Improved Analyses of Human mtDNA Sequences Support a Recent African Origin for *Homo sapiens*

David Penny, Mike Steel,† Peter J. Waddell,* and Michael D. Hendy‡*

*School of Biological Sciences, Massey University; †Department of Mathematics and Statistics, University of Canterbury; and ‡Department of Mathematics, Massey University

New quantitative methods are applied to the 135 human mitochondrial sequences from the Vigilant et al. data set. General problems in analyzing large numbers of short sequences are discussed, and an improved strategy is suggested. A key feature is to focus not on individual trees but on the general "landscape" of trees. Over 1,000 searches were made from random starting trees with only one tree (a local optimum) being retained each time, thereby ensuring optima were found independently. A new tree comparison metric was developed that is unaffected by rearrangements of trees around many very short internal edges. Use of this metric showed that downweighting hypervariable sites revealed more evolutionary structure than studies that weighted all sites equally. Our results are consistent with convergence toward a global optimum. Crucial features are that the best optima show very strong regional differentiation, a common group of 49 African sequences is found in all the best optima, and the best optima contain the 16 !Kung sequences in a separate group of San people. The other 86 sequences form a heterogeneous mixture of Africans, Europeans, Australopapuans, and Asians. Thus all major human lineages occur in Africa, but only a subset occurs in the rest of the world. The existence of these African-only groups strongly contradicts multiregional theories for the origin of *Homo sapiens* that require widespread migration and interbreeding over the entire range of *H. erectus*. Only when the multiregional model is rejected is it appropriate to consider the root, based on a single locus, to be the center of origin of a population (otherwise different loci could give alternative geographic positions for the root). For this data, several methods locate the root within the group of 49 African sequences and are thus consistent with the recent African origin of *H. sapiens*. We demonstrate that the time of the last common ancestor cannot be the time of major expansion in human numbers, and our results are thus also consistent with recent models that differentiate between the last common ancestor, expansion out of Africa, and the major expansion in human populations. Such a two-phase model is consistent with a wide range of molecular and archeological evidence.

Introduction

The "out-of-Africa" (or "mitochondrial Eve") hypothesis for the origins of modern humans (*Homo sapiens sapiens*) is a bold idea that has attracted considerable attention since being proposed in its present form (Cann et al. 1987). The hypothesis can be considered in four largely independent parts (Di Rienzo and Wilson 1991; Wilson et al. 1991):

H1: the most recent common ancestor of *H. sapiens sapiens* lived about 200,000 yr ago;
H2: *H. sapiens sapiens* arose from *H. erectus* in a single region;
H3: Africa is the most probable region for this transition; and
H4: there was later a spread out from Africa, eventually replacing earlier *Homo* groups.

We have called these the "when," "who," "where," and "how" questions (Waddell and Penny 1995). These hypotheses were initially put forward to explain observations such as the higher genetic diversity in mitochondria from African populations (Greenberg et al. 1983; Johnson et al. 1983), the evolutionary tree derived from RFLP (restriction fragment length polymorphisms) data (Cann et al. 1987), and the diversity among nuclear allele frequencies (Nei and Roychoudhury 1982). The hypotheses received support from an independent analysis of fossil and subfossil human remains (Stringer and Andrews 1988), particularly from the lack of objective evidence for intermediates between Neanderthals and modern *H. sapiens*. More recently, mitochondrial DNA (mtDNA) sequences from 135 individ-

Key words: evolutionary trees, great deluge algorithm, *Homo sapiens*, human evolution, mitochondrial DNA, out-of-Africa hypothesis.

Address for correspondence and reprints: David Penny, School of Biological Sciences, Massey University, Palmerston North, New Zealand. E-mail: D.Penny@massey.ac.nz.

Mol. Biol. Evol. 12(5):863–882. 1995.

0737-4038/95/1205-0015$02.00

uals were suggested to support the overall hypothesis (Vigilant et al. 1991).

However, these more recent results (Vigilant et al. 1991) have led to a variety of criticisms (Goldman and Barton 1992; Hedges et al. 1992; Maddison et al. 1992; Templeton 1992, 1993; Thorne and Wolpoff 1992) particularly resulting from the difficulties of analyzing large numbers of sequences. Some of the criticism has not specified which aspect of the overall hypothesis is disputed, but two themes are apparent. One line of questioning (Thorne and Wolpoff 1992) has been against H1, the recent date of the common ancestor of all living humans; the other theme (Goldman and Barton 1992; Hedges et al. 1992; Maddison et al. 1992; Templeton 1992, 1993) appears to accept H1 and questions the strength of support for H3, the African location. The lack of analysis of the logical structure of the overall hypothesis has made it difficult to know precisely which aspect is being questioned. This is compounded by concentrating on the mtDNA and ignoring the prior work that stimulated the hypotheses.

The first line of questioning is from those who prefer the "multiregional" model (Wolpoff et al. 1984) for the origin of *H. sapiens,* itself a bold hypothesis (Popper 1959). (Hereafter we restrict the name *Homo sapiens* to modern humans and do not consider directly relationships to other groups such as Neanderthals.) This multiregional model suggests that *H. sapiens* evolved during the last million years from *H. erectus* over the full range of this species (Africa, Europe, and Asia). Under this model, each of the many mutations crucial to the development of *H. sapiens* must have either occurred in parallel in different parts of the earth (by some unknown mechanism) or arisen just once and then spread over the three continents by virtue of a high rate of selection, migration, and interbreeding. This latter version we will call the "universal migration" or "panmixis" version of the multiregional theory, and it is a testable theory.

A problem with the multiregional model has been the difficulty in formulating a quantitative model of mutation and migration, though recently Thorne et al. (1993) proposed some qualitative models. Supporters of the multiregional model generally work with morphological data where it is difficult to get agreement on cladistic features of skulls and to define a genetic model of morphological change for quantitative testing. It is not yet possible to relate differences in skull morphology to specific genetic differences. Indeed, in some cases (Leamy 1993) a high morphological variability (between the left and right sides of the skull) may reflect a low genetic diversity from inbreeding rather than indicate high genetic diversity, as might be expected from "first principles." However, we show here that the universal migration or panmixis variant of the multiregional model is subject to tests by using DNA sequences. Thus it is a useful scientific model (sensu Popper 1959).

The second line of criticism (Goldman and Barton 1992; Hedges et al. 1992; Maddison et al. 1992; Templeton 1992, 1993) questions the analysis of the sequence data, specifically the statistical significance of the results. We agree with the criticisms to the extent that the original analysis was inadequate. The problem is not with the experimentalists who are both collecting the data and putting forward challenging hypotheses but rather with the theorists who have not developed and tested appropriate methods for a quantitative analysis with large number of sequences. As such, the criticisms are more a reflection of the analytical techniques available than of the hypotheses (both out-of-Africa and multiregional). The major challenge for theorists is to provide better methods applicable to large data sets. Our aim is to develop some new techniques for analyzing large numbers of sequences and then test these on the human mtDNA data set (Vigilant et al. 1991). In order to do this, we first describe the problems we see in analyzing large data sets and then propose a strategy to solve them. In developing this strategy, several new techniques are described.

Background: Problems and Strategies

Well-known problems for any quantitative analysis of a large number of sequences (especially of short sequences) include the following.

P1. *Large number of trees.* The number of potential phylogenetic trees grows exponentially with the number of taxa, t. For $t = 135$ taxa there are $\sim 10^{265}$ rooted binary trees (compared with about 10^{70} elementary particles in the entire universe). This limits the study to computationally efficient (Penny et al. 1992) methods for evaluating each tree. Only heuristic methods (in the operations research sense) can be used for searching the "tree space." There are $(t-3)$ internal edges (132 in this case) on any tree, and we consider these as separate parameters of the model to be estimated. (We use a mathematical terminology for trees consisting of nodes connected by internodes or edges; Penny et al. 1992.)

P2. *Limited range of tree values.* This problem is best illustrated with the parsimony (minimal-length) tree selection criterion where the lengths of trees for these data is of the order $\sim 10^3$. So, by an elementary counting argument, the number of trees of any length must average $\sim 10^{264}$. Although far fewer trees are expected near the global optimum, there

may still be billions of trees of minimal length, making "the best tree found" of unknown significance until more information is available.

P3. *Short sequences and equivalent trees.* Each binary tree has more internal edges ($t-3 = 132$) than parsimony sites for this data. Therefore, even the best trees are expected to have some internal edges of zero parsimony length, and these edges can be contracted to give a nonbinary tree. Consequently, large numbers of apparently different trees of the same length are simply rearrangements around internal edges of the tree that are not supported directly by any nucleotide substitutions.

P4. *Nonindependence of trees.* Heuristic search methods traverse the tree space by making slight modifications to existing trees; consequently, the trees found in a single search are *not* independent. Standard methods for representing common features in the trees (such as various forms of consensus) do *not,* therefore, represent the overall structure of the solution space.

P5. *Many local optima* (Hendy et al. 1988; Maddison 1991). Our preliminary work (unpublished data) using simplistic searches from random starting trees found an extraordinary number of local optima over 100 steps longer than the shortest trees, even for only half the sequences. Hence heuristic methods such as simple hill climbing (with or without steepest ascent) are not expected to perform well with random starting trees, though more advanced programs include mechanisms for escaping from local optima.

P6. *Locating the root.* The position of the root is vital for hypothesis H3. In order to locate the biological root of a tree, one requires additional information (Farris 1972; Penny 1976; Steel et al., 1993) such as an outgroup, an assumption of equal rates (the molecular clock), or an assumption of the nucleotide frequency at the root. If the outgroup is too distant, it can distort the underlying unrooted tree, even with five taxa and equal rates (Hendy and Penny 1989). Outgroups also have a biased tendency to join the unrooted tree at edges with two or more substitutions (Hendy et al. 1980).

P7. *Sites evolve at different rates.* Seldom do all positions evolve at the same rate; mtDNA include a small number of positions that are hypervariable in that they have changed more frequently than expected for all sites changing at equal rates (see, e.g., Wakeley 1993). Such sites tend to mask more informative sites that change more slowly (Hendy et al. 1980; Kuhner and Felsenstein 1994).

P8. *Rooting tree from a single genetic unit.* Because the mtDNA does not recombine, it is inherited as a single unit. However, under a multiregional model, the "last common ancestor" for each gene may have arisen in different parts of the earth. In this sense each gene could, in principle, indicate a different location for the root of the tree. Consequently, mtDNA, or any other gene, *cannot* be used to root the tree as a whole (i.e., the entire genome) unless the multiregional model can be excluded.

P9. *Unresolved tree.* Because the tree is not fully resolved (i.e., is nonbinary), there can be marked differences in length from including, or excluding, a hypervariable site. This makes it difficult to compare results from different programs which may or may not include some sites, for example, depending on how programs handle sites with deletions.

