by single-strand conformational polymorphism heteroduplex analysis in 250 worldwide distributed chromosomes (Zietkiewicz *et al.* 1997).

The analyzed population samples were subsequently typed by allele-specific oligonucleotide hybridization as described (Ziętkiewicz *et al.* 1997).

The haplotypes reported here were derived as a subset of an extended population sample of 1289 chromosomes, which included additional chromosomes of Euroasiatic origin (to be reported elsewhere—E. Ziętkiewicz, V. Yotova and D. Labuda, unpublished results). Converting autosomal genotype data into haplotypes typically requires pedigree analysis or additional experimental effort like allele-specific amplification to determine the phase (*e.g.*, Fullerton *et al.* 1994; Harding *et al.* 1997; Clark *et al.* 1998). Computer-assisted approaches rely on subtracting known haplotypes from the multiple heterozygote genotypes, which eventually initiates a cascade of haplotype solving (Clark 1990). In the X-linked *dys44* segment the derivation of haplotypes was straightforward in hemizygous males as well as in homozygous and single heterozygous females. These haplotypes represented 67% of the chromosomes in the sample analyzed (in Figure 1, these haplotypes have names beginning with an uppercase B). In the next step the remaining multiple-heterozygous female samples were resolved assuming the most likely combinations of two unambiguously solved haplotypes. Compared to the algorithm described by Clark (1990) our approach additionally took advantage of the population frequencies of the unambiguously assigned haplotypes. Briefly, an algorithm was implemented (ACCESS database) creating a table with the genotypes in question presented, whenever possible, as a sum of two known haplotypes in all combinations. All unique solutions were accepted, whereas in the case of multiple choices, choosing the most likely combination was based on two criteria: (i) the combined frequencies (product) of the contributing haplotypes and (ii) these haplotypes' hierarchical occurrence, with the priority given to the local population, then to the closest group, and finally to the continental cluster (when, for a given sample, these two criteria favored two different solutions, the genotype was left for "manual" solving). This phase resulted in resolving 25% of all chromosomes (or three-fourths of multiple heterozygous genotypes). The remaining genotypes (8.2% of all chromosomes) could not be presented as a sum of two known haplotypes, but in a significant portion a contribution of one of the known haplotypes could be recognized. The choice of a new haplotype from among all possible combinations that included a known haplotype was done "manually." The most likely solutions were chosen, taking into account the population frequency of the known haplotype and the recurrent appearance of the new derived haplotype. Simplicity of the new haplotype structure was used as an additional criterion: a new haplotype was chosen if it could be derived preferably by a single (or if not, by a double) recombination between known haplotypes, whose frequencies in populations were taken into account in the case of multiple choices. In the portion of genotypes solved in this way (5.6% of the total), *i.e.*, by indirectly inferring novel haplotypes, these were assigned names beginning with a lowercase b. The remaining genotypes were left unsolved since (i) no contribution of any known haplotype could be recognized, (ii) several solutions were equally likely using the criteria above, or (iii) no simple likely connection (through recombination or mutation) to known haplotypes could be assigned. As a result 17 samples (34 chromosomes, 2.6% of the total) were excluded from further analyses [for this reason site 86 (see Ziętkiewicz *et al.* 1998) was not considered here]; the number of reported haplotypes is thus conservative. Haplotypes found in more than one continental group of populations were called common; those restricted to one continental group or to a population were called specific.

Quantitative analyses: Principal component (PC) analysis using allele frequencies at 35 *dys44* segregating sites in different populations or groups of chromosomes was carried out using the program Population Structure Statistics by Harpending (Harpending *et al.* 1996), kindly provided by M. Batzer. A maximum likelihood tree from the allele frequencies was constructed using CONTML (and evaluated using BOOTSTRAP) program from Phylip package 3.5 (Felsenstein 1993). The tree was rooted by setting at 1.0 the frequencies of ancestral alleles inferred by comparison with orthologous positions in chimpanzee, gorilla, and/or orangutan DNA: the human allele identical by state with at least two of the great ape orthologues was considered ancestral. Given the mutation rate in the order of 10^{-9} per nucleotide site per year, there is a very small chance that a recurrent event would have taken place since the divergence of these species and the human lineage 5–12 million years ago. In other words, the probability is negligible that a human allele identical with the corresponding site in the great apes results from a back mutation or that three identical mutations occurred independently after the separation of these lineages. In practice, therefore, in the case of simple nuclear DNA polymorphisms considered here, the identity by state of a human allele and the corresponding positions in great apes is tantamount to their identity by descent (Ziętkiewicz *et al.* 1998). Nucleotide diversity, h, was calculated from allele frequencies (gene counting) at 35 segregating sites in the total sample, in Africans and non-Africans, or in population subgroups containing common or specific haplotypes. The F_{ST} statistic was evaluated as $(h_T - h_S)/h_T$, where h_T corresponds to the nucleotide diversity of the entire population (calculated from the weighted allele frequencies) and h_S to the average of the contributing subpopulations (Hartl and Clark 1989). If two subpopulations diverge and remain genetically isolated, F_{ST} is expected to grow with time: $F_{ST}(g) = 1 - \exp(-g/2N)$, where N is the long-term effective population size of the diverging populations and g is the number of generations since the divergence (Wright 1969); in the case of *dys44*, $1.5N$ rather than $2N$ has to be used in the formula above, correcting for the ratio of X chromosomes to autosomes.

