Using mitochondrial and nuclear DNA markers to reconstruct human evolution

Lynn B. Jorde,[1]* Michael Bamshad,[2] and Alan R. Rogers[3]

Summary

Molecular genetic data have greatly improved our ability to test hypotheses about human evolution. During the past decade, a large amount of nuclear and mitochondrial data have been collected from diverse human populations. Taken together, these data indicate that modern humans are a relatively young species. African populations show the largest amount of genetic diversity, and they are the most genetically divergent population. Modern human populations expanded in size first on the African continent. These findings support a recent African origin of modern humans, but this conclusion should be tempered by the possible effects of factors such as gene flow, population size differences, and natural selection. BioEssays 20:126–136, 1998. © 1998 John Wiley & Sons, Inc.

INTRODUCTION

For millennia, humans have pondered their origins. This is reflected in the many stories that have evolved to explain when and where modern humans first appeared. The Gilgamesh Epic of the ancient Babylonians and the Judeo-Christian Garden of Eden are but two examples. Indeed, creation stories seem to be a nearly universal feature of human cultures.

The most direct account of our past comes from the fossil record. Skeletal remains have been instrumental in establishing the evolution of archaic humans in Africa, and they also have provided important information about the appearance of modern *Homo sapiens.* However, the fossil record is spotty, and many critical gaps remain.

[1]Department of Human Genetics University of Utah.
[2]Department of Pediatrics, University of Utah Health Sciences Center.
[3]Department of Anthropology, University of Utah, Salt Lake City, UT.
Contract grant sponsor: NSF; Contract grant numbers: DBS-9310105 and SBR-9514733; Contract grant sponsor: NIH; Contract grant number: RR-00064.
*Correspondence to: Dr. Lynn B. Jorde, Department of Human Genetics, University of Utah, Salt Lake City, UT 84112.

Genetic data offer another way of viewing human evolution. As soon as it became possible to measure genetic variation in humans, biologists began to use the data to assess the affinities and origins of populations. Major demographic events (population migrations, bottlenecks, and expansions) leave imprints, in the form of altered gene frequencies, on the collective human genome. Because these imprints are transmitted to succeeding generations, the modern human genome contains an indelible record of our evolutionary past. If enough genetic variation is examined in enough individuals, our evolutionary history should be largely decipherable.

Until recently, evolutionary studies were limited by a paucity of useful genetic polymorphisms. Although blood groups and protein polymorphisms yielded invaluable insights about human population history, the systems were relatively few in number, often rather uniform among populations, and sometimes affected by natural selection. This situation has changed dramatically during the past decade or so, with the advent of literally thousands of new genetic systems that can be easily assayed by using the tools of molecular biology. In addition to their large number, these new systems have the advantage that they provide information about different aspects of the human evolutionary picture.

TYPES OF GENETIC DATA

The nuclear genome, excluding the Y chromosome, contains the great majority of human polymorphisms. These include

the "classical" systems (blood groups and protein polymorphisms) as well as single-nucleotide variants, usually ascertained as restriction fragment length polymorphisms (RFLPs). It is estimated that the mutation rate for single nucleotides is approximately 10^{-7} to 10^{-8} per generation.[1] The slow mutation rate of these polymorphisms provides a means of visualizing the ancient history of our species, but these polymorphisms are likely to be relatively uninformative about recent history. Another class of nuclear polymorphisms consists of tandemly repeated DNA sequences in which the repeated sequence is two to five base pairs in length. These are termed microsatellite systems, and the number of repeats in each system varies from one individual to the next. Thousands of microsatellite polymorphisms have now been identified in the human.[2] Their mutation rate is much higher than that of single nucleotides, approaching 10^{-3} per generation.[3] Because of this high mutation rate, microsatellites have the potential to provide information about recent evolutionary events. In particular, they can assign dates to events that have occurred since the origin of humans with greater accuracy than the slowly evolving RFLPs.

The mitochondrial genome offers a very different perspective on human evolution. Because mitochondrial DNA (mtDNA) is inherited only through the maternal cytoplasm, variation in mtDNA provides a record of the maternal lineage of our species. In contrast to autosomal DNA, there is no recombination in mtDNA. Hence, the difference between any two mitochondrial sequences represents only the mutations that have taken place since each sequence was derived from a common ancestor. The noncoding mitochondrial control region is especially rich in polymorphisms, and the mutation rate is estimated to be approximately 2 to 3×10^{-7} per nucleotide per generation.[4]

Just as mtDNA yields a record of maternal human lineages, Y chromosome DNA (exclusive of the recombining pseudoautosomal region) provides a record of the paternal lineage. It has been relatively difficult to identify polymorphisms in the Y chromosome, but a growing collection of Y-chromosome polymorphisms is now available.[5,6]

GENETIC DATA AND THE ORIGIN OF MODERN HUMANS

Genetic data have been used to address many questions about the history of our species. For example, gradual changes in gene frequencies across Europe support the hypothesis that agriculture was spread through the migration of agricultural peoples, not simply through the cultural diffusion of the technology itself.[7] Similar studies have addressed conflicting hypotheses about the origins of modern Europeans[8] and Native Americans.[9] Genetic studies of worldwide Jewish populations have shown that they are more similar to one another than to surrounding populations, indicating that the Jewish population has maintained a
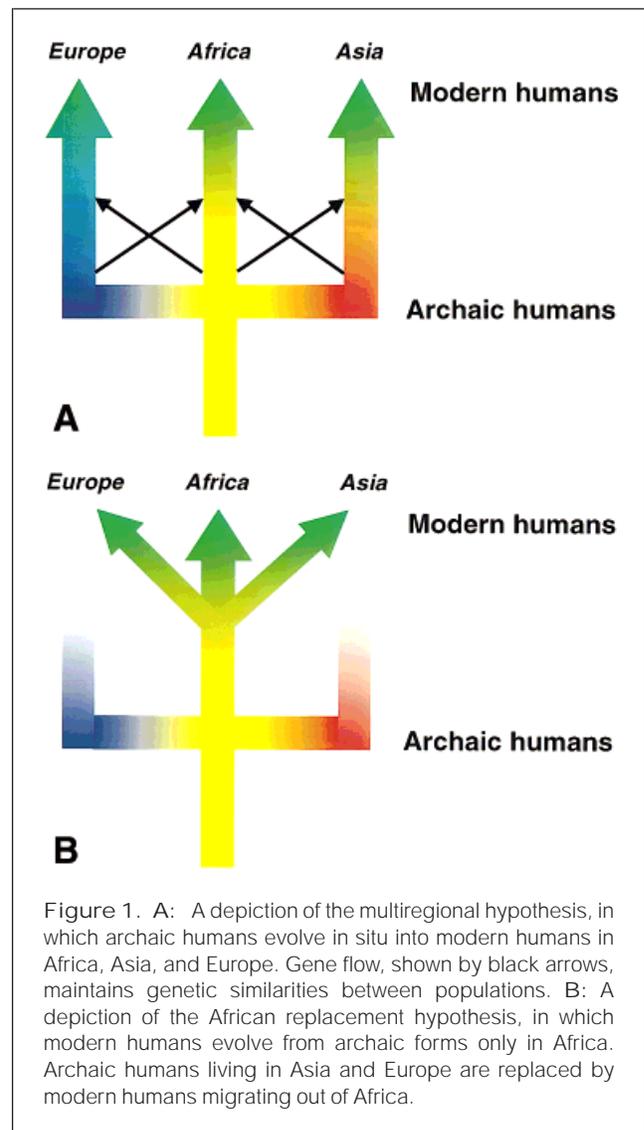