Given these problems, it is clear that newer methods of tree analysis would be necessary to make further progress in the study of human evolution (Wilson et al. 1991; Goldman and Barton 1992; Maddison et al. 1992; Ross 1992). There is now interest in studying the properties of the set of trees close to the overall optimum (the landscape) rather than focusing on a single "best" tree (Hendy et al. 1988; Maddison 1991; Page 1993). Our overall strategy, outlined below, included a mixture of using existing techniques and developing new techniques that we hope will be useful in other studies as well.

S1. *Analyze the landscape.* Our principal aim was to obtain information on the landscape of trees (Hendy et al. 1988; Maddison 1991; Kauffman 1993) that have been found independently and are close to the global optimum. This landscape consists of the trees, the distances between them (S5), a neighborhood for each tree, and a measure of how well each tree describes the data (S11).

S2. *Heuristic search procedure.* We used a new search procedure, the Great Deluge (also called threshold accepting) to locate optima in the landscape (Dueck 1990; Dueck and Scheuer 1990; D. Penny and M. Steel, unpublished data). It has proved useful for a variety of scientific applications as it is effective in escaping from local optima. As implemented here, the Great Deluge has both stochastic and deterministic phases. The stochastic phase is a random walk in the tree landscape (S1), at each stage moving to any neighboring tree whose fit to the data lies above a constantly rising "waterline" (hence the name Great Deluge). The main advantage of the program was our ability to vary parameters in order to measure properties of the landscape.

S3. *Restrictive definition of local optima.* A tree is accepted as a local optimum if the program has been

unable to find a tree *at least equally good* after 50 iterations of the stochastic phase and cannot find a *better* tree after trying all (neighboring) alternatives by three methods of changing the tree (crossovers [nearest neighbor interchange]), cut and paste (subtree pruning and re-grafting), and taxon swapping (interchanging pairs of taxa) (D. Penny and M. Steel, unpublished data). (An additional test is required to prevent oscillating within a small subset of trees.) Using three methods of traversing the tree space reduces the arbitrariness from selecting any one of them, since the search space is different for each procedure (D. Penny and M. Steel, unpublished data).

S4. *Independence of local optima.* Each run began from a random tree, and only one local optimum was retained. This allows us to consider each optimum as being found independently, and using a random starting tree avoids the criticism (Goldman and Barton 1992) that with a random addition tree the order of addition of taxa could, in some unknown way, bias the results.

S5. *A suitable tree comparison metric.* On this data many trees differ only on rearrangements of edges with no implied substitutions (zero-length edges) and are effectively the same tree for this data. However, existing tree metrics (including partitions, path lengths, and quartets; Steel and Penny 1993) give a range of values for these trees. We introduce a *sites metric* which compares the number of substitutions required (Fitch 1971) to fit each site (column) onto two trees T_1 and T_2. The distance is taken as either the sum over all sites of the absolute differences, $|l_1-l_2|$ (the *linear* form) or as the square root of the sum of squares of these differences (the *quadratic* form) where l_i is the parsimony length for a site on a specified tree, T. The linear form is related to techniques used for a different problem (Templeton 1983; Prager and Wilson 1988; Kishino and Hasegawa 1989), although a tree comparison metric comparing distance matrices (Lapointe and Legendre 1992) was the immediate stimulus for the sites metric. The sites metric is a true metric (rather than a pseudometric) only if two trees that have zero distance (the same parsimony lengths for every site) are regarded as equivalent trees; thus we focus initially on how the data fit the trees rather than on the trees themselves.

S6. *Improved predictability by using the median tree.* The median tree is a form of consensus tree (Penny et al. 1982; Barthélemy and McMorris 1986; Swofford and Olsen 1990). Although trees of low parsimony length are generally better predictors than longer trees (Penny and Hendy 1985), there are still a large number of trees of any parsimony length. The median tree is useful in identifying the best predictors among optima of a given length.

S7. *Convergence of optima.* Optima of a given parsimony length, L, were tested to determine whether the average intertree distance gets smaller as L decreases. If several unrelated major peaks of optimal trees existed, then this average distance should be large and stay positive; if there is only one major peak (albeit with many small side peaks), then the average distance between trees should converge toward zero, even if the DNA sequences are not long enough to resolve the global optimum. For this test we also examined the distribution of distances between trees to search for dispersed subsets of optima.

S8. *Locating the root by several methods.* The methods used have different rationales and require different conditions to be met in order to obtain the root. They are the standard outgroup method (Farris 1972) using a chimpanzee sequence, two forms of midpoint rooting (the edge of the tree with the longest average path length), splitting taxa into two disjoint subsets so as to minimize the largest within-group distances (Guénoche et al. 1991), and an ingroup method based on datings of colonization of different continents (Bowcock et al. 1991).

S9. *Downweighting hypervariable sites.* Hypervariable sites were detected in early runs when the distribution of rates at different sites did not fit the model of all sites evolving at the same rate. Downweighting these sites, and repeating the searches, led to a new set of local optima that are much more similar to each other (by the tree comparison metric). The lower weighting of hypervariable sites brings out more evolutionary information in the data while still allowing the faster-evolving sites to help resolve the fine structure of the tree.

S10. *Homogeneity (or dispersion) of groups throughout the optima.* Optima were analyzed for subtrees containing the largest number of people from any continent or ethnic group. If, for example, Europeans were scattered throughout the tree, then the largest subsets containing only Europeans would be quite small. The largest subset was first determined for the four geographic groups—Africans, Asians, Australopapuans, and Europeans—and then for several subgroups of Africans. This allows a direct test of the multiregional model in that there should be no large groups exclusively found in any region.

S11. *Parsimony was selected as the optimality criterion.* It is necessary to use a "global" optimization criterion that considers all possible trees (Penny et al. 1992) rather than a local optimization method (such as neighbor joining) that always selects the same tree(s) for a data set. The choices were maximum likelihood, parsimony, compatibility, closest tree, and deviations from additivity in observed distances. Parsimony was selected because its calculation, for a single tree with t taxa, requires order t time to calculate ($O[t]$); that is, it is linear with respect to the number of taxa. In addition, it now has some well-studied statistical distributions (Archie and Felsenstein 1993; Steel 1993; Steel and Charleston 1995). A new property is described later. The main disadvantage of parsimony is a potential lack of consistency on sequences not corrected for multiple changes (Felsenstein 1978; Hendy and Penny 1989). (Parsimony *is* consistent if appropriate adjustments are made for multiple substitutions; Steel et al. 1993.) A test was made to estimate if this data set was in a range where uncorrected parsimony was likely to be inconsistent. Parsimony minimizes the number of mutations required to fit the data onto a tree, and consequently shorter trees (with fewer mutations) are preferred over longer trees. Maximum likelihood was not used as it has some undesirable mathematical properties; as yet it is far worse than linear, ($O[t]$), to evaluate on a single tree, and there is no guarantee that current implementations will converge to the global optimum, even for a single tree of four taxa (Steel 1994).

S12. *A model consistent with many types of evidence was developed.* A useful hypothesis or model (Popper 1959) should be consistent with a range of types of evidence, a "consilience of induction," in the words of Whewell (1967, p. 469). A model should be consistent with data additional to that used to develop the model. In the present case, archeological and anthropological information is appropriate.

The above strategies are used to reconsider the criticisms of previous studies; other questions are covered by Stoneking (1994). Given the above strategy and techniques, the first step was to determine some properties of local optima. Our expectation, from an evolutionary model, is that optima will be much more similar to each other than to random trees and that better (shorter) optima will be more similar to each other than to worse (longer) optima. This is tested in the next section; if there is not considerable structure among the optima, then no evolutionary conclusions are possible (Penny et al. 1982).

Results

Local Optima

The first test involved 260 runs with all sites weighted equally (uniformly); the "equalW" or unweighted data set. The result is 260 different local optima (trees). The parsimony algorithm is easily adjusted to parallel computation (Penny and Penny 1990), but in this case only the simplest form of parallelism, running on different machines, was used. Each run consists of establishing an initial starting tree and then making a search through the space of all trees in an attempt to find a better tree. An arbitrary tree was selected at random (with all trees equally likely; Steel and Penny 1993) to start each run because although using "random addition trees" (taxa added in random order but to the best position available) gives much shorter starting trees, we could not be sure if any structure in the resulting local optima was an artifact introduced by the starting trees (Goldman and Barton 1992). In the present case, each of the 260 local optima (table 1*a*, equalW) were found independently.

These 260 optima are significantly shorter than randomly selected trees. The left-hand portion of table 1*b* shows the average length and standard deviation of the random trees. Clearly, each of the 260 local optima are significantly shorter than random trees and, importantly, are much more similar to each other than are random trees (right-hand side of table 1). However, the 260 optima still differ from each other, and a further test is whether any sites are genuinely hypervariable and masking some of the evolutionary information in the data. This can be tested by downweighting the hypervariable sites and finding new optima which should be more similar to each other (i.e., further from random trees than optima from the unweighted data set).

The shortest of the 260 trees was used to test for hypervariable sites, and, like previous authors (e.g., Wakeley 1993), we found sites required widely different numbers of mutations to fit onto the tree. It is first necessary to determine whether this distribution is expected, or consistent with, the model where all sites were equally free to change (i.i.d.). The "lost Bealey theorem" (Steel et al. 1995) provides for the shortest tree an upper bound on the distribution of parsimony lengths at different sites, assuming that all sites free to change do so identically and independently (i.i.d., the Cavender-Farris model; Penny et al. 1992). These upper bounds on the numbers of sites with 0, 1, 2, 3, . . . , and so forth, changes will exhibit a Poisson-style rate of decay as the numbers of changes per site increases. The distribution, together with some additional calculations, are shown in figure 1.

For this data set there were 596 noninformative sites; 421 had no substitutions on the optimal tree, and 75 sites required just a single substitution. The number of sites with two, three, or more changes (fig. 1) is much higher than predicted if all sites were free to vary; the results are thus inconsistent with the i.i.d. model since the upper bound provided by the theorem is strongly violated. We consider two modifications to the i.i.d. assumption. The first is that a proportion of sites are unable to change because of functional constraints but that other sites (the "effective sequence length") behave identically as expected under i.i.d. A simple calculation gives the corresponding upper bounds on the distribution of lengths across sites for 300 and 321 invariant sites. These fit the observed distribution better, but still not satisfactorily (fig. 1). For example, by assuming 300 invariant sites we can fit the expected distribution for one or two substitutions, but the observed tail of the distribution is still much higher than expected. Assuming 321 invariant sites is consistent for four substitutions. Varying the number of sites free to change does *not,* in itself, give a good fit between data and model. A second modification to the i.i.d. assumption is that a small number of "hypervariable" sites (perhaps a dozen or so) have evolved more rapidly than the others; they show more substitutions than expected if all sites changed at the same average rate. These combined features of functionally constant sites, plus some hypervariable sites, are a much better fit to the data, and we explore their effect.