RESULTS

Excess of *dys44* variability in Africa: In *dys44* segment, 36 simple polymorphisms were previously ascertained in a sample of 250 X chromosomes from all over the world (Ziętkiewicz *et al.* 1997, 1998). Following extended genotyping of these polymorphisms in more than a thousand chromosomes (not shown), 21 new alleles were found distributed worldwide, 13 were confirmed unique to sub-Saharan Africa (hereafter referred to as Africa), and only 2 were restricted to populations from other continents (Japan and Papua New Guinea). The new information reported here was obtained by analyzing polymorphic sites in their genomic context, *i.e.*, as haplotypes representing combinations of alleles on a single chromosome. *Dys44* haplotypes shown in Figure 1 were built from 35 sites, *i.e.*, excluding a solitary African site 86 (see materials and methods) and counting a single three-allelic polymorphism as two separate sites, 95 and 96. Derivation of haplotypes required substantial additional experimental effort; inversion of an allele due to misreading or mistyping was of little consequence for frequency evaluation in a large sample

but, in most cases, it would have led to a novel false haplotype. The analysis therefore included verification of the data and retyping of a variety of samples, resulting in elimination of spurious haplotypes and confirmation of the reported ones.

Sixty-four haplotypes were observed among 868 chromosomes representing 19 populations from different continents. Haplotypes found in populations from more than one continent are referred to as common, as opposed to specific ones, found only in a single continental group. Fourteen haplotypes were classified as common (Figure 1a). The chromosomes carrying common haplotypes represented 82% of the total sample, and among these, 75% chromosomes in Papuans from New Guinea (PNGs), 93% in Asians, 93% in Europeans, 96% in Amerindians, and only 58% in Africans. Among the common haplotypes all but one (b052—see Figure 1a) were shared between Africans and non-Africans. The remaining 50 haplotypes (Figure 1b) were specific to the continental groups: 28 were found on chromosomes from Africans, and 22 were found on chromosomes representing PNGs, Asians, Europeans, and Amerindians. The greater African diversity relative to other continental groups was therefore seen in the increased number of *dys44* segregating sites as well as in the excess of specific haplotypes (42 and 8% of all chromosomes from Africa and outside, respectively); see *Summary statistics* below.

Discontinuity in *dys44* haplotypes: Inspecting the structure of *dys44* haplotypes reveals a striking dissimilarity between African-specific and the remaining chromosomes. In non-Africans, the groups of common and specific haplotypes appear to be closely related: specific ones can be most parsimoniously derived (i) by mutations occurring on the most-frequent common haplotypes (sites 48 and 87, on B001 and B002, respectively) or (ii) by recombinations involving two or three common haplotypes from among the 8 most frequent (Figure 1b, right column). In contrast, specific haplotypes in Africans are structurally disparate from the common ones. First, 19 out of the 28 African-specific haplotypes are easily recognized by the presence of African-only polymorphisms (see Figure 1b). Second, the structure of the haplotypes on which these polymorphisms reside cannot be related in a simple way to those that are shared with other continents. Only 5 of the African-specific haplotypes (B024, B032, B037, b063, and B027) could be proposed to result from recombinations or a mutation (at site 64) involving exclusively common haplotypes. The remaining ones cannot be related in a simple way to the set of common haplotypes; in contrast, they can be easily derived through a single mutation or through a single or double recombination by implicating other haplotype(s) from the African set. To illustrate the possibility of such a structural relationship, five frequent African-specific haplotypes (B007, B010, B012, B018, and B017, indicated by dashes in the last column in Figure 1b) were arbitrarily selected to derive the less frequent ones. At three instances (B022, B047, b059) multiple genetic events, through a number of possible pathways, would have to be postulated to relate these with any other known haplotypes—these African haplotypes were therefore left unassigned (denoted by a question mark in the last column in Figure 1b).

Initially, the low world frequency of the new alleles at African-specific sites suggested the recent origin of the underlying mutations (Zietkiewicz *et al.* 1998). However, their haplotype context contradicts such an interpretation. First, the same new allele is found associated with many different haplotypes, causing the reduction of linkage disequilibrium with the adjacent polymorphisms. At four sites the same new allele was associated with at least 3 (up to 11) haplotypes, at four sites with 2 different haplotypes, and only four sites were uniquely found associated with a single haplotype. This indicates that the corresponding chromosomes must have undergone multiple historical recombinations. Second, different African-specific sites tend to cluster within the same haplotype, with up to four new African alleles found on the same haplotype (*e.g.*, B047—4 alleles; B012, B018, B021, and b062—3 sites). Such clustering could have resulted (i) from succession of new mutations within the same chromosome lineage or (ii) from recombinations juxtaposing mutations that arose on independent chromosomes. Both mechanisms are expected to require a substantial evolutionary time to generate the observed structure of the African-specific haplotypes. The mutation rate in *dys44* was estimated at 2.3×10^{-4} per generation per segment (Zietkiewicz *et al.* 1998) and the recombination rate can be assumed to be in the same order of 10^{-4} per generation per segment (Abbs *et al.* 1990).

According to the qualitative analysis above, the specific chromosomes in continental groups other than Africans can be directly derived from the common haplotypes, indicating evolutionary continuum, while the African-specific chromosomes appear to represent a subset with an evolutionary history distinct from the rest of the sample. A series of tests were carried out to examine these observations independently and to quantify the differences between the continental pools of common and specific chromosomes.