Figure 1. A: A depiction of the multiregional hypothesis, in which archaic humans evolve in situ into modern humans in Africa, Asia, and Europe. Gene flow, shown by black arrows, maintains genetic similarities between populations. **B:** A depiction of the African replacement hypothesis, in which modern humans evolve from archaic forms only in Africa. Archaic humans living in Asia and Europe are replaced by modern humans migrating out of Africa.

genetic identity since the Diaspora.[10] Classical studies of blood groups showed that the Romany Gypsies were genetically more similar to North Indians than to nearby European populations, verifying the Gypsies' oral tradition that they originated in India.[11] Many similar examples are summarized by Cavalli-Sforza et al.[12]

One of the most contentious (and possibly the most interesting) issues addressed by geneticists has been the origin of anatomically modern humans. Most current controversy revolves around two competing hypotheses. The "multiregional hypothesis" states that modern *Homo sapiens* evolved from more archaic forms over the course of a million years in several different locations in the Old World. The relative genetic homogeneity observed in the human species was maintained by natural selection and by gene flow between populations (Fig. 1a).The "African replacement"

hypothesis (also known as the "Garden of Eden" or "Noah's Ark" hypothesis) suggests that modern humans arose in Africa approximately 100,000 to 200,000 years ago and dispersed throughout the Old World to replace archaic human species completely (Fig. 1b). Many population geneticists support the African replacement hypothesis in its pure form[12–16] or in a variant form in which modern humans originated in Africa but may not have completely replaced archaic forms elsewhere.[17] Other geneticists question the African replacement hypothesis,[18,19] and some favor the multiregional hypothesis.[20,21] Many paleontologists support the multiregional hypothesis,[22,23] but there is also substantial support among paleontologists for the African replacement hypothesis[24,25] or a variation thereof.[26] These hypotheses have raised a number of important questions about human genetic variation. This review will address these questions and the ways in which genetic data have helped to answer them.

TABLE 1. F_{st} Values for the Human Population Divided Into Three Major Continental Groups (Africans, Asians, and Europeans)

Data type	F_{st}	Reference
Blood groups	0.16	[77]
Protein polymorphisms	0.11	[77]
RFLPs	0.11	[66]
Dinucleotide microsatellites*	0.11	[52]
Trinucleotide microsatellites	0.04	[78]
Tetranucleotide microsatellites	0.04	[66,67]
Microsatellites + RFLPs	0.15	[79]
Mitochondrial DNA (HVS1)	0.14	Bamshad et al., unpublished

*This study was based on eight worldwide populations rather than three continental populations, so the F_{st} value is relatively high.

How Do We Explain the Relative Lack of Genetic Diversity in Humans?

Although humans seem to vary a great deal in physical appearance, we are actually a genetically homogeneous species. Chimpanzees, our nearest evolutionary relatives, exhibit variation in 1 of 5 mitochondrial control region nucleotides, whereas humans exhibit variation in only 1 of 17.[27] Similarly, chimpanzees show considerably more variation in their nuclear genome than do humans,[28] and *Drosophila pseudoobscura* manifests approximately 10 times as much nucleotide variability as do humans.[29] Importantly, this lack of genetic variability in humans is compatible with the hypothesis that modern humans originated and/or expanded from a small population fairly recently.

There are two possible explanations for low nucleotide diversity in humans. Either humans have recently undergone a major bottleneck in population size, or the effective population size of the human species has been small for a long time. Some have argued that the sharing of alleles of the major histocompatibility complex (MHC) in humans and chimps argues against a human bottleneck, because such alleles would have been lost to genetic drift in a bottleneck.[17] Analyses of the β-globin and apolipoprotein C-II loci have produced similar conclusions.[19,30] However, a moderate bottleneck of longer duration would allow the persistence of such alleles.[31] This bottleneck corresponds to an effective size of 10,000 individuals, which is in accord with the estimates afforded by the apolipoprotein and β-globin data,[19,31] as well as mtDNA sequences,[32] Y chromosome data,[33] a series of nuclear DNA sequences,[34] and nuclear protein polymorphism data.[35] The MHC data argue for a larger effective population size of approximately 100,000,[17] but this analysis has been criticized because it fails to account adequately for the effects of natural selection and recombination.[36] Furthermore, even the MHC data are compatible with a reduction of the human population to an effective size of approximately 10,000 sometime during the Pleistocene.[34] Most genetic data, then, support a relatively small effective population size for humans, consistent with a moderate bottleneck and a low level of genetic diversity.

In addition to overall genetic diversity, it is useful to examine the distribution of genetic diversity within and between major continental populations (sometimes referred to as "races"). Table I presents estimates of the F_{st} statistic, which measures the proportion of genetic variation that can be attributed to between-population differences. For protein polymorphisms and nuclear RFLPs, approximately 10% of genetic variation exists between populations. The lower microsatellite figure is a consequence of the high mutation rate of these systems. High mutation rates increase within-population variance, thus decreasing the relative proportion of between-population variance.[37] The F_{st} statistic for mtDNA is somewhat larger than that of the nuclear systems, reflecting a more rapid between-population divergence rate as a result of the lower effective population size of mtDNA (i.e., for a mating pair, four different copies of each autosomal nucleotide can be transmitted to the offspring, whereas only one copy of a Y chromosome or mtDNA nucleotide is transmitted).
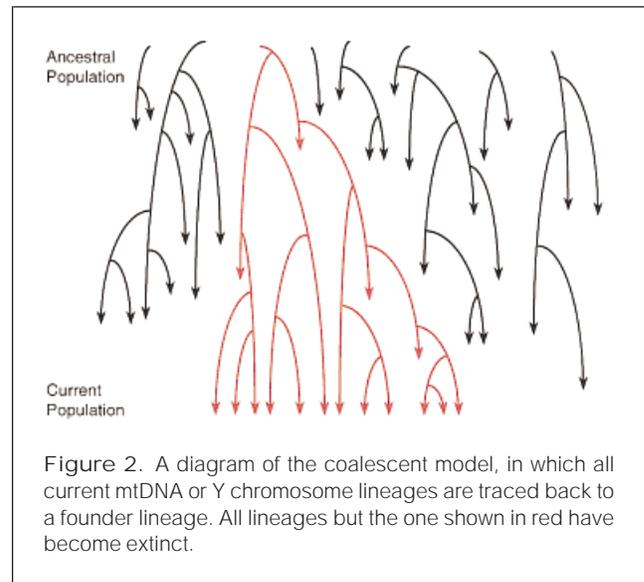
Despite some variability, all of these figures show that the great majority of genetic diversity exists within, rather than between, major continental populations. One interpretation of this result is that gene flow between the major continents has been sufficiently large to maintain genetic similarity (which could be consistent with the multiregional hypothesis). Another interpretation is that the major populations separated relatively recently, with little time for divergence to occur (which would support the replacement hypothesis).