The problem that hypervariable sites cause when inferring trees is that they can overwhelm the slower-evolving sites (Hendy et al. 1980). Hendy et al.'s (1980) paper summarizes a large body of work on establishing minimal-length trees by resolving incompatibilities between sites. If hypervariable sites are downweighted, they can still help resolve the fine structure of the tree without overwhelming the more slowly evolving sites. The validity of the new model can be tested by comparing the new optima found after giving these hypervariable sites a lower weighting. If a particular weighting scheme is valid (or at least closer to the real biological situation), then the best optima should be more similar to each other, and more dissimilar to random trees, than the 260 optima in the unweighted data set (equalW).

Reducing the Effect of Hypervariable Sites

The sites that required the most substitutions for the unweighted data set were downweighted as follows. The 5% of sites requiring most substitutions were weighted as 0.4 (low weighting), the next 20% of sites weighted as 0.6 (medium weighting), and the remainder left unchanged at a weight of 1.0 (unweighted). These values allow two medium sites in agreement to count more than a single unweighted site and a medium and

Table 1
Parsimony Lengths and Distances between Trees for (a) Local Optima and (b) Random Trees

a.	Lengths of Local Optima				Intertree Distances		
Data Set	Average	SD	Shortest	Number	Linear ± SD	Quadratic ± SD	Number
EqualW	369.33	4.14	364.0	260	32.40 ± 6.64	7.21 ± 1.21	250
SmallW	266.78	1.96	264.2	160	18.58 ± 5.98	5.04 ± 1.10	160
MediumW	253.64	1.58	251.6	580	18.93 ± 5.54	5.08 ± 1.04	580
RandomW	308.87	7.10	294.8	40	42.81 ± 9.40	9.69 ± 2.04	40
b.	Lengths of Random Trees				Random Tree Distances		
Data Set	Average	SD	Shortest	Number	Linear ± SD	Quadratic ± SD	Number
EqualW	1,055.57	15.68	974	200,000	49.94 ± 10.7	12.99 ± 2.69	500
SmallW	767.29	10.59	708.4	200,000	49.80 ± 10.5	12.94 ± 2.75	500
MediumW	744.03	10.33	684.2	200,000	50.32 ± 10.5	12.99 ± 2.72	500
RandomW	900.03	29.34	778.8	200,000	50.42 ± 10.5	13.17 ± 2.75	500

Note.—Table 1*a* presents a summary of the lengths of local optima and the distances between them under the two versions of the sites metric ("linear" and "quadratic"). The data sets are unweighted (equalW), and those with low weighting (smallW), medium weighting (mediumW), and random weighting (randomW, which has the same weights as mediumW but randomly reassigned among sites). The number of mutations required to fit each tree was determined using parsimony (Fitch 1971). Local optima are for independent runs of the Great Deluge; each run searched at least 10^5 trees. For comparisons, 200,000 random trees were generated with all binary trees equally likely (Steel and Penny 1993) for each of the four data sets, and a summary of the corresponding tree lengths and intertree distances is shown in the lower half of the table (*b*). The results show that for the four data sets, the local optima are significantly shorter than random trees; standard deviations for local optima decrease as weighting increases; lengths and variability of local optima for randomW are higher than those of mediumW; local optima are much more similar to each other than random trees; downweighting the hypervariable sites (smallW and mediumW) reduces the differences and the variability of local optima but not of random trees; and randomly reassigning weights (randomW) reverses both these last effects, but weighting does not affect the difference between random trees (a control).

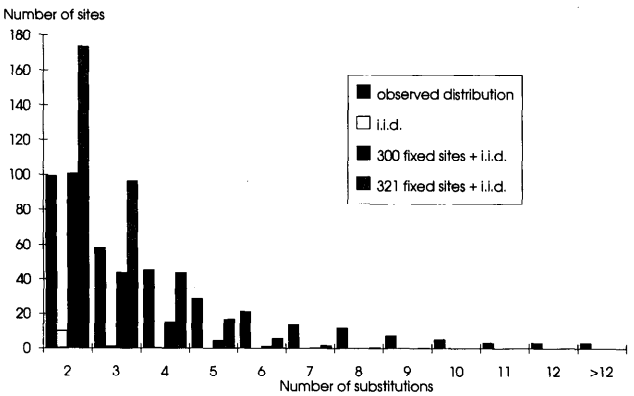


FIG. 1.—Numbers of substitutions per site on the optimal tree. The histogram shows first the observed numbers of sites requiring two or more changes on the best tree for the unweighted data set (equalW). Then upper bounds are predicted for how many sites will have a given number of substitutions on the optimal tree on the basis that 421 sites were constant, and 75 sites required just one substitution. The comparison is done taking three values for the number of 421 constant sites that were invariant (unable to change state). The three cases considered were no constrained sites (all sites free to vary), 300, then 321 sites constrained not to vary. The first of these violates the upper bounds quite strongly; 300 sites constrained fits the observed frequency of sites up to two substitutions; 321 sites fits the observed number for four substitutions but not the number of sites with more substitutions. Varying only the number of sites free to vary cannot fit the observed distribution. These results demonstrate that not all sites are equally free to vary; consequently some hypervariable sites can mislead slower-evolving sites.

low site in agreement to equal one unweighted site. This modified data set with the small amount of downweighting is referred to as the "smallW" data set. This weighting is different from "weighted parsimony" (Hillis et al. 1994), where the weighting is on character state changes, not on nucleotide sites. With this differential weighting, we tested the prediction that downweighting would result in local optima becoming more similar to each other than with equal weighting.

To test this prediction, we then ran the data set "smallW" 160 times through the Great Deluge search; results are shown in the second row of table 1 *a* and *b*. These new trees are again shorter than random trees for the same weighting. Because of the different weights, the lengths of the optima are not comparable with the unweighted optima, but the distances between optima are. The important result is that optima found with this data set are indeed much more similar to each other than without weighting, approximately 19 as compared to 32 when measured on the linear form of the sites metric (right-hand side of table 1 *a*). The conclusion is that the weighting is useful, and the optima can only be more similar if the weighting detects more structure (nonrandomness) in the data. As a control, the small amount of weighting does not change the similarity between randomly selected trees (right-hand side of table 1 *b*).

A second step was then to extend the weightings to give sites with a medium number of substitutions a weighting of 0.8 and the two sites with only one purine and one pyrimidine (transversion substitutions) a weighting of 1.6. This is conservative compared with studies in which transversions are often weighted by up to 4–10 times more than transitions. After 580 runs of the Great Deluge, this data set ("mediumW") gave results similar to smallW in that the optima are a similar distance apart and again much more similar to each other than for the unweighted data set (table 1 *a*). The distribution of lengths of the 580 trees from the mediumW data is shown in figure 2. Altogether there were 1,000 separate runs for the three data sets. The lengths of the optima are slightly less dispersed than for smallW (SD in table 1 *a*). Again the optima are much more similar to each other than for the unweighted data and little different than for smallW (table 1 *a*). Downweighting the hypervariable sites is thus detecting more structure (nonrandomness) in the data.

It was considered possible that the increased similarity between optima from smallW and mediumW (compared to equalW) was in some unknown way an artifact of the lower weights. This seems unlikely as the random trees are still the same distance apart (table 1 *b*), but to check this possibility we made 40 Deluge runs with exactly the same weights as mediumW but with the weights randomly reassigned between sites for each run. This is data set "randomW." The results show that randomly reassigning the weights leads to much longer trees, their lengths being many standard deviations longer than the original mediumW (table 1 *a*). In addition, the local optima from randomW are not nearly as similar as either the weighted or unweighted data sets (table 1). Thus the increased similarity of trees from the small amount of weighting does not appear to be an artifact due to weighting; the downweighting is revealing

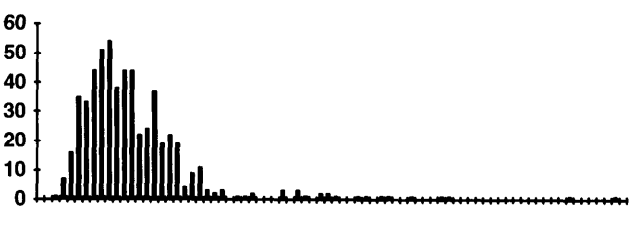


FIG. 2.—Number of local optima versus lengths for 580 runs with the mediumW data set. The optimum with the shortest length (251.6) was found only once; seven optima one step longer (251.8) were found, and so forth. The most frequent class was trees of length 253.0 with 54 optima. Although shorter optima are generally better predictors of behavior with other data (Penny and Hendy 1985), more consistent results are obtained by also considering the average consensus values for the optima.

more evolutionary information. The original prediction—that the weighting does give much more similar optima than the unweighted data—is supported and consistent with earlier findings (Hendy et al. 1980).

The conclusion of this section is that there is considerable structure among the optima, even though we have no guarantee of finding trees that are the global optima. The local optima are much shorter than random trees and, more importantly, much more similar to each other. Downweighting the hypervariable sites increases these effects. The next step is to compare the local optima themselves rather than comparing them to random trees. This will test the prediction that shorter optima should be more similar to each other than to longer optima.

Landscape

We require information on the distribution of these optima in the multidimensional space of trees. If the optima were scattered more or less randomly throughout the entire space of trees, then there would be no justification for any evolutionary interpretation. An alternative, expected under an evolutionary mechanism (stochastic change with divergence), is that the better optima are converging toward a single global optimum, even if the sequences were too short to allow the global optimum to be determined uniquely. With such short sequences we expect many internal edges of the tree to be unresolved, that is, to be of zero length (no mutations). Because of these zero-length edges, many rearrangements of the tree will be possible. If the general conclusion about converging toward the global optimum were correct, then better optima should be more similar to each other than poorer optima (which require more mutations); shorter trees are expected to be more similar than longer trees. The distances between optima were measured using both forms of the sites metric and the results classified according to the lengths of the optima.

For trees of a given length, distance values were compared with other optima of the same length, with better optima (requiring fewer mutations), and with poorer optima. The expectation was that if the better optima are converging toward a single solution, then the average distance between optima of a given length L will decrease as L decreases; optima of any given length L would be more similar to each other than to poorer optima (optima with a longer length, L); consequently, the distance from optima of a given length L to the average of all poorer optima will increase as L increases; and, most strikingly, optima of a given length L will, on the average, be more similar to better optima than they are to themselves. The first prediction is perhaps the most important in that, if correct, the others are expected to follow. However, the final prediction—that trees at any optimality level are more similar to better optima than to themselves—is more striking and helps us understand the overall landscape.