PC analysis: This analysis has the advantage of allowing the graphical presentation of multidimensional data in a reduced number of dimensions (Cavalli-Sforza *et al.* 1994). First we used the frequencies of 35 *dys44* polymorphisms to infer genetic affinities among the populations represented in our sample. Second, to examine whether the effect of continentally restricted sites, "weighting" especially heavily in the African-specific sample, could alone be responsible for the observed effect, the analysis was repeated considering only 21 worldwide distributed polymorphisms. A remarkable separation of the African-specific haplotypes from the

rest of the sample was seen using both total and the reduced number of sites. Figure 2 showing the results for 21 sites illustrates well this effect, which obviously was even more dramatic with 35 polymorphisms (not shown). On the PC plot in Figure 2a geographically related groups of populations cluster together; the variance attributable to the first and the second component (50.4 and 28.2% to the total variance, respectively) reflected differences between the continental groups and was enhanced by the presence of outliers represented by population isolates such as Karitiana from Brazil, Highland Papuans from New Guinea, and M'Buti Pygmies from Congo. To examine the relatedness of the common and specific haplotypes, the groups of common and specific chromosomes in the populations were considered separately (specific chromosomes in non-African populations were pooled continentally because of small per population numbers). This stratification led to a rearrangement of the PC plot (Figure 2b). African-specific chromosomes are clearly separated from the cluster of common African haplotypes found in the center of an even greater cluster including all other non-African groups, both common and specific. This analysis confirms conclusions from structural comparisons above (Figure 1 and previous section), indicating that the African-specific haplotypes represent lineages distinct from those of the common haplotypes and the closely related non-African-specific haplotypes.

Phylogenetic analysis: Subsequently, the frequencies of the polymorphisms in *dys44* were used to construct the maximum likelihood tree of the groups of common and specific haplotypes as in the PC analysis. The tree shown in Figure 3 was constructed considering the 21 worldwide-distributed sites (essentially the same tree, only with deeper separation among the groups of African-specific chromosomes, was obtained using all 35 polymorphisms, not shown). In the tree, the non-African populations clustered continentally, with bootstrap analysis supporting grouping of European (82/100) and Amerindian populations (71/100). The separation of the African-specific haplotypes from the remaining groups, including common African haplotypes, was supported by the highest bootstrap value in the tree (88/100), thus further reaffirming the distinctiveness of the African-specific haplotypes. In addition, the African-specific haplotypes were the closest to the root of the tree and more dispersed, suggesting that they represented relatively older lineages.

Summary statistics: Table 1 compiles nucleotide diversity values that were obtained for all chromosomes in the total sample, in non-Africans and Africans (columns), and for the corresponding pools of chromosomes carrying common and specific haplotypes (rows). The overall nucleotide diversity $h = 0.000977$ for the total sample is similar to the values obtained with other loci (Li and Sadler 1991; Harding *et al.* 1997; Clark *et al.* 1998; Harris and Hey 1999; Jaruzelska *et al.*

a

Haplo ID	2	3	5	8	10	12	14	15	18	20	25	30	32	33	35	38	40	45	48	50	55	58	59	64	65	70	71	72	85	87	88	90	93	95	96	Continental occurence	Count N-Afr.	Count African
B001		1					1					1			1			1		1		1	1		1						1	1	1		All	197	33
B002	1			1			1					1										1					1								<>Amin	106	67
B004	1			1			1															1					1					1	1	1	All	82	2
B003		1					1					1			1			1		1		1	1		1	1					1	1	1	All	62	11
B006					1							1																1					1	<>PNG	55	3
B005							1			1		1			1		1			1		1	1		1						1	1	1	<>Amin	26	11
B009		1					1			1		1			1			1		1		1	1		1	1					1	1	1	Afr, Euro	1	13
B008							1			1		1			1			1		1	1	1			1						1	1	1	Afr, Euro	11	2
B011		1					1			1		1			1		1	1		1		1	1		1						1	1	1	Afr, Euro	8	2
B016		1					1			1		1					1			1		1	1		1	1					1	1	1	Afr, Euro	2	4
B019		1					1					1			1		1			1		1	1		1	1					1	1	1	Afr, Amin	2	2
b052		1			1							1																	1					Amin, Eur	3	0
B031		1					1					1			1		1			1		1	1		1						1	1	1	Afr, Euro	1	1
B039		1					1			1		1			1		1			1		1	1		1						1	1	1	Afr,Euro	1	1

Figure 1.—*Dys44* haplotypes occurring on more than one continent are listed as common (a) while those restricted to a single continent as specific (b). The haplotypes (built of 35 segregating sites indicated at the top) are defined by the presence (number 1) or absence (empty space) of a new allele at each of the sites (allele shared with the chimpanzee, gorilla, and/or orangutan was considered ancestral). The order of polymorphisms corresponds to the physical map; new alleles whose occurrence is continentally restricted are boxed (Ziętkiewicz *et al.* 1998). The symbol <> indicates "all except." In the putative genealogical reconstruction of specific haplotypes (b, last column) the haplotype name followed by the name of a segregating site indicates the background haplotype on which this polymorphism occurred; the haplotype name followed by another haplotype name(s), separated by /, represents a list of the proposed parental recombining haplotypes. Five frequent African-specific haplotypes (B007, B010, B012, B017, and B018) arbitrarily selected to derive the less frequent ones through mutation or recombination are indicated by dashes in the last column. Three African-specific haplotypes (B022, B047, b059) where multiple genetic events, through a number of possible pathways, would have to be postulated to relate these with any other known haplotypes are denoted ? in the last column.