How Old Is Our Most Recent Common Ancestor?

If we compare DNA sequences from two individuals, we can count the number of nucleotide differences between them. Suppose that the two sequences differ at three nucleotide positions. We can then infer that at least three mutations have taken place since the two DNA sequences were derived from a common ancestor who lived at some time in the past (note that this is a minimum estimate because of the possibility of multiple mutations at the same position). Because mutation is the most clock-like of the evolutionary processes, an estimate of the nucleotide mutation rate will allow us to infer the date of this common ancestor (assuming that natural selection is not affecting DNA sequence variation). Extending this principle to the human species, it is possible to sample a large number of humans, examine their DNA sequences for differences, and then work backward to an ancestor from whom all existing variation could have sprung (this is termed the "coalescent"). To understand this concept, it is important to keep in mind that some genetic variation is naturally lost in each generation because of the extinction of some lineages (Fig. 2). Eventually, all lineages but one will be lost. Lineage extinction will tend to occur more quickly, and the coalescence date will be more recent, in smaller populations than in larger ones.

Because of its exclusively maternal inheritance, lack of recombination, and relatively high mutation rate, mtDNA lends itself well to estimating the age of the most recent common (female) ancestor. The coalescent principle can be used to estimate the number of generations that separate this individual from the present. To assign a date, the rate of mutation per generation must be estimated. The mtDNA molecular clock has been "calibrated" by measuring the genetic differences between populations or species that have been separated by a known length of time (e.g., chimps and humans diverged between five and nine million years ago, providing a yardstick with which to estimate the mtDNA mutation rate). Application of the coalescent approach has led to estimates of approximately 100,000 to 200,000 years as the age of the female ancestor common to all living humans[4,38,39] (more precisely, this represents the age of the *mitochondrial molecule* ancestral to all presently existing mitochondrial molecules).

Coalescence dates also have been estimated from Y chromosome data. By using several Y-specific polymorphisms, Hammer[33] obtained a coalescence estimate of 188,000 years (with 95% confidence limits of 51,000 to 411,000 years). Dorit et al.[40] found no sequence variation in an intron of the Y-specific *ZFY* gene and inferred a coalescence time of 270,000 years (with 95% confidence limits of 0 to 800,000 years). The latter study has been criticized because of methodological flaws[41] and because this part of the gene may not be selectively neutral.[42] The

Figure 2. A diagram of the coalescent model, in which all current mtDNA or Y chromosome lineages are traced back to a founder lineage. All lineages but the one shown in red have become extinct.

possibility that Y chromosome variation has been reduced by periodic selective "sweeps" must be kept in mind.[43] It is interesting, however, that coalescence estimates based on male-specific Y chromosome lineages are in rough agreement with those based on mtDNA. This result argues in favor of selective neutrality for the mtDNA and Y chromosome data, because it is unlikely that natural selection would affect two independent genetic systems in exactly the same way.

In the nuclear genome, recombination (the exchange of homologous chromosome material during meiosis) complicates coalescence estimates because differences in individual nuclear DNA sequences can arise both from mutation and recombination. Thus, coalescence estimates for autosomal genes are based on single sequences unlikely to undergo significant recombination. An analysis of human β-globin sequence yielded a coalescence estimate of approximately 800,000 years.[19] Because the effective size of autosomal DNA is 4 times higher than that of the Y chromosome and mtDNA, nuclear lineages are expected to take 4 times as long to become extinct. Hence, this figure is consistent with the estimates based on Y chromosome and mtDNA data.[32]

The meaning of these coalescence estimates has been misinterpreted in many ways. Perhaps owing to the popular designation of the female common ancestor as "mitochondrial Eve," some have assumed that only one woman lived 200,000 years ago. Figure 2 shows why this is incorrect: many other women doubtless lived at the same time; their mitochondrial lineages simply became extinct. What remains today are the mtDNA descendants, modified by mutation, of the woman whose mitochondrial lineage did not become

extinct. The same interpretation can be applied to the Y chromosome coalescence figure.

Another misconception is that the coalescence date represents the time of emergence of modern humans. It does not, for a population of modern humans could have existed well before the coalescence date. Also, the mixture of anatomically modern human and archaic human populations could extend the convergence date backward in time (indeed, this is a reasonable interpretation of the β-globin coalescence date). In any case, the coalescence date itself must be taken with a large grain of salt, because mtDNA and the Y chromosome, as single nonrecombining sequences, are subject to a large degree of stochastic error.[44] Incorporation of this error leads to estimates of the mtDNA coalescence as early as 700,000 years ago.[45] In addition, the date of the human-chimp divergence is debatable, leading to variability in the mutation rate estimate. It also has been argued that the inclusion of nucleotide transitions in the mtDNA estimates may lead to error, because the transitions tend to "saturate" over time (i.e., a C may change to a T and then back again to a C, underestimating genetic divergence). When only mtDNA transversions (which occur less frequently) are used, coalescence estimates of 400,000 to 800,000 years are obtained.[46] A further difficulty is that one commonly used estimator of the coalescence date[21] is valid only in the special case in which only two individuals are compared; otherwise, this method yields inflated estimates, with confidence intervals that are too large.[47] Finally, direct estimates of mtDNA mutation rates, in which mtDNA sequences were compared in mothers and their offspring, have indicated that the rate may be 10 to 20 times higher than estimates based on phylogenetic studies.[48,49] This mutation rate could yield a coalescence time of only 10,000 years. However, it is likely that most of these mutations do not persist for more than a few generations, because they are lost through the bottlenecks that occur as mitochondrial molecules are replicated each generation.[48] They would then have little effect on coalescence date estimates.