All four predictions are supported (fig. 3). Results for the 580 optima from "mediumW" (table 1) on both the linear and quadratic forms of the sites metric reveal that better optima (shorter lengths) are more similar to each other than are optima with larger lengths. Because the values of comparisons of optima of equal lengths are independent, a rank correlation test can be made, and the increase in similarity of shorter trees is highly significant. Most strikingly, optima of a given length are *more similar* to optima above them than they are to themselves. In addition, a weighted regression of tree length against independent distances between trees of a given length suggests that the points are curving downward as the tree length becomes shorter. For example, the probability of the coefficient of a square-root component in the regression was <0.005, with or without exclusion of the data points for the two shortest sets of trees, lengths 251.8 and 252.0, respectively. This suggests that the shortest possible tree is close to the length of the best trees found in this study. However, we do not attempt to predict its length from the x intercept of a regression line because the residuals of the two shortest sets of trees are the largest and in opposite directions. We plotted the distribution of distances between trees as an additional test, and the resulting histogram showed no evidence for a bimodal distribution.

Our simplified interpretation of the landscape from these results is shown in figure 4*a* and *b*. We imagine the landscape in three dimensions as a large volcano, with better optima near the top converging toward the global optimum. For reasons outlined later, we call this the Kilimanjaro landscape.

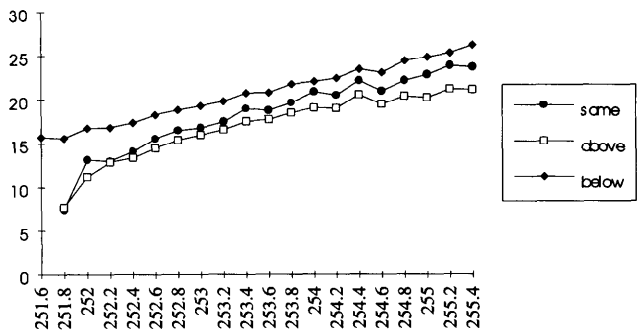


FIG. 3.—Comparison of distance values between trees (sites metric) versus length of local optima. The X-axis is the parsimony length of the optima (see fig. 2). Values on the Y-axis are average distances under the sites metric (linear form) for optima of a given length, to optima of the same length (*circles*), to better optima (*squares*), to worse optima (*diamonds*). The most important points are that longer (worse) optima are more dissimilar than shorter optima and that optima are, on the average, more similar to better optima than to themselves. An interpretation of these results is shown in fig. 4.

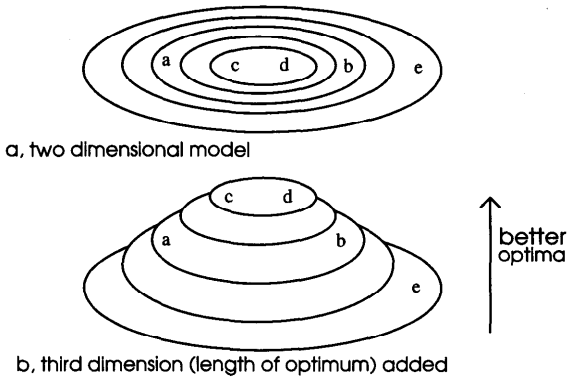


FIG. 4.—Optima of different lengths converging. *a*, Model representing a "landscape" in two dimensions with the optima of shorter lengths being included within the range of those with longer lengths. The distance between two optima *a* and *b* of the same length can be greater than the maximum between *a* and any better optimum, for example, *a* to either *c* or *d*. Conversely, the maximum and average distance between *a* and optima at longer lengths, such as *e*, can be greater than the distances within the same level as *a*. In practice there are 120 sites (dimensions) contributing to the value for the sites metric, not just the two dimensions represented here. *b*, Representation in three dimensions, with the third dimension being the "height" of the optimum (which is not included in the value for comparisons under the sites metric). This represents the results of shorter optima "converging" to a single global optima, even though we may not have reached a global optimum.

Thus there is considerable information in the optimal trees, despite some earlier doubts (Goldman and Barton 1992; Templeton 1992, 1993) about the "significance" of the shorter trees. This structure allows evolutionary hypotheses to be tested. We need to determine whether the best optima distinguish between the out-of-Africa and the multiregional models. The prediction of the multiregional model we are testing here is the panmixis model; there was continued genetic intermixing of populations over the whole of Africa, Europe, and Asia in the early stages of evolution of *Homo sapiens*.

Properties of the Best Trees

We initially identified the 10 "best" optima: the shortest 8 plus 2 additional trees that were longer but excellent consensus (median) trees. In general, shorter trees are better predictors than longer trees (Penny and Hendy 1985), and so the shortest eight were included. Using tree comparison metrics as an additional criterion allows more information to be included. In this study there was only a single shortest tree (length 251.6), but it was not the best median tree; five of the seven trees one step longer were better median trees. Indeed, 1 of the 54 trees of length 253.0 was a very good median tree even though, on the average, trees of this length were considerably worse than the shorter trees. Figure 4 is consistent with an occasional slightly longer tree still being central among the population of all local optima.

Because of both the variation among trees reported previously (Vigilant et al. 1991; Hedges et al. 1992; Templeton 1992, 1993) and the variation between trees at any one optimality level, we assumed initially there may not be much agreement on details of these 580 local optima trees. For example, the positions for the !Kung and the single Naron sequence differed considerably in trees reported previously. Despite these reservations, we determined the largest cluster for each of the 10 trees that was made up exclusively of African, Asian, European, or Australopapuan (New Guinean plus one Australian sequence). We then extended this by finding the largest cluster allowing one, then two, and then three sequences from any other group. For example, we found the largest group of Asian sequences that had three non-Asian sequences.

The most striking feature of this analysis (table 2) is the large size of an exclusively African group, particularly when up to two additional sequences were allowed. All 10 best trees had the same 23 African sequences as

Table 2
Largest Numbers of Sequences in Continental and/or Ethnic Groups

Group	No. in Group	No. from Other Groups			
		0	1	2	3
African	78	23.0	5.0	47.0	0.6
European	15	3.7	5.0	2.2	3.6
Asian	17	4.3	5.4	7.5	8.0
Australopapuan	25	2.9	3.8	2.5	1.9
!Kung	16	10.0	16.0	0.0	0.0
Afroamerican	8	2.0	1.0	1.0	1.0
East Pygmy	15	9.0	3.0	0.0	3.6
Herero	8	4.0	4.7	0.9	1.4
West Pygmy	14	8.8	0.0	11.6	7.2
Yoruban	12	2.1	2.1	4.0	4.0
Hadza	4	0.0	2.9	0.0	0.1

NOTE.—Sequences were first grouped into four geographic classes (the first four groups listed) and their largest clusters found in each of the 10 best trees; results are averages for these trees. Clusters could consistent exclusively of members from a single group or have one to three sequences from any other group(s). All 10 trees had the same cluster of 23 African sequences; the largest African cluster with one non-African sequence was five sequences (which implies the cluster of 23 must join with a subtree containing two or more non-African sequences). The main features are the large African clusters is that groups outside this cluster of 49 sequences are heterogeneous (more dispersed throughout the trees). The study was repeated by subdividing African sequences further into the seven groups listed next. The most striking feature is the large size of the !Kung group, in which all 16 sequences join together with the other San group, a single Naron sequence. The eastern and western Pygmies each form relatively large clusters on the trees but are more diverse overall. An important control is that the Afroamericans do not form a single large cluster, reflecting the varied sources of their ancestors within Africa. The overall conclusion is that there *cannot* have been continued genetic interchange over the whole range of early *Homo erectus/Homo sapiens* populations, thus contradicting multiregional models requiring high migration between continents.

the largest group and similarly had the same 47 African sequences plus two "Asian" sequences (numbers 23 and 26; all numbers are those in the Vigilant data set). We refer to this as the "49 group." The composition of this cluster of 49 was identical in every case (sequences 1–48 plus 76, the single Naron sequence—a San group from southwest Africa [Schapera 1960] similar to the !Kung). Two other large groups were eastern Pygmies and another of western Pygmies. An important internal control is that the sequences of Afroamericans are scattered among the Africa groups; they do not form any large groups. In contrast, there were no large clusters of other continental groups, and, more strikingly perhaps, allowing one to three members from outside the group did not lead to a large increase in the size of such clusters. In other words, allowing one to three immigrants scarcely increased the maximum group size. The study was repeated for the next 15 trees of length 252.0 with virtually identical results.

The large African grouping can be investigated further by considering smaller populations within it, and so the above analysis was repeated with the African sequences subdivided into seven subpopulations (excluding two groups of a single sequence each). The most striking feature of the results (table 2) is that the !Kung, together with the single Naron sequence (number 76), are united in all 10 trees (the Naron are a related group from southwest Africa; Schapera 1960). The results are highly significant. The probability that a randomly chosen binary tree (under all trees equiprobable; Steel and Penny 1993) has a *given* split of size k, $n-k$ is

$$\frac{b(k+1)\cdot b(n-k+1)}{b(n)}, \quad \text{where } b(s) = (2s-5)!!$$

$$(2s-5)!! = 1\cdot 3\cdot 5\ldots(2s-5).$$

The probability that a *pair* of randomly chosen binary trees share *at least one* split of size k, $n-k$ is bounded above by

$$\binom{n}{k}\left[\frac{b(k+1)b(n-k+1)}{b(n)}\right]^2.$$

Thus, given that these groups (the 17 !Kung + Naron sequences and the 49 group) were found in one tree, the probability of finding the same group in a second tree is $<10^{-20}$ for 17 sequences and $<10^{-37}$ for a group of 49. There was certainly more common structure in the optima than we expected, and, even if have not reached the global optimum, we can predict many of its properties.

We consider that these results eliminate any "universal migration" form of the multiregional model, a model that requires alleles to spread throughout the range of *Homo erectus* (Africa, Europe, and Asia). Note in particular that this conclusion is independent of which region of the world had the largest group; a very large Asian group of sequences that was not dispersed would lead to the same rejection of the multiregion model. A possible ad hoc modification would be to argue that Africa was different in that only males, not females, dispersed. This would mean mitochondrial genes (maternally inherited) would disperse little, but nuclear alleles could spread more widely, albeit more slowly, than if both sexes dispersed. However, there is a large body of anthropological studies of social organization against such an ad hoc modification (Murdock 1960). Female dispersal between bands/villages is, as in our nearest relatives, almost universal in humans. There are exceptions in some agricultural groups with a strong matrilineal emphasis (Murdock 1960). Male dispersal does occur occasionally, but it is even less common than neither male nor female marrying outside the village (Murdock 1960), which would not aid genetic dispersal.

The results from the best 25 trees were more striking than expected, but we still anticipated these groups would be quickly broken up as we searched longer optima. Nevertheless, we continued examining longer optima to test when either the 49 group, or the !Kung group, was broken up (either by including other sequences or losing some of the 49). The 35 trees of length 252.2 had the same patterns, though one of the trees included a single additional African sequence in the 49 group (number 56, a Herero sequence). The study was then extended to the best 40% of optima (length 253.0). Indeed *all* the best 40% of trees have the same major groupings with occasionally (<2%) sequence 56 being included in the 49 group. It appeared to be a lower cost to include some additional African sequences in the 49 group than to exclude any of them. Just one of the 47 trees of length 253.4 had no partition with the full 49 sequences; it omitted sequences 24 and 25 (both Yoruban sequences). Essentially all the best 300 trees had the 49 grouping. The results of the !Kung + Naron are even more striking. In order to find a tree that did not have the full !Kung group, we had to go right down the tail of the distribution and a tree of length 258.0 which omitted two of the !Kung plus the Naron sequence from this group of 49.