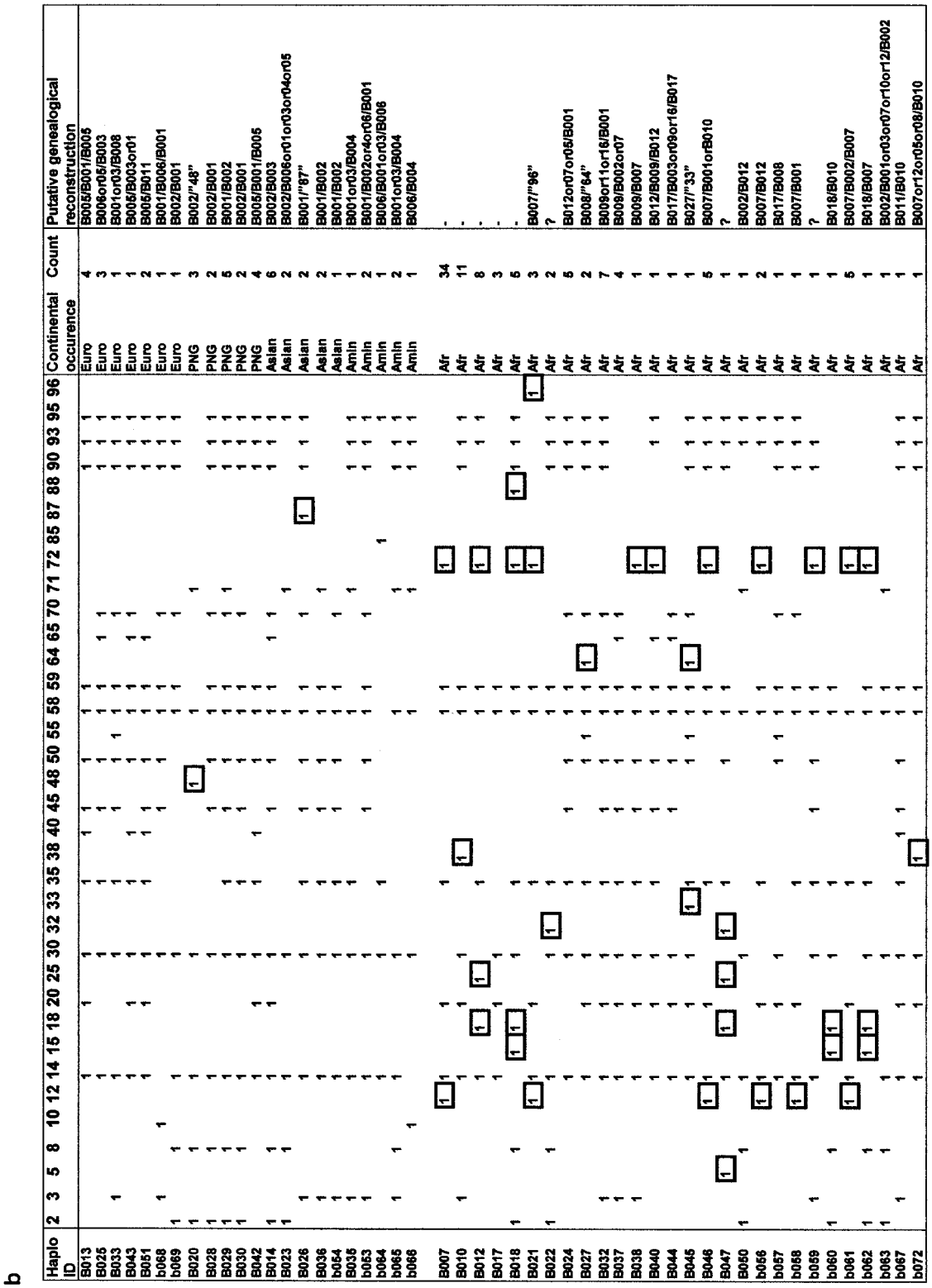

FIGURE 1.—*Continued.*

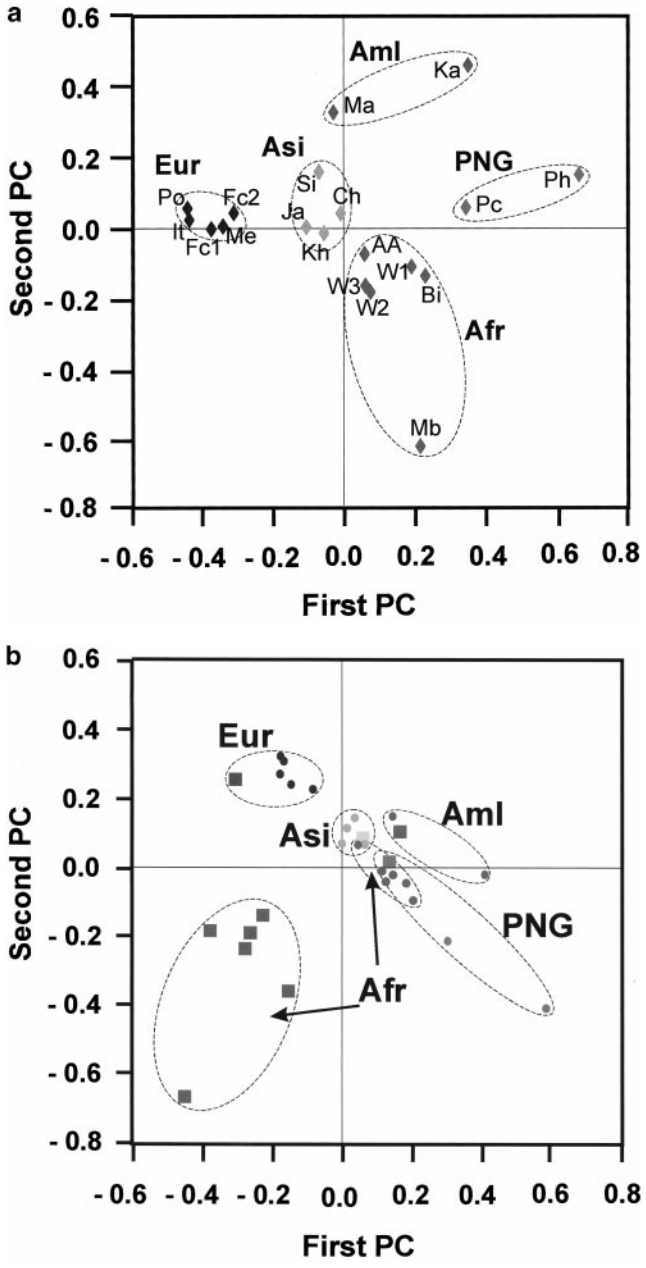

FIGURE 2.—Principal component analysis using allele frequencies at 21 worldwide-distributed *dys44* polymorphisms. (a) Within each population, all chromosomes are considered together. (b) Chromosomes carrying common and specific haplotypes indicated by circles and squares, respectively, are considered separately; specific chromosomes in non-African populations were grouped into continental pools to increase sample size. Afr (Africans): W1, W2, W3, populations from West Africa; Fc1, Fc2, French-Canadians from Montreal and North-Eastern Quebec, respectively; Mb, Mbuti Pygmies; Bi, Biaka Pygmies; AA, African Americans. Eur (Europeans): Po, Polish; It, Italian; Mt, Montreal; Sa, Saguenay/Lac-St-Jean; Me, mixed European origin. Asi (Asians): Ch, Chinese; Ja, Japanese; Si, Siberians; Kh, Khalka Mongolians. AmI (Amerindians): Ma, Maya; Ka, Karitiana. PNG (Papuans from New Guinea): Pc, Coastal; Ph, Highland.