Considering these uncertainties, what can the coalescence date tell us about the origin of modern humans? More important than the age of the most recent common ancestor is the information the coalescence time gives us about the effective size of the human population. For mtDNA, the coalescence time in generations, t, is directly related to the effective female population size (N_f) as $t = 2N_f$. The preponderance of mtDNA estimates suggests a coalescence date of 100,000 to 200,000 years ago, and this is now bolstered by similar estimates from Y chromosome data, as well as a compatible β-globin figure. This tells us that the effective population size of humans has been relatively small throughout most of our history: only a few thousand breeding females.[32] If the population had been significantly larger, more lineages would have existed, and the elimination of all

but one would have taken much longer than 100,000 to 200,000 years. It has been argued that such a small population would be unable to cover most of the Old World and exchange enough migrants to maintain relative genetic uniformity.[31,32,50] Thus, a small population size is inconsistent with the multiregional hypothesis.
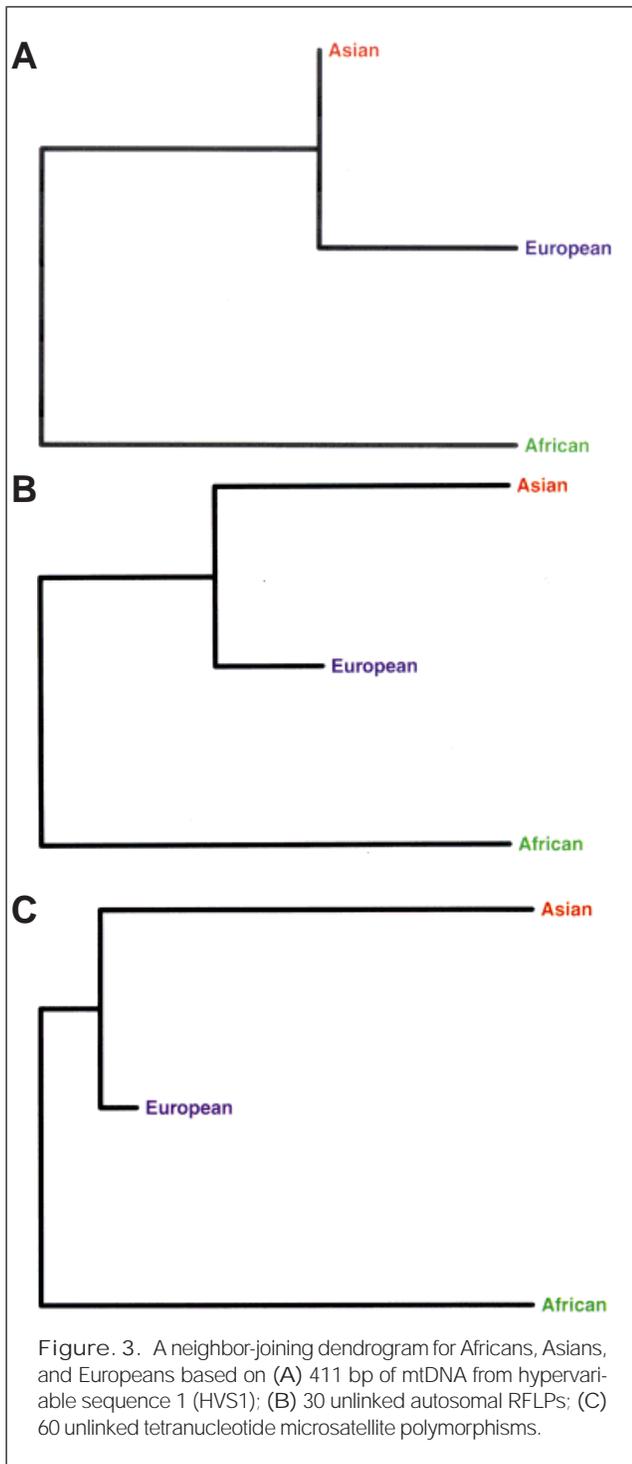
How Similar Are the Major Human Populations to One Another?

Some early blood group studies of the major human continental populations (Africans, Asians, and Europeans) showed that Asians were the most divergent population, with Europeans and Africans showing greater similarity to one another.[51] However, these results appear to have been caused by the inclusion of systems that were affected by natural selection. Selectively neutral polymorphisms are more likely to provide accurate portrayals of population relationships and history. Figure 3 shows a series of dendrograms (diagrams that summarize the extent of genetic divergence between populations) estimated from more appropriate genetic systems. For mtDNA, autosomal microsatellites, and autosomal RFLPs, Africans are the most divergent population. Similar results have been obtained from dinucleotide microsatellites,[16,52] blood groups, and protein polymorphisms,[44,53,54] and nuclear as well as mtDNA RFLPs from other populations.[19,38,55,56] Most estimates of the date of divergence of African and non-African populations are in the neighborhood of 100,000 to 150,000 years or so,[4,51,57,58] although a recent estimate of only 15,000 years was obtained from a single minisatellite locus.[59]

Our confidence in the divergence patterns shown in these dendrograms, or "trees," can be increased by adding a "root." If a root can be placed reliably, it orients the dendrogram to show which population split took place first. Statistical methods can be used to root a dendrogram based only on the populations under study, but it is usually preferable to define a root by using an outgroup population or species. The outgroup is genetically distant from all of the populations under analysis, and the position at which it joins the remainder of the dendrogram is the root. Chimpanzees are typically used as an outgroup species when worldwide human populations are analyzed.

An attempt to root a human mtDNA dendrogram placed the root within African populations,[39] but this analysis was correctly criticized on methodological grounds.[21] Nevertheless, a subsequent reanalysis of the mtDNA data using more appropriate methods again placed the root within African populations.[60] In addition, analyses of several large nuclear data sets consistently placed the root between Africans and other populations, indicating that the first major split separates these two groups of humans.[61]

One difficulty in using a chimpanzee outgroup is that rapidly mutating systems, such as nuclear microsatellites

Figure. 3. A neighbor-joining dendrogram for Africans, Asians, and Europeans based on **(A)** 411 bp of mtDNA from hypervariable sequence 1 (HVS1); **(B)** 30 unlinked autosomal RFLPs; **(C)** 60 unlinked tetranucleotide microsatellite polymorphisms.

inserted into the nuclear genome, this sequence mutated at the much slower nuclear rate (estimated to be at least 10 times slower than the mtDNA rate). An analysis of this inserted sequence again placed the root within African populations.[62]

Increased African divergence, in combination with a relatively recent African/non-African separation date, is often regarded as evidence for a scenario in which a small portion of the African population emigrated to colonize the rest of the world approximately 100,000 years ago.[13,58] However, greater African divergence could be simply the result of genetic drift and gene flow patterns. Furthermore, we must guard against too literal an interpretation of dendrograms, particularly in terms of divergence date estimation. Human evolution has not simply been a series of population fissions followed by isolation.[14] A more accurate picture of human evolution would look something like a trellis, with multiple connections between branches. These connections represent gene flow between populations, which causes them to become genetically more similar to one another and thus produces underestimates of divergence dates. Because gene flow has occurred at unknown rates between the major populations, divergence dates are only approximate lower limits. Furthermore, a larger African effective population size could mimic the observed divergence pattern even if all three populations had diverged from one another at the same time.[18,19] Thus, the conclusions that Africans are the most divergent population and that the separation from non-Africans may have occurred approximately 100,000 years ago are consistent with an African origins hypothesis. However, uncertainties about gene flow patterns and population size differences caution against placing too much reliance on these conclusions.