Identical results were found from the equivalent analyses on the 160 optima from the smallW data set, which is expected given that these optima are similar distances apart to the mediumW optima. With the unweighted data set (equalW) the results are more diverse as would be expected from the optima being more scat-

tered. Nevertheless, about 40% of the best 62 optima have the group of 49, and another 20% either omit, or include, one additional sequence. Again, about two-thirds have the 16 !Kung sequences either grouped together, or with one additional sequence, usually the Naron. The results emphasize that the downweighting is enhancing an evolutionary signal that is to some extent being masked by the rapidly evolving sites. Giving equal weighting to rapidly evolving sites results an occasional tree that would support almost any result. The high consistency of the results with smallW and mediumW is striking. Although we cannot guarantee how close we are to the global optima, this high consistency between local optima, particularly compared with earlier studies (Hedges et al. 1992; Templeton 1992, 1993), supports the robustness of the conclusions. The !Kung grouping may also be of anthropological interest in that it supports a distinct group of southern African alleles consistent with a small group colonizing southern Africa in the last 50,000–100,000 yr (Deacon 1992).

A comment is necessary on the two "Asian" sequences in the Africa 49 group. The sequences are similar to other African sequences in the 49 group so they are not ancient lineages. In this case there is insufficient information to distinguish between several explanations. We can expect (Waddell and Penny 1995) to find some modern African sequences within Asia as a result of an early slave trade into Asia. However, the two sequences are from people from Chinatown in San Francisco who describe themselves as Chinese-Americans. It is unclear whether these particular individuals have an Afro-american maternal lineage, given that there were Afro-americans living in Chinatown in the last century (M. Stoneking, personal communication). Given their close similarity to other sequences in the group (they are not ancient lineages), we will, for the purposes of rooting the tree and until further information is available, refer to the whole 49 group as African.

It is now appropriate to reexamine the data in table 2, particularly the African group as a whole. The largest exclusively Africa group has either 23 sequences, or 49 when the other two sequences are included. This is in marked contrast to the size of the largest group, five, with one non-African sequence. (Clearly the group of 23 must join with others to give the group of 49; otherwise a larger group of "African plus one other" would occur.) The number in the African group with one non-African sequence is very similar to the results for Europeans, Asians, and Australopapuans. The division is thus *not* African versus non-African sequences; rather, it is the group of 49 African sequences versus all others—the others including some African sequences. Such a distinction cannot be made when putting all Africans into a single group as is done with gene frequency data. This conclusion of an African group versus all other sequences is probably the central result of the whole reanalysis of the human mtDNA data. There is one large group of African sequences, and all other sequences (African, European, Asian, and Australopapuan) form a second major group. Recent results comparing European and African populations are in agreement (Pult et al. 1994).

The results of the size of African groups in the optima were perhaps the most unexpected part of the study and the most important in distinguishing between a multiregion/single-region origin of modern humans (hypothesis H2). This analysis shows the importance of keeping just one tree from each run. Each tree has been found independently, and so expected properties of the global optimum can be determined. However, from the viewpoint of rejecting the multiregional model, the geographic location of such a large group is immaterial. Even if the grouping had occurred elsewhere in the world, it would still be evidence against any model requiring panmixis.

Root of the Trees

As discussed earlier (P8), a single "gene" cannot be used to locate the root of the tree for the whole genome until the multiregional model is excluded. But because the only testable version of that model appears excluded, we will estimate the position of the root and take its geographic position as our best estimate of the ancestral population. Several strategies are possible for rooting trees (Smith 1994), and four have been considered for this data. The two main methods are outgroup and midpoint rooting. The use of a chimpanzee sequence as an outgroup has been used earlier, though not on these optima. The second approach, midpoint rooting, does not require an outgroup but does depend on approximately equal rates of evolution to estimate the position of the root.

Two variants of midpoint rooting have been used. The usual one is to find the midpoint of the longest path between two taxa in the tree, the *longest path* version. This variation uses only a single path on the tree and as such is expected to be sensitive to both random sampling effects (resulting from short sequences) as well as to sequencing errors. A more robust approach, also used here, is finding the edge of the tree with the longest *average* path length. Each edge of a tree partitions t taxa into subsets with x and $t-x$ taxa. There are $x(t-x)$ paths through each edge of the tree, and so the longest average path is expected to be a better estimate of the root than using only a single path. This is the *all paths* version of midpoint rooting.

Our reasoning is as follows. Let T be a binary phylogenetic tree with weights on the edges representing the expected number of substitutions along it. The tree is rooted on one of the edges to give a root vertex ρ. The weight for any edge e is $w(e)$ and for any pair of taxa i and j; the distance between them, $d(i, j)$, is the sum of the weights of edges on the path between them. Under the molecular clock hypothesis, the distance $d(\rho, l)$ from the root vertex ρ to any leaf l (taxon, or in this case a sequence) is expected to be constant for all leaves l, and we call this value the height h_0 of T. For an edge e of T, let $\delta(e)$ be the average (over $l \in L$, $l' \in L'$) of $d(l, l')$, where $\{L, L'\}$ is the bipartition of the set of leaves of T formed by deleting edge e.

We arrive thus at proposition 1: Under the molecular clock hypothesis, $\delta(e)$ is uniquely maximized by the edge e of T, whose midpoint corresponds to the root ρ. A proof is given in the Appendix.

A third method (Guénoche et al. 1991) does not require a tree but partitions the taxa into two subsets in order to optimize a function of the pairwise distances between all pairs of taxa. An efficient algorithm is available for this problem (Guénoche et al. 1991). Again, this partitioning into two subsets is a good estimate of the root as long as the lineages are evolving at similar rates. Finally, we examine results from a form of "in-group" or "subgroup" rooting by using estimates (in this case) from archeological evidence for the time of colonization of regions by modern humans (Bowcock et al. 1991). Although the method is not "algorithmic" in the same way as the other approaches, it does allow useful checks. By estimating several divergence times within a tree and comparing these, we can check that the molecular clock is a good approximation and narrows down the position of the root so that a better estimate is possible.

The methods have different requirements for their validity. Outgroup rooting requires a good tree and sequences from an outgroup but does not assume a molecular clock. Midpoint rooting requires a good tree and depends on the molecular clock. The third method does not depend on a tree but partitions the taxa based solely on a matrix of genetic distances—it does assume a molecular clock. If in-group rooting has several points in the tree where direction of change is established, there may be less sensitivity to small deviations from the assumptions.

For the following reasons, a molecular clock is a reasonable assumption in studies such as this. The number of neutral substitutions is proportional to both the proportion of mutations that are neutral and the overall mutation rate (which in turn depends on properties of enzymes). Within a short time scale we expect the proportion of neutral mutations to be the same and the properties of the enzymes involved in DNA synthesis and repair also to be the same. Both properties could change over long-term studies of different major lineages. Thus, although it is reasonable to assume equal rates of change for this data set, the conclusions do not depend on it. Overall, if methods give a similar answer, then we are more confident of the result—the position of the root has survived different tests.

The results for the first method for estimating the root, outgroup rooting with the chimpanzee sequence, are quite striking; all 245 best optima have an African root. In this case an "African root" means both that on one side of the root all sequences are African and that at least one of the first branching points on the other side of the root is also exclusively African. Even more striking was that the second and third best positions for rooting the tree were also exclusively African.

These results should be treated with care. Others (Hedges et al. 1992; Maddison et al. 1992) have already pointed out that using such a distant outgroup is a difficulty. We agree with this in that a distant outgroup can lead to problems, even when the molecular clock is valid (Hendy and Penny 1989). This effect was minimized by determining the trees with the outgroup omitted, then adding the chimpanzee to all possible positions, and also by selecting the three best places for rooting the tree. In spite of these qualifications, the results are impressive, particularly as the African root was suggested for earlier data sets (Cann et al. 1987).

Similar results were found with both forms of midpoint rooting, the midpoint of the longest path in the tree (longest path), and the edge with the largest average path length (all paths). Again, an African root was found in the 245 trees. The all-paths version gave an African root within the group 49 sequence; the single-path version gave an African root, but sometimes it was just outside this group, especially with the longer trees.

The third approach used did not resolve the taxa clearly into two subsets because, although the first few steps were clear, there quickly became a large number of ties. From that point, decisions were arbitrary. The large number of ties is not unexpected with a small number of sites in relation to the number of entries in the pairwise distance matrix, a feature commented on earlier when selecting a tree comparison metric. The fourth approach, in group rooting, is discussed again later, and the tree shown later in figure 6 is consistent with the same root. The support for an African root is thus strong, given this data.

Human Population Expansions

The time of the last common ancestor of the human mitochondrial genome cannot also be the beginning of

the major expansion in numbers of early humans. Most calculations of the rate of the loss of alleles from a population assume a constant average population size and are not applicable for a continuously expanding population. For reasons outlined below, a model of continued expansion is unlikely to lead to just one DNA sequence from the initial population remaining in the modern population. We consider a possible alternative with an early phase of constant average population size, followed later by an expansionary phase.

In considering how many lineages of the mtDNA survive from an earlier population, we use the usual independence assumptions common in branching theory (Arthreya and Ney 1972), and so the number of surviving lineages has a Poisson distribution. We show that the determining parameter (namely, the mean) depends critically on the dynamics of human expansion—from an initial, localized, founding group—to a much larger, geographically dispersed population. In particular, a uniform exponential rate of population growth (*lower half* of fig. 5) leads to the survival of a larger number of maternal lineages than the two-phase model (fig. 5), even given the same initial and final populations and over the same time scale.

To apply classical results from branching theory (Arthreya and Ney 1972), we must assume that in any generation mothers independently leave a random number of female offspring, though the expected number offspring is not necessarily constant with time. We further assume that the spread of the mtDNA through the population was by standard stochastic processes; that is, the process was neutral (Kimura 1983; Nei 1987). This is reasonable given that mtDNA is only one part in 10^5 of total DNA, it is not linked to nuclear genes, and there is no reason to suspect that improved mitochondrial function was responsible for the expansion of *Homo sapiens* numbers.

A simple model—scenario 1—takes a constant expected rate of population increase per generation (the expected population thus grows exponentially by a constant percentage each generation). Let k denote the number of ancestral females in the founding population; N, the size of the final population; and T, the number of generations over which this expansion occurred. A second model—scenario 2—starts from the same time and population but allows a two-phase behavior. For T_1 generations the population fluctuates about a mean (it has zero expected growth), and then for T_2 generations the expected population grows, on the average, each generation by a constant percentage, which must be higher than the growth in scenario 1 because it gives rise to the same population value (N) but in half the time ($T/2$). It is this second model (fig. 5) that we can fit to the data.