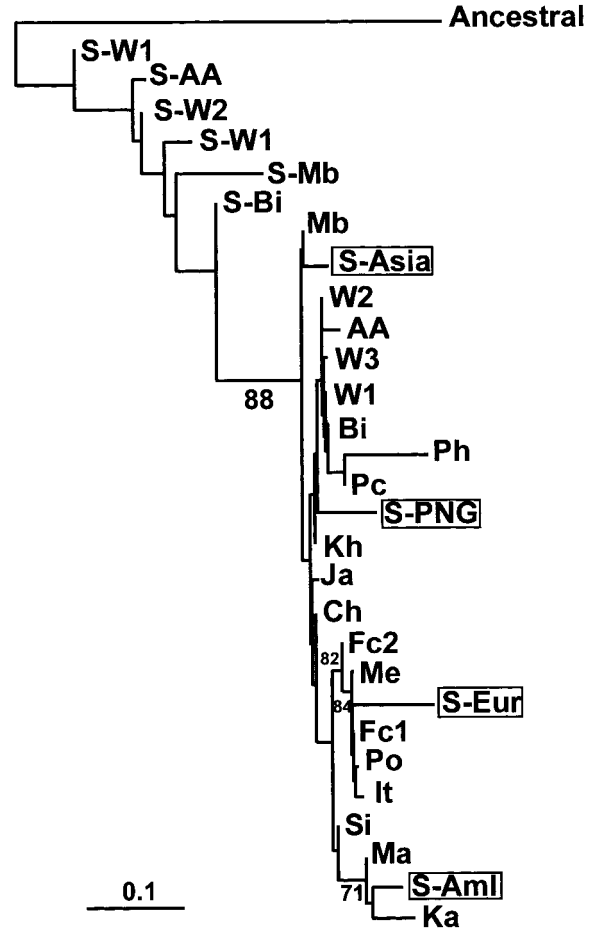

FIGURE 3.—Maximum likelihood tree using allele frequencies at 21 worldwide-distributed polymorphisms in pools of common and specific chromosomes from the populations under study as indicated in the legend to Figure 2 (S- denotes "specific"). Bootstrap values of more than 70/100 replicates are indicated on the tree.

1999) and leads to the effective population size estimate of 10,900; the h values for different subgroups listed in Table 1 yield similar estimates of the effective population size. However, in the pool of African-specific chromosomes characterized by the overall h value of 0.000885, the contribution of the worldwide polymorphisms is only 0.000627 (*i.e.*, one-third less than in the pool of common chromosomes), which is compensated for by the relatively high contribution of the continentally restricted sites (0.000258).