Which Major Population Is the Most Diverse, and What Does This Mean?

Evolutionary theory predicts that an older, "source" population will typically have greater genetic diversity than a population derived more recently from it.[63] Thus, it is not surprising that genetic diversity has been compared extensively among the major continental populations in an attempt to determine whether one continent contains the source population of modern humans.

These studies have yielded somewhat contradictory results, depending on the category of polymorphisms analyzed. Blood groups, protein polymorphisms, and RFLPs usually show greatest variability in Europeans. However, this is probably because these polymorphisms were first ascertained in European populations on the basis of their variability.[64,65] This ascertainment bias does not apply to nuclear or mitochondrial DNA sequence that is assayed uniformly in populations, nor does it have much effect on highly polymorphic microsatellite systems.[65] As Table II shows, mtDNA

and mtDNA, can reach mutational saturation in less time than the five million or so years that separate humans and chimpanzees. An interesting solution to this problem involves the use of a mitochondrial sequence that was inserted into human chromosome 11 sometime in the past. Once

TABLE 2. Heterozygosity (Gene Diversity) Values for Africans, Asians, and Europeans*

Data type	African	Asian	European	Reference
Blood groups	0.164	0.145	**0.179**	[80]
Protein poly-morphisms	0.179	0.164	**0.186**	[80]
RFLPs	0.297	0.327	**0.379**	[16]
RFLPs	0.322	0.377	**0.432**	[66]
Dinucleotide microsatel-lites	**0.807**	0.685	0.730	[16]
Trinucleotide microsatel-lites	**0.850**	0.820	0.807	[78]
Tetranucleo-tide micro-satellites	**0.760**	0.697	0.732	[66,67]
Tetra- and tri-nucleotide microsatel-lites	**0.748**	0.698	0.733	[81]
Mitochondrial DNA (HVS1)	**0.023**	0.015	0.010	Bamshad et al., unpublished

*The highest population value for each data type is shown in bold type.

Figure 4. An unrooted network based on mtDNA (HVS1) from 15 populations living in Africa (green), Asia (red), and Europe (blue). Genetic distances were estimated by using the Kimura 2-parameter model, and the neighbor-joining method was used to make the network.

sequence, Y chromosome polymorphisms, and highly polymorphic microsatellites all show greater diversity in Africa than in Asian or Europe, although only some of these studies yielded statistically significant results. Elevated African diversity is particularly apparent in the mtDNA control region, as seen in a network diagram of African, Asian, and European populations (Fig. 4). African populations have by far the longest branch lengths in this network, reflecting increased genetic diversity within and between these populations.

Elevated African diversity, although consistent with an African origin, does not necessarily prove that modern humans originated in Africa. It is possible that modern humans could have originated elsewhere (e.g., Asia) but that a subsequent population bottleneck reduced Asian diversity below that of Africa.[31] It is also possible that higher African diversity is simply the result of greater effective population size in Africa.[18,19]