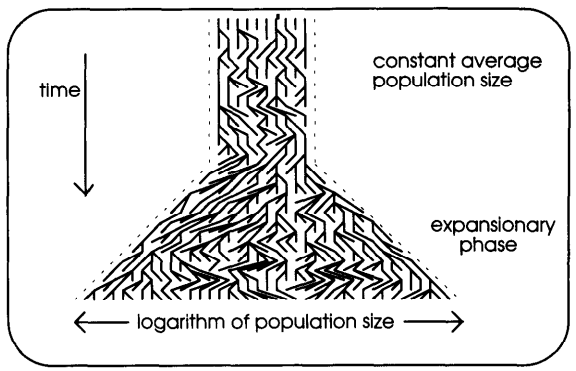


FIG. 5.—Two-stage model. The first stage is a constant average population size; then a second phase has an exponential expansion in population numbers. During the first phase of constant average population size, there is a random loss of most of the neutral alleles present in the initial population. Once the population starts a continued expansion, there is a much lower chance of random loss of alleles. The example shown here initially has 10 alleles, 7 of which are lost in the stationary phase and 2 more during the expansionary phase. However, for the calculations the human population started with a population that included 5,000 females.

We assume that k, T_1, and T_2 are large (>50), $k < N$, and $\ln(N/k)/T \ll \min\{1, \sigma^2\}$ where σ^2 denotes the variance of the offspring distribution. All of these assumptions are reasonable in the present setting and lead to proposition 2 (the proof is given in the Appendix): Under the above assumptions, the number of founding females whose maternal lineages survive T generations is given by a Poisson distribution of mean λ, where

(1) $\lambda > 2k \ln(N/k)/\sigma^2 T$, under scenario 1 (constant exponential growth) and

(2) $\lambda < 2k/\sigma^2 T_1$, under scenario 2 (two-phase model).

We use some estimated values based on earlier work; k = 5,000 (Hudson 1990; Takahata 1993), T = 200,000 yr/(20 yr/generation) = 10,000 generations (Waddell and Penny 1995), $T_1 = T_2$, $N = 5 \times 10^6$, and take $\sigma^2 = 1$. The value k (number of females) is toward the lower estimate of Takahata (1993) and fits the model better. With these parameters, we obtain that $\lambda > 3 \ln(10) = 7$, under scenario 1, while $\lambda < 2$ under scenario 2. Thus, with a population expanding continually, we do not expect, under the parameters estimated earlier (Hudson 1990; Takahata 1993), for there to be just a single variant remaining in the population. Because the distribution has a Poisson distribution, the probability of observing just a single variant under scenario 1 is approximately 10^{-3}. Thus, it is very unlikely that the last ancestor of human mtDNA occurred at the same time as the major expansion in human numbers.

Although these calculations are based on idealized assumptions, they suggest that a neutral allele fixed in the present human population must have arisen well before the major expansion of human numbers. Perhaps the main problem with the assumptions strengthens this conclusion—once the population spreads from a localized region, it is even less likely that all variants in an expanding population will be lost.

This two-phase model (fig. 5) appears consistent with the data—a stationary phase when the average population size is constant and the number of mitochondrial variants present in the original population was reduced, followed by a second expansionary phase when population numbers increase and a few final variants present from the beginning are lost. In addition, as discussed earlier and later under in-group rooting, the times it gives are in reasonable agreement for the first anatomically modern humans outside Africa: approximately 90,000 yr in the Near East (Aitken and Valladas 1992; Schwarcz and Grün 1992) and approaching 60,000 yr for Australia (Roberts et al. 1990, 1994). It is possible to alter the estimates of early population size, and a smaller initial population size would support a smaller time period before expansion. However, these estimates (Hudson 1990; Takahata 1993) are based on genetic diversity in populations (particularly HLA) and so are constrained and not arbitrary. This emphasizes that, overall, models have to be consistent with a variety of types of data; a criticism of just one type of data lacks force if the original conclusion was based on several lines of evidence.

Our calculations would not apply to a locus, such as HLA, where positive selection is expected (Takahata 1993). In such cases the models predict, as is observed, that HLA diversity would predate by a long time the last common mtDNA ancestor. It is a major, but common, misunderstanding to expect that all genes will converge to the same time and/or individual. Conversely, results from the HLA locus (Klein et al. 1993), given frequency-dependent selection, do not contradict the present model.

Some possible events could alter this conclusion. A hitchhiker event (a neutral allele closely linked with a gene which is increasing in frequency through positive selection) could occur. However, mitochondrial DNA is not linked to nuclear markers; although some early advantage through a sex-linked allele may be possible, there is no evidence for this more complex model. It is also desirable to know the equilibrium frequency of alleles for a given mutation rate and population size and use this value rather than assuming all female members of the population have different mtDNA sequences. Another alternative is a small effective bottleneck in population size resulting from the chromosome fusion that formed human chromosome 2 (Seuánez 1979; Burrows 1994). Because this is a unique event and the new, fused chromosome would initially have had no genetic variation, it is desirable to measure the level of genetic variation at the point of fusion for this chromosome. This chromosome fusion may have happened any time after the separation of *Pan* and *Homo* lineages and the level of diversity in modern populations should help identify the time the chromosome fusion occurred. Again, there is no evidence for such an event affecting frequencies of alleles of nuclear genes. The main conclusion we draw from this section is that the rapid expansion of human numbers must be more recent than the last common ancestor of any particular gene.

Discussion

Our searches have identified considerable evolutionary structure in the sequences; local optima are very much shorter than randomly selected trees, and these optima are much more similar than random trees. Some nucleotide positions are hypervariable; downweighting these results in the new optima being even more similar than with equal weighting (thus extracting more information from the data). There is strong support for the four related predictions on the average similarities of optima: the average similarity between optima increases with better optima (shorter-length trees); optima of any given length are more similar to each other than they are to poorer optima; similarly, better (shorter) optima of any given length are more similar to each other than optima of any longer length; and, most strikingly of all, optima of a given length L will, on the average, be more similar to better optima than they are to themselves. The distribution of distances between trees gives a unimodal distribution. Taken together, all these results support the conclusion that the optima are converging to a single global optimum, albeit with multiple side peaks.

A large cluster of 49 sequences is found repeatedly in the best optima, and such a large regional cluster is overwhelming evidence against the universal migration version of the multiregional model, which requires continued genetic intermingling over the entire range of *Homo erectus*. The sequences appear to fall into two groups, one "exclusively African" with 49 sequences, the other a "general" group of Africans, Europeans, Asians, and Australopapuans (86 sequences). Thus, Europeans, Asians, and Australopapuans appear to be a subset of the diversity within Africa. An analogy is that sequences of Native Americans (or of Polynesians) are subsets of Asian lineages, and these Asian lineages are subsets of those occurring in Africa. The use of sequences detects the split within Africa between the exclusively

African and the general groups, whereas the use of population frequencies masks this split and finds only an African/non-African grouping (Nei and Roychoudhury 1993). Only with significantly worse optima did the large African group, or its !Kung plus Naron subgroup, start to break up. This pattern contradicts any multiregional model that relies on migration and interbreeding over the whole range of *H. erectus*.

The root of the tree is best placed within the African group of 49 sequences. This placement is supported by a variety of criteria that require different conditions to be met to identify the correct root. Because of the African location of the root and the convergence of shorter optima, we call this the Kilimanjaro landscape. However, the time of the last common ancestor of the mtDNA must be earlier than the time of expansion of human numbers. In order to fit the data, we suggest at least a two-stage model with much of the initial variation lost during the first phase of constant average population size, and then a phase of exponential growth. Overall there is strong support for all four aspects of the out-of-Africa model but contradiction of a crucial prediction (panmixis) of the multiregion model.

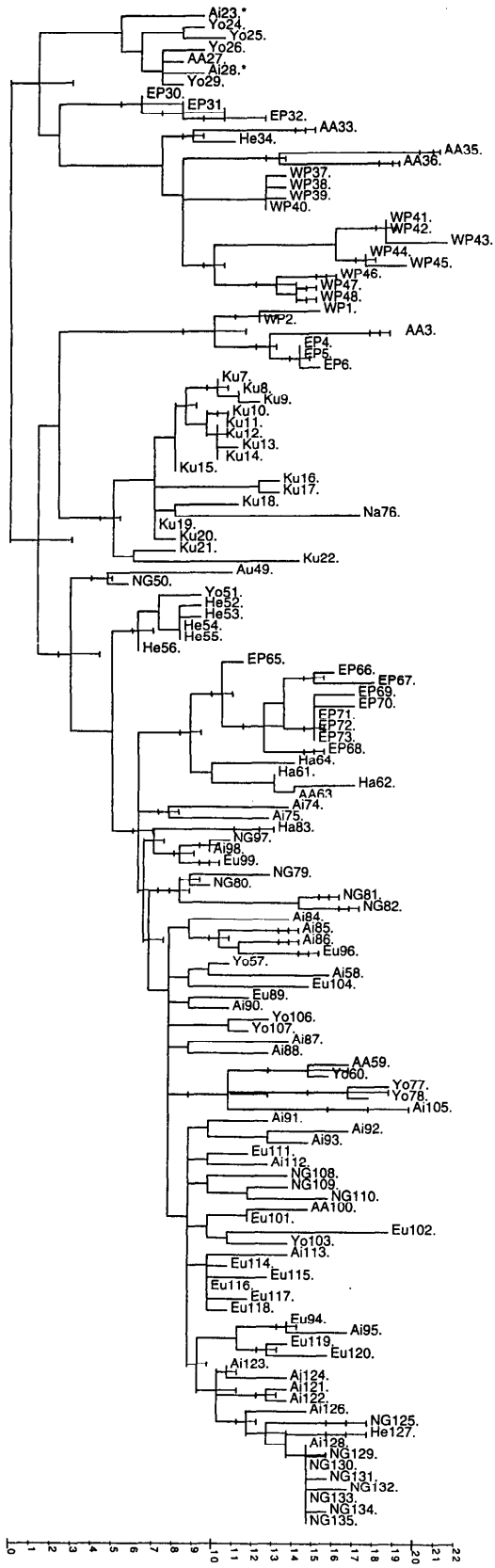
By focusing on the general properties of the landscape of all possible solutions (Kauffman 1993), we have been able to avoid the distraction of whether a particular tree is shorter than earlier studies. Because of the consistency of our results (for example, the !Kung, together with the Naron sequence, being united) we are confident that we have a very good representation of the underlying tree structure in the data. There has been too much concern with the absolute length of trees and not enough on the general properties (landscape) among optima. Perhaps we have contributed to this by showing that shorter trees are generally better predictors than longer trees (Penny and Hendy 1985). In the present study the single shortest tree was not as good a predictor of overall results as some trees one step longer.