The magnitude of the population variance that is due to the difference among subgroups can be measured by a statistic F_{ST} evaluated from nucleotide diversities (MATERIALS AND METHODS); if two subpopulations diverge and remain genetically isolated, F_{ST} is expected to grow with time (WRIGHT 1969). The world pool of common haplotypes subdivided into Africans and non-Africans was characterized by a very low F_{ST} value of 0.016, revealing little separation between Africa and the rest of the world. However, when the pool of specific

TABLE 1

Nucleotide diversities of the total chromosomal sample and the separate pools of common and specific chromosomes

Haplotypes	Population	Total	Non-African	African
All	No. chromosomes	868	606	262
	h ($\times 10^3$)	0.977 (0.052)[a]	0.897 (0.002)[a]	1.051 (0.148)[a]
Common	No. chromosomes	709	557	152
	h ($\times 10^3$)	0.911	0.895	0.923
Specific	No. chromosomes	159	49	110
	h ($\times 10^3$)	1.016 (0.221)[a]	0.832 (0.025)[a]	0.885 (0.258)[a]

[a] The contribution of the continentally restricted sites (see boxed positions in Figure 1) to the overall nucleotide diversity (h) value is given between parentheses.

haplotypes was subdivided in a similar way, the F_{ST} value was 0.167. Not surprisingly, this effect was due to the specific haplotypes in Africans: stratification of all the African chromosomes into two pools, of common and specific haplotypes, led to the substantial F_{ST} value of 0.141, while similar stratification of non-African chromosomes yielded F_{ST} of 0.029. Considering common and specific haplotypes in Africans to represent genetic pools of independent subpopulations, one may estimate (assuming N of ~10,000) the period of their separation at 2260 generations or ~45,000 years (see MATERIALS AND METHODS). This estimate uses the effective population size based on a number of nuclear data sets and the assumption of panmixia. If the proposed model of ancient structure is true, the effective population size would be lower, and the estimate for the time of separation should be considered an upper bound. On the other hand, this has to be taken with extreme caution due to a great variance associated with such an estimate.

DISCUSSION

Admixture with ancient lineages: Similar to other systems, the *dys44* segment reveals greater diversity among sub-Saharan Africans. However, neither the older age of African populations (CANN *et al.* 1987; TISHKOFF *et al.* 1996) nor their greater effective population size (RELETHFORD and HARPENDING 1995; ROGERS and JORDE 1995) following divergence from the homogeneous ancestral stock (ZIETKIEWICZ *et al.* 1998) can alone explain the pattern and distribution of *dys44* haplotypes described above. Examination of the structure of *dys44* haplotypes, supported by PC analysis, maximum likelihood trees, and summary statistics, reveals a sharp discontinuity between groups of common and African-specific chromosomes. Although the analysis presented here should be considered exploratory, these results suggest that the genetic pool of sub-Saharan Africans represents two lineages that had evolved separately for some period of time and eventually hybridized. The F_{ST} value obtained for the African sample stratified into subpopulations of common and specific haplotypes, considered to represent these lineages, suggests that the time of their separation could have lasted ~50,000 years. Given the underlying assumptions (WRIGHT 1969) and uncertainty in N estimate, this should be considered rather as an indication of a time range between 10^4 and 10^5 years. A period of 20,000–100,000 years corresponds to that between the appearance of the first anatomically modern humans and the Upper Paleolithic expansion. The earliest fossils displaying modern characteristics, from Omo Kibbish in Ethiopia, Klassies River Mouth in South Africa, and Skhul and Oafzeh in Israel, indicate that humans inhabited widely disparate regions within Africa and the Middle East as early as ~120,000–100,000 years ago (BRÄUER 1989; LAHR and FOLEY 1994), *i.e.*, during the last interglacial. The following periods of arid climate, due to glacial maxima, could have led to the fragmentation of population inhabiting this geographic area (LAHR and FOLEY 1999). This would promote independent evolution of separated local populations. On the other hand, according to archeological evidence, the Upper Palaeolithic revolution began in Africa and outside only around 50,000–40,000 years ago (KLEIN 1995), coinciding with the range expansion and important demographic growths. The evidence for the latter is found in the mitochondrial DNA (DI RIENZO and WILSON 1991; ROGERS and HARPENDING 1992; WATSON *et al.* 1997; KIVISILD *et al.* 1999; QUINTANA-MURCI *et al.* 1999) supported by microsatellite data (DI RIENZO *et al.* 1998; KIMMEL *et al.* 1998). *Dys44* haplotype data fit a scenario in which previously separated and independently evolving lineages mix together. The minimal model requires at least two such lineages, a local one that stayed in Africa and a second one that expanded within this continent and outside, which presumably occurred in the Upper Paleolithic.

The contribution of archaic lineages to the genetic makeup of present-day populations has been previously debated in the context of early Eurasian populations, descendants of *H. erectus*, facing the recent out-of-Africa expansion, or in the context of the putative hybridization between Neanderthals and Upper Paleolithic populations in Europe (BRÄUER 1989; THORNE and WOLPOFF