Additional information, however, can be gained by examining the distribution of haplotypes (groups of polymorphisms closely linked on the same autosome or on the Y or mitochondrial genomes) in populations. In general, such studies show that a greater diversity of haplotypes is found in Africa. More importantly, the haplotypes found outside Africa are a *subset* of those found in Africa.[5,33,57,60] This is most consistent with an African origin of modern humans, followed by a reduction in haplotype diversity as subsets of the
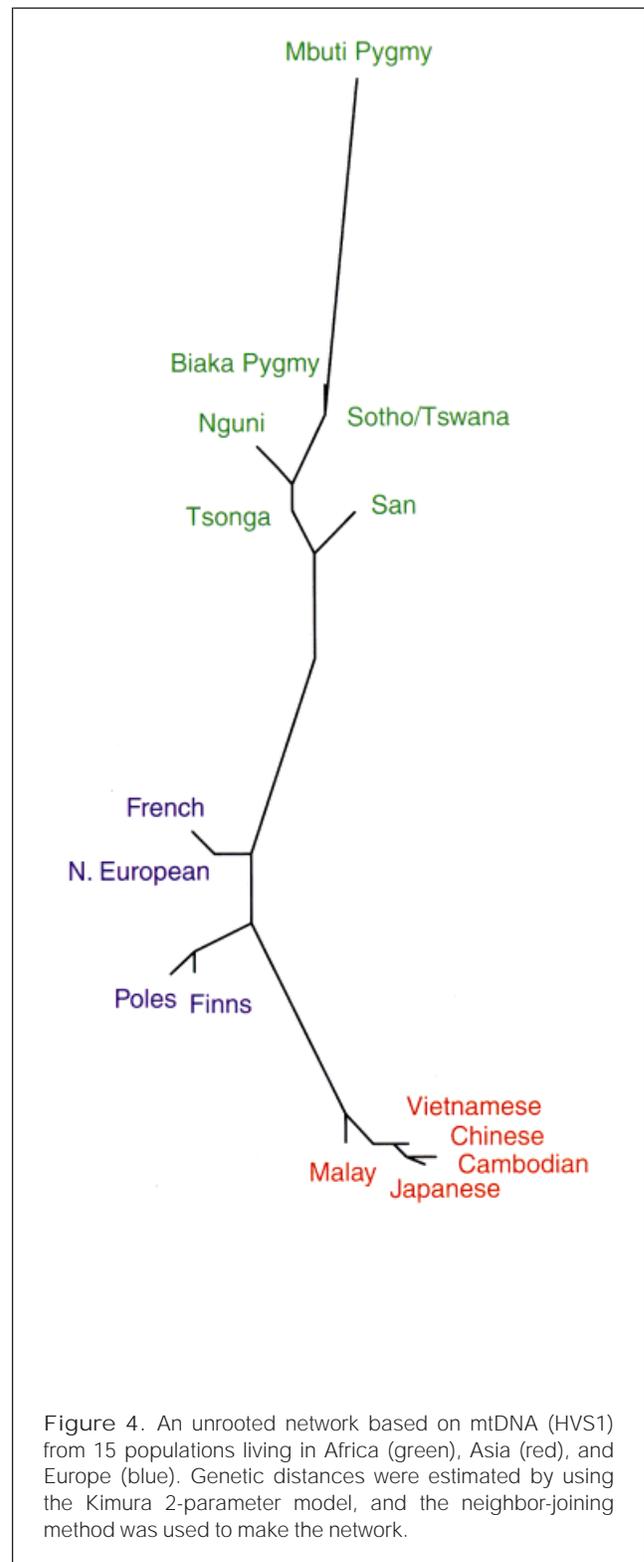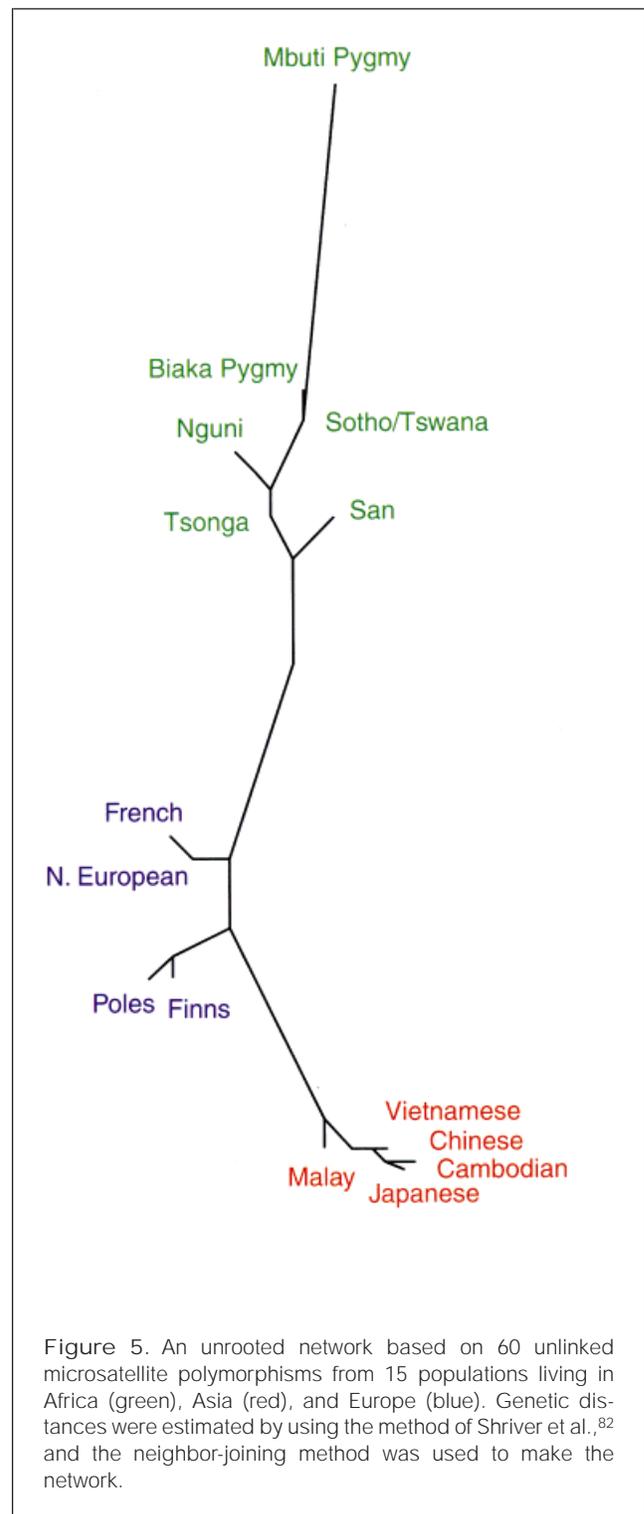
African population left to colonize Asia and Europe. If modern humans had originated, say, in Asia and then undergone a bottleneck in situ, one would still expect to find many Asian haplotypes that are not a subset of those from Africa. Thus, it is difficult to explain this pattern without invoking an African origin of modern humans.
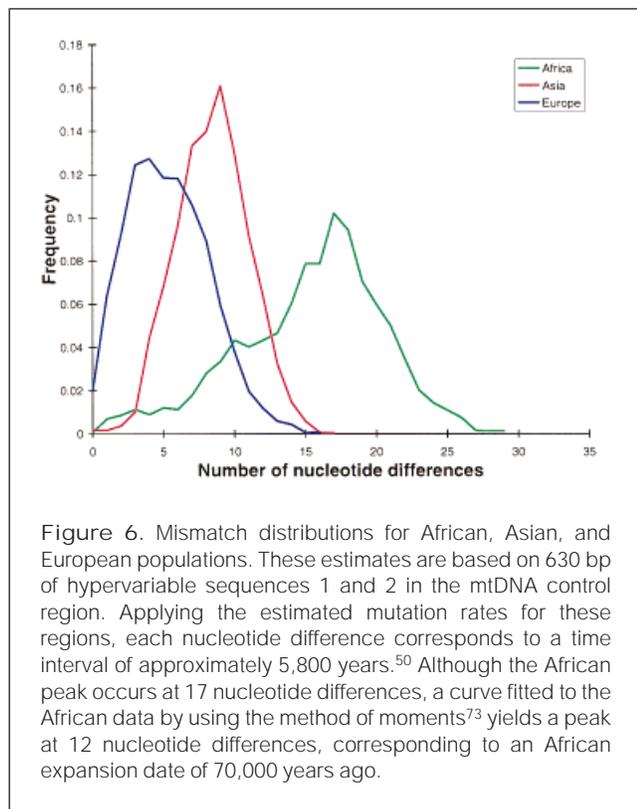
It is interesting to note that African mtDNA diversity is several times greater than that of other populations, whereas African nuclear diversity is only slightly greater (Table II). This difference also is seen by comparing a dendrogram based on microsatellites (Fig. 5) with the mtDNA dendrogram in Figure 4. There are a number of possible explanations for this difference.[66] One possibility is that Africa underwent a population expansion earlier than Europe or Asia. Because the mtDNA effective size is one-fourth that of nuclear genes, it responds more quickly to such events, resulting in a larger ratio of African to non-African mtDNA diversity.[67] A second possibility is that natural selection has reduced mtDNA variability in non-African populations, perhaps as a result of adaptation to altered climatic conditions. Because there is no recombination in the mitochondrial genome, natural selection at any of the vital mitochondrial genes would affect variation throughout the genome, including the commonly studied control region. Because several studies support a departure from selective neutrality for mtDNA,[68–70] this possibility deserves serious attention. Further comparisons among mtDNA, Y chromosome, and autosomal data will likely provide a definitive answer.

When Did Human Populations Expand in Size?