The deluge search strategy has been useful, and this quality is illustrated by the high consistency of the groups it finds (tables 1 and 2). Several ways of improving its current implementation were developed during the study, but these were not used so as to maintain comparability for the different data sets (equalW, smallW, and mediumW). Although we now have some idea of the landscape in terms of the high numbers of local optima and their overall similarity (Penny et al. 1982; Hendy et al. 1988; Maddison 1991; Steel 1993; D. Penny and M. Steel, unpublished data), we still need to know more on the pathways between optima (Page 1993) and how the optima would appear with alternative selection criteria—for example, maximum likelihood.

The restriction of using random starting trees may not be necessary except when it is crucial to demonstrate there is no effect of starting position. Other programs, such as PAUP (Swofford 1993), are suitable for a similar study by retaining only one optimum per run, thus allowing an analysis of the landscape. We have found that using alternative methods of changing trees (different tree spaces; D. Steel and M. Penny, unpublished data) is particularly effective, and this option could also be incorporated, as could alternating stochastic and deterministic phases to the search strategy.

Our analysis was carried out *without* drawing the trees to study them, both to prevent subjective bias and emphasize the objective nature of the analysis. However, we could not resist the temptation to examine one of the shortest trees, that shown in figure 6. It is one of the second shortest trees (L=251.8) and is the best consensus trees (averaged over both the linear and quadratic forms). It is drawn with zero-length edges contracted and so is shown as a nonbinary tree. After drawing the tree (using MacClade 3.1; Maddison and Maddison 1992), it became clear that most of the zero-length edges were outside the core group of 49 African sequences. The observation of many zero-length edges among groups outside Africa is significant from two aspects: it is in agreement with the two-phase model presented earlier (fig. 5), and it explains the distribution of distance values used to infer a recent expansion (Rogers and Harpending 1992; Harpending et al. 1993), particularly of populations outside of Africa. This agreement between the distribution of edge lengths in the optimal trees in this study with the distribution of distance values (Rogers and Harpending 1992) is encouraging and represents another case in which there is reinforcement between different approaches to analyzing this data set.

Our results do not contribute any new information to the debate over the time of the last common ancestor of human mtDNA. Recent work (Hasegawa et al. 1993; Ruvolo et al. 1993) still favors a best estimate of about 200,000 yr ago, but an estimate using three complete human mtDNA sequences is even more recent (145,000 yr BP; Horai et al. 1995), thus reinforcing our conclusion (below) that more attention should be paid to the lower bound on the time estimates. For our model we have split this into two equal periods: 75,000–100,000 yr for a constant average population size and 75,000–100,000 for the expansionary phase. These values will have quite high variances and also depend on estimates of the size of the population that included the common ancestor. These values need corroboration from other data but are encouraging in that models are becoming more precise, include information from a wider range of disciplines, and are potentially more testable (Popper 1959). But until a similar study considers all the sources of variation as for human-chimpanzee times of divergence



(Waddell and Penny 1995), we prefer to consider the two-phase model as a qualitative model. Griffiths and Tavaré (1994*a*, 1994*b*) have recently extended coalescence methods to expanding populations; which will improve our understanding of dates of human evolution. Their study of Nuu Chah Nulth people gave a more recent origin when expansion in population size was considered. An expansion, such as the Polynesians (Lum et al. 1994), would be another good test.

Several investigators concentrate on the upper bound on the time of the last common mtDNA ancestor as one approach to testing a multiregional model; however, there has been less focus on the lower bound, the most recent time possible for human expansion out of Africa. There is a rather loosely defined cultural-archeological model (the "cultural explosion" model) that favors a very recent major expansion in human numbers, even as recently as 40,000 yr ago (the 40,004 BC model; Noble and Davidson 1991). This would be after the development of many modern artifacts and cultural attributes, including speech (as distinguished from language; Corballis 1993). Dates such as the colonization of Australia (at least 55,000 yr BP; Roberts et al. 1990, 1994) perhaps favor a somewhat earlier expansion (though it is not clear that these very early fossils have left descendants among modern populations). Our two-phase model, with an exponential increase in population numbers within the last 100,000 yr, including a slightly later expansion out-of-Africa, appears consistent with ideas of a recent expansion of modern humans (Noble and Davidson 1991). Despite earlier controversy (Cavalli-Sforza 1991), it can be shown (by using a different tree comparison metric to the one used here) that trees from gene frequencies and languages are very similar (Penny et al. 1993). However, our results would also be consistent with a third, even more rapid, expansion which may have occurred after the initial expansion of *H. sapiens* out of Africa.

Any useful hypothesis on the origins of *H. sapiens* needs to be consistent with a range of evidence, including nuclear and mitochondrial sequences, stochastic mechanisms of mutation, other genetic data such as allele frequencies, archeological finds, times of arrival of *H.*

FIG. 6.—A best local optimum tree. It was selected as being both one of the shortest trees and one of the median trees. *AA,* Afroamerican; *Ai,* Asian; *Au,* Australian; *Ep,* eastern Pygmy; *Eu,* European; *Ha,* Hadza; *He,* Herero; *Ku,* !Kung; *Na,* Naron; *NG,* New Guinean; *Wp,* western Pygmy; *Yo,* Yoruban. The lengths of edges on the tree are proportional to the number of changes required, with the error bars indicating the range of possible lengths consistent with the minimal length. Edges of zero length have been collapsed, and so the tree is nonbinary. The scale is number of mutations.

sapiens in different regions, appearance of "modern" artifacts in the archeological record, and the similarity of trees from languages and genetic information (Cavalli-Sforza 1991; Penny et al. 1993). We think a two- or three-phase model (fig. 5) is consistent with this wide range of classes of evidence, and the model is thus strengthened by such a "consilience of induction" (Whewell 1967). Conversely, the multiregional model, apart from being contradicted by the evidence produced here, is weakened by being based on such a limited range of features.

One piece of information not explained is the apparent morphological similarity between fossil *H. erectus* and modern *H. sapiens* in the same region (Thorne and Wolpoff 1992). We have already commented on the difficulty of quantitative genetic tests on morphological data. It may be that the authors of the multiregional model were just unlucky in that the similarities are only due to chance, but an alternative is that, even given a recent African origin of *H. sapiens,* there could have been some gene flow between *H. erectus* and *H. sapiens* in some regions. There is no evidence for this from the maternally inherited mtDNA, which is the expected direction of genetic movement (in chimpanzees and gorillas [Pusey and Packer 1987], as well as for humans [Schapera 1960], female transfer is the norm). However, the possibility needs to be tested with nuclear sequences (Bowcock et al. 1994).

There are recent claims for fossil *H. erectus* in Java 1.8 million yr ago (Swisher et al. 1994) and early *H. sapiens* in China 200,000 yr BP (Tiemei et al. 1994). The Java claim (Swisher et al. 1994) does not distinguish between the two main hypotheses in that both accept an early dispersal of *H. erectus.* The claim (Tiemei 1994) of early *H. sapiens* in China certainly would be anomalous in being so early but, for a number of reasons, must still be treated with care until a full quantitative analysis of the morphological features is available. A recent quantitative analysis of morphological characters does favor a very recent expansion of humans from a single region (Waddle 1994). It is a long-standing problem that it is not yet possible to determine directly whether a fossil group is ancestral to any extant group. Increasingly, models of human evolution must be consistent with a wide range of evidence: gene frequencies, DNA sequences, timing and distribution of fossils, archeological findings, and languages. The time has gone when isolated fossil finds are sufficient to build theories of human origins, even though they are an essential component of any theory. It seems at times that there has scarcely been a fossil find that was not initially interpreted as a direct line to *H. sapiens,* but virtually none of the claims withstand further analysis. A qualitative attempt has been made to explain the genetic data (Thorne et al. 1993). We suggest that, having failed a major test in the present work (the universal migration version), the multiregional model cannot be taken seriously until it makes a real attempt to produce a quantitative model that accounts for the DNA sequence information. In many respects the archeological evidence has been less subject to continued revisions than has paleontological data, and archeological data appear to be in better agreement with the DNA sequence data.

To conclude, the out-of-Africa hypothesis has been subjected to numerous tests and is now the only hypothesis supported by a wide range of data. Alternatives such as the multiregional model fail crucial tests, in this case predicting gene dispersal over the whole range of human precursors. The support for the out-of-Africa hypothesis comes from mtDNA sequence and RFLP data, as well as from frequencies of nuclear alleles. It is consistent with much archeological work and the similarity of trees from genetics and languages. More work, particularly with nuclear sequence information, will clarify the hypotheses still further. We expect that in the next decade a fairly complete picture of human migration over the last 100,000 yr will emerge.

APPENDIX
Proofs of the Propositions
Proof of Proposition 1

Clearly if an edge e contains the root, then $\delta(e) = 2h_0$. Now, suppose e does not contain the root ρ. In this case we may represent T as in figure 7 where $r \geq 1$. For $0 \leq i \leq r + 1$, let L_i denote the leaf set of T_i, $L = L_0 \cup L_1 \cup \ldots \cup L_r$, $L' = L_{r+1}$, and $h_i = d(v_i, l_i)$, where $l_i \in L_i$ (by the molecular clock hypothesis, this value is the same for all choices of l_i from L_i). Then,

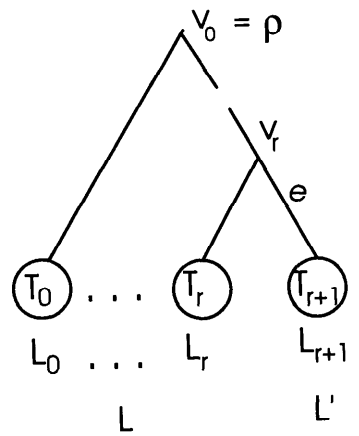


FIG. 7.—For this tree, the root ρ is the midpoint of an edge different to e, and so $\delta(e)$ is less than maximal (refer to the proof of proposition 1 in the Appendix).

$$\delta(e) = \frac{1}{|L||L'|} \sum_{x \in L, x' \in L'} d(x,x')$$

so that

$$\delta(e) = \frac{1}{|L|} \sum_{i=0}^{r} |L_i| 2h_i$$

(by the molecular clock hypothesis)

$< 2h_0$, since $h_i < h_0$ for $i > 0$,

thereby completing the proof.

Proof of Proposition 2

By independence and the assumption that k is large, the distribution is Poisson. Thus we need only calculate $\lambda = kp(T)$, where $p(T)$ is the probability that any individual founding female has a surviving line after T generations.

Consider first scenario 1. If μ denotes the mean of the offspring distribution, then the expected value of N is $k\mu^T$, so we may estimate μ as $(N/k)^{1/T} = \exp[(1/T)\ln(N/k)] \cong 1 + \ln(N/k)/T$, since $T \gg \ln(k/N)$. Now, $p(T)$ converges, in a monotonically decreasing fashion, to $1 - x$, where $x < 1$ is the solution to the equation $F(x) = x$, in which F is the probability-generating function for the offspring distribution (see Arthreya and Ney 1972). Expanding $F(z)$ as a power series about $z = 1$ gives $F(1-\varepsilon) = 1 - \mu\varepsilon + 0.5\varepsilon^2(\sigma^2 - \mu + \mu^2) + O(\varepsilon^3)$, so that the solution to the equation $1 - \varepsilon = F(1-\varepsilon)$, for ε small, is given approximately by $\varepsilon \cong 2(\mu - 1)/\sigma^2 - \mu + \mu^2)$.