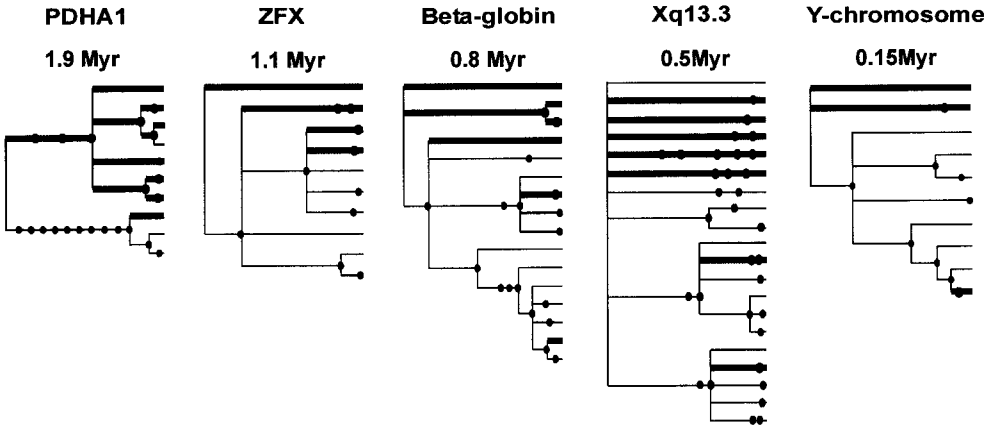

FIGURE 4.—Archaic African lineages. Comparison of gene trees of PDHA1, ZFX, β-globin, and Y chromosome, redrawn from HAMMER *et al.* (1998), JARUZELSKA *et al.* (1999), HARDING *et al.*, (1997), and HARRIS and HEY (1999), respectively. The coalescent tree of Xq13.3 was calculated according to GRIFFITHS and TAVARÉ (1994, 1999) using the data from KAESSMANN *et al.* (1999). The heights of the trees in million years are shown at the top. The African-only branches are indicated by thick lines.

1992; KLEIN 1995; KRINGS *et al.* 1997; DUARTE *et al.* 1999; TATTERSAL and SCHWARTZ 1999). In our data the only compelling evidence for mixing of separate lineages is found within the African sample. While not excluding the possibility of such events outside Africa, our model differs from others by proposing admixture between the expanding population (represented by common haplotypes) and the local (specific) African lineages. This new interpretation not only embraces findings from other genetic systems but also provides a better explanation. Expansion of the lineages represented by common chromosomes outside Africa can be perceived as the recently-out-of-Africa bottleneck (TISHKOFF *et al.* 1996); their expansion within Africa can lead to an interpretation of a contribution of other continents to the genetic make-up of African populations (HARDING *et al.* 1997; HAMMER *et al.* 1998). The joint contribution of the "expanding" and "local" lineages, which increases genetic diversity of the present-day African genetic pool, can be interpreted as the indication of greater antiquity (CANN *et al.* 1987) or larger long-term effective population size of Africans (RELETHFORD and HARPENDING 1995; ROGERS and JORDE 1995). Given structural discontinuity in the African genetic pool, seen clearly in *dys44* data but also apparent from the examination of gene trees from other systems (see below), these genetic findings can be most parsimoniously explained within the frame of the proposed model.

Other data: The interpretation above finds support at other nuclear loci tested for polymorphisms in genomic samples of different geographic origin. In the studies of three X-linked DNA segments, PDHA1 (HARRIS and HEY 1999), ZFX (JARUZELSKA *et al.* 1999), and Xq13.3 (KAESSMANN *et al.* 1999), of a fragment of the β-globin gene (HARDING *et al.* 1997), as well as the Y chromosome (HAMMER *et al.* 1998), owing to the absence of recombinations (or by neglecting their relatively minor contribution), it was possible to analyze the underlying gene trees by coalescent approach making use of the full data (GRIFFITHS and TAVARÉ 1994). In the corresponding trees, redrawn in Figure 4, the oldest branches (starting at or near the root of the coalescent trees) represent almost uniquely African chromosomes, thus essentially displaying the same pattern as in Figure 3. It is noteworthy that similar topology, with the African-only branches separated at the root of the tree from the remaining, mixed non-African and African branches, was found with the mitochondrial DNA tree as well (CANN *et al.* 1987; CHEN *et al.* 1995). In Figure 4 the discontinuity between African-specific branches and the rest of the world is exceptionally deep in the PDHA1 tree (see, however, DISOTELL 1999; HARDING 1999) and is least pronounced in the Xq13.3 tree, where the ancestral haplotype is relatively very frequent and occurs outside Africa as well. In all these trees the split between the oldest (locally evolving) African-only branches and the rest of the world takes place at the deepest nodes of the tree, as in Figure 3. The height of these trees (see Figure 4), dating this split accordingly, differs substantially. However, these coalescent estimates were obtained using the simple model of a homogenous, constant-size population. When we take into account the present data that suggest admixture involving independent lineages, the simple model used in these calculations above does not hold, and the resulting time estimates (Figure 4) will have to be reevaluated under different demographic and evolutionary scenarios.

Combining *dys44* polymorphisms into a haplotype provided important information that could not be gained by analyzing them separately. The high contribution of recombinations into *dys44* haplotype diversity, while preventing an application of standard methods of coalescent analysis that make use of the whole data in the gene tree, allowed us to look at the history of the analyzed segment from a different perspective. However, it seems that optimal analytical tools with which to investigate jointly recombination and mutational data under complex demographic scenarios have not yet been developed. We hope that our results will stimulate the development of new and perhaps more appropriate methods of formal genetic analysis for studying the history of human populations.

We thank A. Di Rienzo and E. Poloni for comments on the earlier version of the manuscript, J-F. Bibeau and D. Gehl for assistance in data analysis and numerous contributions of DNA samples, M. Jamba for typing the Mongolian DNA, and R. Ballarano for secretarial assistance. This work was supported by the CGAT Program and the Medical Research Council of Canada.