Genetic data indicate that the human population once numbered only 10,000 or so breeding individuals. If so, it must have expanded significantly in size at some point in history. The coalescent process, described above, can be applied to mtDNA data to estimate the date of this expansion. To see how, consider that a random pair of individuals are less likely to be offspring of the same mother in a large population than in a small one. Consequently, genealogical connections between individuals tend to be long in large populations and short in small ones. When a small population expands in size, the result is often a star-like genealogy: lines of descent remain distinct back to the point where the population was small and then coalesce rapidly. Thus, coalescent events are concentrated in the generations just before the expansion. The result is that many pairs of individuals in the modern sample would trace their common ancestor to an individual who lived in the generations just prior to the expansion. If we then examine the number of nucleotide differences among all pairs of individuals in a modern sample, we will find that the most common number of differences corresponds to the number of mutations separating individuals from common ancestors who lived

Figure 5. An unrooted network based on 60 unlinked microsatellite polymorphisms from 15 populations living in Africa (green), Asia (red), and Europe (blue). Genetic distances were estimated by using the method of Shriver et al.,[82] and the neighbor-joining method was used to make the network.

just before the expansion. This produces a ''peak'' in the distribution of nucleotide differences.[31,71–73] The timing of the population expansion is indicated by the peak in the distribution, and the approximate expansion date can be estimated

Figure 6. Mismatch distributions for African, Asian, and European populations. These estimates are based on 630 bp of hypervariable sequences 1 and 2 in the mtDNA control region. Applying the estimated mutation rates for these regions, each nucleotide difference corresponds to a time interval of approximately 5,800 years.[50] Although the African peak occurs at 17 nucleotide differences, a curve fitted to the African data by using the method of moments[73] yields a peak at 12 nucleotide differences, corresponding to an African expansion date of 70,000 years ago.

by using known mtDNA mutation rates. In this remarkable fashion, the signature of an event that took place tens of thousands of years ago can be read from modern genetic data.

Figure 6 shows the distributions of mtDNA differences (termed ''mismatch distributions'') for African, Asian, and European populations. The African mismatch peak occurs furthest to the right, with a mode at approximately 17 nucleotide differences. This corresponds to an expansion date of roughly 99,000 years ago (see Fig. 6 legend). The Asian peak, occurring further to the left, corresponds to an expansion date of approximately 52,000 years ago. The European peak, occurring leftmost in the plot, indicates an expansion date of only a little more than 23,000 years ago. The European and Asian expansion dates are intriguing in that they coincide roughly with the replacement of Mousterian cultural assemblages by the more elaborate and refined Upper Paleolithic technology.[74]

Because of the large stochastic error associated with mtDNA-based estimates, these are very approximate dates. Some analyses of nuclear data have not supported a population expansion,[19,34] but it is questionable whether a relatively recent demographic event could be detected with small, slowly mutating nuclear DNA sequences. Analysis of 60 rapidly mutating nuclear microsatellite polymorphisms, on the other hand, corroborates the mtDNA expansion time

estimates,[67,75] Moreover, an earlier African expansion is consistent with the difference in mtDNA and nuclear diversity ratios discussed above.

CONCLUSIONS

The findings reviewed in this article show that human populations exhibit remarkably little genetic diversity and that most of this diversity is allocated within, rather than between, major continental populations. Multiple types of genetic systems (autosomal, mitochondrial, and Y-specific) are in broad and reassuring agreement that modern humans are a relatively young population, that Africans are genetically the most diverse group of humans, and that the largest genetic divergence among Old World populations occurs between Africans and non-Africans. Estimates of population expansion dates indicate that modern human populations expanded in size first in Africa. These various lines of evidence are most consistent with a recent African origin of modern humans, but other possibilities, involving the effects of natural selection, effective population size, and gene flow patterns, cannot yet be excluded.

Although the majority of evidence appears to favor an African origin of modern humans, it is less clear that these modern humans completely replaced non-African populations. There is now limited evidence, in the form of β-globin alleles[19] and Y chromosome haplotypes[5] for ancient non-African contributions to the present-day gene pool. Thus, neither the African replacement model (in its pure form) nor the multiregional model is fully consistent with the data. An African origin, with some mixing of populations, appears to be the most likely possibility. More data, particularly in the form of nuclear DNA sequences, will be needed to resolve this issue.

Although much useful genetic information has been collected, assimilation and habitat destruction are rapidly eliminating many unique human populations. A primary goal should be to reverse this process whenever possible. A second goal, incumbent upon the genetics community, is a prompt and culturally sensitive collection of genetic data from these populations.[76] Technological developments, spawned by the Human Genome Project, will then enable us to assemble large quantities of autosomal, mitochondrial, and Y-specific data on a truly representative sample of the human species. Such data are needed for definitive tests of the replacement, multiregional, and other hypotheses. Paleontological findings, which directly provide a temporal perspective that genetic data cannot replace, must also be incorporated into hypothesis testing. Ultimately, these data will tell us the story of our origins.

REFERENCES

1 Nei, M. (1987) *Molecular Evolutionary Genetics.* New York: Columbia University Press.

2 Dib, C., et al. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**:152–154.

3 Weber, J.L., and Wong, C. (1993) Mutation of human short tandem repeats. *Hum Mol Genet* **2**:1123–1128.

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

5 Hammer, M.F., et al. (1997) The geographic distribution of Y chromosome variation. *Genetics* **145**:787–805.

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

7 Sokal, R.R., Oden, N.L., and Wilson, C. (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* **351**:143–145.

8 Piazza, A. (1993) Who are the Europeans? *Science* **260**:1767–1769.

9 Forster, P., Harding, R., Torroni, A., and Bandelt, H.-J. (1996) Origin and evolution of Native American mtDNA variation: A reappraisal. *Am J Hum Genet* **59**:935–945.

10 Livshits, G., Sokal, R.R., and Kobyliansky, E. (1991) Genetic affinities of Jewish populations. *Am J Hum Genet* **49**:131–146.

11 Boyd, W.C. (1963) Four achievements of the genetical method in physical anthropology. *Am J Physiol Anthropol* **65**:243–252.

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

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

14 Harpending, H., Relethford, J., and Sherry, S. (1995) Methods and models for understanding human diversity. In A.J. Boyce and C.G.N. Mascie-Taylor (eds): *Molecular Biology and Human Diversity.* Cambridge: Cambridge University Press, pp. 283–299.

15 Stoneking, M. (1993) DNA and recent human evolution. *Evol Anthrop.* **2**:60–73.

16 Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., and Cavalli-Sforza, L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**:455–457.

17 Ayala, F.J. (1995) The myth of Eve: Molecular biology and human origins. *Science* **270**:1930–1936.

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

19 Harding, R.M., et al. (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* **60**:772–789.

20 Spuhler, J.N. (1993) Population genetics and evolution in the genus Homo in the last two million years. In C.F. Sing and C.L. Hanis (eds): *Genetics of Cellular, Individual, Family, and Population Variability.* New York: Oxford University Press, pp. 262–297.

21 Templeton, A.R. (1993) The "Eve" hypothesis: A genetic critique and reanalysis. *Am Anthropol* **95**:51–72.

22 Wolpoff, M.H., Thorne, A.G., Smith, F.H., Frayer, D.W., and Pope, G.G. (1994) Multiregional evolution: A world-wide source for modern human populations. In M.H. Nitecki and D.V. Nitecki (eds): *Origins of Anatomically Modern Humans.* New York: Plenum Press, pp. 175–199.

23 Frayer, D.W., Wolpoff, M.H., Thorne, A.G., Smith, F.H., and Pope, G.G. (1993) Theories of modern human origins: The paleontological test. *Am Anthropol* **95**:14–50.

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

25 Wood, B. (1996) Human evolution. *Bioessays* **18**:945–954.

26 Aiello, L.C. (1993) The fossil evidence for modern human origins in Africa: A revised view. *Am Anthropol* **95**:73–96.

27 Morin, P.A., Moore, J.J., Chakraborty, R., Jin, L., Goodall, J., and Woodruff, D.S. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**:1193–1201.

28 Crouau-Roy, B., Service, S., Slatkin, M., and Freimer, N. (1996) A fine-scale comparison of the human and chimpanzee genomes: Linkage, linkage disequilibrium and sequence analysis. *Hum Mol Genet* **5**:1131–1137.

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

30 Xiong, W., et al. (1991) No severe bottleneck during human evolution: Evidence from two apolipoprotein C-II deficiency alleles. *Am J Hum Genet* **48**:383–389.

31 Rogers, A.R., and Jorde, L.B. (1995) Genetic evidence on the origin of modern humans. *Hum Biol* **67**:1–36.

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

33 Hammer, M.F. (1995) A recent common ancestry for human Y chromosomes. *Nature* **378**:376–378.

34 Takahata, N., Satta, Y., and Klein, J. (1995) Divergence time and population size in the lineage leading to modern humans. *Theor Popul Biol* **48**:198–221.

35 Nei, M., and Graur, D. (1984) Extent of protein polymorphism and the neutral mutation theory. *Evol Biol* **17**:73–118.

36 Hickson, R.E., and Cann, R.L. (1997) Mhc allelic diversity and modern human origins. *J Mol Evol* **45**:589–598.

37 Jin, L., and Chakraborty, R. (1995) Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* **74**:274–285.

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

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

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

41 Fu, Y.-X., and Li, W.-H. (1996) The age of the common ancestor of human males estimated from ZFY intron sequence data. *Science* **272**:1356–1357.

42 Burrows, W., and Ryder, O.A. (1997) Y-chromosome variation in great apes. *Nature* **385**:125–126.

43 Whitfield, L.S., Sulston, J.E., and Goodfellow, P.N. (1995) Sequence variation of the human Y chromosome. *Nature* **378**:379–380.

44 Nei, M., and Livshits, G. (1989). Genetic relationships of Europeans, Asians and Africans and the origin of modern Homo sapiens. *Hum Hered* **39**:276–281.

45 Nei, M. (1992) Age of the common ancestor of human mitochondrial DNA. *Mol Biol Evol* **9**:1176–1178.

46 Wills, C. (1995) When did Eve live? An evolutionary detective story. *Evolution* **49**:593–607.

47 Fu, Y.-X., and Li, W.-H. (1997) Estimating the age of the common ancestor of a sample of DNA sequences. *Mol Biol Evol* **14**:195–199.

48 Parsons, T.J., et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* **15**:363–367.

49 Howell, N., Kubacka, I., and Mackey, D.A. (1996) How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* **59**:501–509.

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

51 Nei, M., and Roychoudhury, A.K. (1982) Genetic relationship and evolution of human races. *Evol Biol* **14**:1–59.

52 Deka, R., et al. (1995) Population genetics of dinucleotide (dC-dA)n.(dG-dT)n polymorphisms in world populations. *Am J Hum Genet* **56**:461–474.

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

54 Cavalli-Sforza, L.L., Piazza, A., Menozzi, P., and Mountain, J. (1988) Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci USA* **85**:6002–6006.

55 Long, J.C., Chakravarti, A., Boehm, C.D., Antonarakis, S., and Kazazian, H.H. (1990) Phylogeny of human β-globin haplotypes and its implications for recent human evolution. *Am J Physiol Anthropol* **81**:113–130.

56 Bowcock, A.M., et al. (1991) Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc Natl Acad Sci USA* **88**: 839–843.

57 Tishkoff, S.A., et al. (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**:1380–1387.

58 Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L., and Feldman, M.W. (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* **92**:6723–6727.

59 Armour, J.A.L., et al. (1996) Minisatellite diversity supports a recent African origin for modern humans. *Nat Genet* **13**:154–160.

60 Penny, D., Steel, M., Waddell, P.J., and Hendy, M.D. (1995) Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens. Mol Biol Evol* **12**:863–882.

61 Nei, M., and Takezaki, N. (1996) The root of the phylogenetic tree of human populations. *Mol Biol Evol* **13**:170–177.

62 Zischler, H., Geisert, H., von Haeseler, A., and Paabo, S. (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**:489–492.

63 Mayr, E. (1970) *Populations, Species, and Evolution.* Cambridge: Harvard University Press.

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

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

66 Jorde, L.B., et al. (1995) Origins and affinities of modern humans: A comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* **57**: 523–538.

67 Jorde, L.B., et al. (1997) Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA* **94**:3100–3103.

68 Nachman, M.W., Brown, W.M., Stoneking, M., and Aquadro, C.F. (1996) Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* **142**:953–963.

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

70 Merriwether, D.A., et al. (1991) The structure of human mitochondrial DNA variation. *J Mol Evol* **33**:543–555.

71 Di Rienzo, A., and Wilson, A.C. (1991) Branching pattern in the evolutionary tree for human mitochondr. *Proc Natl Acad Sci USA* **88,** 1597–1601.

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

73 Rogers, A.R. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* **49**:608–615.

74 Klein, R.G. (1992) The archeology of modern human origins. *Evol Anthropol* **1**:5–14.

75 Rogers, A.R., and Jorde, L.B. (1997) Unpublished data.

76 Cavalli-Sforza, L.L., Wilson, A.C., Cantor, C.R., Cook-Deegan, R.M., and King, M.-C. (1991) Call for a worldwide survey of human genetic diversity: A vanishing opportunity for the Human Genome Project. *Genomics* **11**:490–491.

77 Nei, M., and Livshits, G. (1990) Evolutionary relationships of Europeans, Asians, and Africans at the molecular level. In N. Takahata and J.F. Crow (eds): *Population Biology of Genes and Molecules.* Tokyo: Baifukan, pp. 251–265.

78 Watkins, W.S., Bamshad, M.J., and Jorde, L.B. (1995) Population genetics of trinucleotide repeat polymorphisms. *Hum Mol Genet* **4**:1485–1491.

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

80 Nei, M., Livshits, G., and Ota, T. (1993) Genetic variation and evolution of human populations. In C.F. Sing and C.L. Hanis (eds): *Genetics of Cellular, Individual, Family, and Population Variability.* New York: Oxford University Press, pp. 239–252.

81 Deka, R., Shriver, M.D., Yu, L.M., Ferrell, R.E., and Chakraborty, R. (1995) Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* **16**:1659–1664.

82 Shriver, M.D., Jin, L., Boerwinkle, E., Deka, R., Ferrell, R.E., and Chakraborty, R. (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* **12**:914–920.