LITERATURE CITED

Abbs, S., R. G. Roberts, C. G. Mathew, D. R. Bentley and M. Bobrow, 1990 Accurate assessment of intragenic recombination frequency within the Duchenne muscular dystrophy gene. Genomics **7:** 602–606.

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.

Bräuer, G., 1989 The evolution of modern humans: a comparison of the African and non-African evidence, pp. 124–154 in *The Origins and Dispersal of Modern Humans: Behavioural and Biological Perspectives*, edited by P. Mellars and C. B. Stringer. Edinburgh University Press, Edinburgh.

Cann, R. L., M. Stoneking and A. C. Wilson, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

Cavalli-Sforza, L. L., P. Menozzi and A. Piazza, 1994 *The History and Geography of Human Genes.* Princeton University Press, Princeton, NJ.

Chen, Y.-S., A. Torroni, L. Excoffier, A. S. Santachiara-Benerecetti and D. C. Wallace, 1995 Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am. J. Hum. Genet. **57:** 133–149.

Clark, A., 1990 Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. **7:** 111–122.

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.

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:** 1269–1281.

Disotell, T. R., 1999 Human evolution: origins of modern humans still look recent. Curr. Biol. **9:** R647–R650.

Duarte, C., J. Maurício, P. B. Pettitt, P. Souto, E. Trinkaus *et al.*, 1999 The early upper paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. Anthropology **96:** 7604–7609.

Felsenstein, J., 1993 PHYLIP (Phylogeny Inference Package), Version 3.5p, distributed by the author. Department of Genetics, University of Washington, Seattle (1989 PHYLIP—Phylogeny Inference Package, Version 3.2. Cladistics **5:** 164–166).

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, 1999 The ages of mutations in gene trees. Ann. Appl. Prob. **9:** 567–590.

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:** 427–441.

Harding, R. M., 1999 More on the X files. Proc. Natl. Acad. Sci. USA **96:** 2582–2584.

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.

Harpending, H. C., S. T. Sherry, A. R. Rogers and M. Stoneking, 1993 The genetic structure of ancient human populations. Curr. Anthropol. **34:** 483–496.

Harpending, H., J. Relethford and S. T. Sherry, 1996 Methods and models for understanding human diversity, pp. 283–299 in *Molecular Biology and Human Diversity*, edited by A. J. Boyce and C. G. N. Mascie-Taylor. Cambridge University Press, Cambridge, UK.

Harris, E. E., and J. Hey, 1999 X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. USA **96:** 3320–3324.

Hartl, D. L., and A. G. Clark, 1989 *Principles of Population Genetics.* Sinauer Associates, Sunderland, MA.

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.

Jaruzelska, J., E. Zietkiewicz, M. Batzer, D. E. C. Cole, J.-P. Moisan *et al.*, 1999 Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. Genetics **152:** 1091–1101.

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

Kaessmann, K., F. Heisig, A. von Haeseler and S. Paabo, 1999 DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat. Genet. **22:** 78–81.

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.

Kivisild, T., M. J. Bamshad, K. Kaldma, M. Metspalu, E. Metspalu *et al.*, 1999 Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr. Biol. **9:** 1331–1334.

Klein, R. G., 1995 Anatomy, behavior and modern human origins. J. World Prehist. **9:** 181–183.

Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking *et al.*, 1997 Neandertal DNA sequences and the origin of modern humans. Cell **90:** 19–30.

Lahr, M. M., and R. Foley, 1994 Multiple dispersals and modern human origins. Evol. Anthropol. **3:** 48–60.

Lahr, M. M., and R. Foley, 1999 Towards a theory of modern human origins: geography, demography and diversity in recent human evolution. Am. J. Phys. Anthropol. **27** (Suppl.): 137–176.

Li, W. H., and L. A. Sadler, 1991 Low nucleotide diversity in man. Genetics **129:** 513–523.

Nei, M., 1995 Genetic support for the out of Africa theory of human evolution. Proc. Natl. Acad. Sci. USA **92:** 6720–6772.

Quintana-Murci, L., O. Semino, H.-J. Bandels, G. Passarino, K. McElreavey *et al.*, 1999 Genetic evidence of an early exit of Homo sapiens from Africa through eastern Africa. Nat. Genet. **23:** 437–441.

Relethford, J. H., and H. C. Harpending, 1994 Craniometric variation, genetic theory and modern human origins. Am. J. Physical. Anthropol. **95:** 249–270.

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. C. 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.

Stringer, C. B., and P. Andrews, 1988 Genetic and fossil evidence for the origin of modern humans. Science **239:** 1263–1268.

Tattersal, I., and J. H. Schwartz, 1999 Hominids and hybrids: the place of Neanderthals in human evolution. Proc. Natl. Acad. Sci. USA **96:** 7117–7119.

Thorne, A. G., and M. H. Wolpoff, 1992 The multiregional evolution of humans. Sci. Am. **266:** 76–83.

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.

Watson, E., P. Forster, M. Richards and H.-J. Bandelt, 1997 Mitochondrial footprints of human expansions in Africa. Am. J. Hum. Genet. **61:** 691–704.

Wright, S., 1969 *Evolution and the Genetics of Populations*, Volume 2, The Theory of Gene Frequencies. University of Chicago Press, Chicago.

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: S. Tavaré