Since $\mu \cong 1 + \ln(k/N)/T$, and $\ln(k/N)/T$ is small compared to $\min\{1, \sigma^2\}$, we have that $p(T) > 1 - x = \varepsilon \cong 2\ln(N/k)/T\sigma^2$. Since $\lambda = p(T)k$, we have the required result.

Considering scenario 2, $p(T) < p(T_1)$, and over the first T_1 generations, $\mu = 1$, and so, by a well-known and classical result (see Arthreya and Ney 1972), $p(T_1) \cong 2/\sigma^2 T_1$, leading to the claimed bound for λ.

Acknowledgments

We thank many people for challenge, comment, discussion, ideas, and information, including Colin Burrows, Dick Byrne, Ann Chowning, Iain Davidson, Blair Hedges, Robert Hickson, Rhys Jones, David Maddison, Mark Stoneking, Chris Stringer, plus one anonymous referee.

Note Added in Proof

The recent discovery in Africa (Zaire) of sophisticated tools about 90,000 yr BC (Yellen et al. 1995) is in agreement with the two-phase model developed here from DNA sequence data. These tools may be twice as old as equivalent ones from outside Africa, and so the times are consistent with expansion from Africa being considerably later than the time of the "last common ancestor." The "cultural explosion" model may be more appropriate than those based on supposed changes in climatic conditions.

LITERATURE CITED

AITKEN, M. J., and H. VALLADAS. 1992. Luminescence dating relevant to human origins. Philos. Trans. R. Soc. Lond. [Biol.] **337**:139–144.

ARCHIE, J. W., and J. FELSENSTEIN. 1993. The number of evolutionary steps on random or minimal length trees from random evolutionary data. Theor. Pop. Biol. **43**:52–79.

ARTHREYA, K. B., and P. E. NEY. 1972. Branching processes. Springer, Berlin.

BARTHÉLEMY, J.-P., and F. R. MCMORRIS. 1986. The median procedure for n-trees. J. Classif. **3**:329–334.

BOWCOCK, A. M., J. R. KIDD, J. L. MOUNTAIN, J. M. HEBERT, L. CAROTENUTO, K. K. KIDD, and L. L. CAVALLI-SFORZA. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc. Natl. Acad. Sci. USA **88**:1597–1601.

BOWCOCK, A. M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, J. R. KIDD, and L. L. CAVALLI-SFORZA. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368**:455–457.

BURROWS, C. J. 1994. Genesis: a model for the origin of the first hominids. Rebus, Christchurch.

CANN, R. L., M. STONEKING, and A. C. WILSON. 1987. Mitochondrial DNA and human evolution. Nature **325**:31–36.

CAVALLI-SFORZA, L. L. 1991, Nov. Genes, peoples, and languages. Sci. Am., pp. 72–78.

CORBALLIS, M. C. 1993. On the evolution of language and generativity. Cognition **44**:197–226.

DEACON, H. J. 1992. Southern Africa and modern human origins. Philos. Trans. R. Soc. Lond. [Biol.] **337**:177–183.

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.

DUECK, G. 1990. New optimization heuristics: the Great Deluge algorithm and record-to-record travel. Scientific Centre Technical Report. (IBM Germany, Tiergartenstrasse 15, D-6900 Heidelberg, Germany.)

DUECK, G., and T. SCHEUER. 1990. Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing. J. Comp. Physics **90**:161–175.

FARRIS, J. S. 1972. Estimating phylogenetic trees from distance matrices. Am. Nat. **106**:645–668.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**: 401–410.

FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**:406–416.

GOLDMAN, N., and N. H. BARTON. 1992. Genetics and geography. Nature **357**:440–441.

GREENBERG, B. D., J. E. NEWBOLD, and A. SUGINO. 1983. Intraspecific nucleotide sequence variability surrounding

the origin of replication in human mitochondrial DNA. Gene **21**:33–49.

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

———. 1994*b*. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. [Biol.] **344**:403–410.

GUÉNOCHE, A., P. HANSEN, and B. JAUMARD. 1991. Efficient algorithms for divisive hierarchical clustering with the diameter criterion. J. Classif. **8**:5–30.

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.

HASEGAWA, M. A., DI RIENZO, T. D. KOCHER, and A. C. WILSON. 1993. Toward a more accurate time scale for the human mitochondrial tree. J. Mol. Evol. **37**:347–354.

HEDGES, S. B., S. KUMAR, K. TAMURA, and M. STONEKING. 1992. Human origins and analysis of mitochondrial DNA sequences. Science **255**:737–739.

HENDY, M. D., L. R. FOULDS, and D. PENNY. 1980. Proving phylogenetic trees minimal with *l*-clustering and set partitioning. Math. Biosci. **51**:71–88.

HENDY, M. D., and D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. **38**:297–309.

HENDY, M. D., M. A. STEEL, D. PENNY, and I. M. HENDERSON. 1988. Families of trees and consensus. Pp. 355–362 *in* H. H. BOCK, ed. Classification and related methods of data analysis. Elsevier Scientific, Amsterdam.

HILLIS, D. M., J. P. HUELSENBECK, and C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenies. Science **264**:671–677.

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.

HUDSON, R. R. 1990. Gene genealogies and the coalescent process. Oxford Surv. Evol. Biol. **7**:1–41.

JOHNSON, M. J., D. C. WALLACE, S. D. FERRIS, M. C. RATTAZZI, and L. L. CAVALLI-SFORZA. 1983. Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. J. Mol. Evol. **19**:255–271.

KAUFFMAN, S. A. 1993. The origins of order. Oxford University Press, Oxford.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of Hominoidea. J. Mol. Evol. **29**:170–179.

KLEIN, J., N. TAKAHATA, and F. J. AYALA. 1993. MHC polymorphisms and human origins. Sci. Am. **269**:46–51.

KUHNER, M. K., and J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal rates. Mol. Biol. Evol. **11**:459–468.

LAPOINTE, F.-J., and P. LEGENDRE. 1992. Statistical significance of the matrix correlation coefficient for comparing independent phylogenetic trees. Syst. Biol. **41**:378–384.

LEAMY, L. 1993. Morphological integration of fluctuating asymmetry in the mouse mandible. Genetica **89**:139–153.

LUM, J. K., O. RICKARDS, C. CHING, and R. L. CANN. 1994. Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum. Biol. **66**:567.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. **40**: 315–328.

MADDISON, D. R., M. RUVOLO, and D. L. SWOFFORD. 1992. Geographic origins of human mitochondrial DNA phylogenetic inference from control region sequences. Syst. Biol. **41**:111–124.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade, version 3. Sinauer, Sunderland, Mass.

MURDOCK, G. P. 1960. Social structure. 2d ed. Macmillan, New York.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., and A. K. ROYCHOUDHURY. 1982. Genetic relationship and evolution of human races. Evol. Biol. **14**:1–59.

———. 1993. Evolutionary relationships of human populations on a global scale. Mol. Biol. Evol. **10**:927–943.

NOBLE, W., and I. DAVIDSON. 1991. The evolutionary emergence of modern human behavior: language and its archaeology. Man **26**:223–254.

PAGE, R. D. M. 1993. On the islands of trees and the efficiency of different methods of branch swapping in finding most-parsimonious trees. Syst. Biol. **42**:200–210.

PENNY, D. 1976. Criteria for optimizing phylogenetic trees and the problem of determining the root of a tree. J. Mol. Evol. **8**:95–116.

PENNY, D., L. R. FOULDS, and M. D. HENDY. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature **297**:197–200.

PENNY, D., and M. D. HENDY. 1985. Testing methods of evolutionary tree construction. Cladistics **1**:266–278.

PENNY, D., M. D. HENDY, and M. A. STEEL. 1992. Progress with evolutionary trees. Trends Ecol. Evol. **7**:73–79.

PENNY, D., and P. PENNY. 1990. Search, parallelism, comparison and evaluation. Pp. 75–91 *in* A. DRESS and A. VON HAESELER, eds. Trees and hierarchical structures. Lecture Notes in Biomathematics 84. Springer, Berlin.

PENNY, D., E. E. WATSON, and M. A. STEEL. 1993. Trees from languages and genes are very similar. Syst. Biol. **42**: 382–384.

POPPER, K. R. 1959. The logic of scientific discovery. Hutchinson, London.

PRAGER, E. M., and A. C. WILSON. 1988. Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. J. Mol. Evol. **27**:326–335.

PUIT, I., A. SAJANTILA, J. SIMANAINEN, O. GEORGIEV, W. SCHAFFNER, and S. PÄÄBO. 1994. Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol. Chem. Hoppe-Seyler **375**:837–840.

PUSEY, A. E., and C. PACKER. 1987. Dispersal and philopatry. Pp. 250–266 *in* B. B. SMUTS, D. L. CHENEY, R. M. SEY-

FARTH, T. T. STRUHSAKER, and R. W. WRANGHAM, eds. Primate societies. University of Chicago Press, Chicago.

ROBERTS, R. G., R. JONES, and M. A. SMITH. 1990. Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. Nature **345**:153–156.

ROBERTS, R. G., R. JONES, N. A. SPOONER, M. J. HEAD, A. S. MURRAY, and M. A. SMITH. 1994. The human colonization of Australia: optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. Quatern. Geochr. **13**:575–584.

ROGERS, A. R., and H. C. HARPENDING. 1992. Population growth makes waves in the distribution of pairwise genetic distances. Mol. Biol. Evol. **9**:552–569.

ROSS, P. E. 1992. Shaking the tree. Sci. Am. **267**(12):18–19.

RUVOLO, M., S. ZEHR, M. VON DORNUM, D. PAN, B. CHANG, and J. LIN. 1993. Mitochondrial COII sequences and modern human origins. Mol. Biol. Evol. **10**:1115–1135.

SCHAPERA, I. 1960. The Khoisan peoples of South Africa: bushmen and Hottentots. Routledge & Kegan Paul, London.

SCHWARCZ, H. P., and R. GRÜN. 1992. Electron spin resonance (ESR) dating of the origin of modern man. Philos. Trans. R. Soc. Lond. [Biol.] **337**:144–148.

SEUÁNEZ, H. N. 1979. The phylogeny of human chromosomes. Springer, Berlin.

SMITH, A. N. 1994. Rooting molecular trees: problems and strategies. Biol. J. Linn. Soc. **51**:279–292.

STEEL, M. A. 1993. Distributions on bicolored binary trees arising from the principle of parsimony. Discrete Appl. Math. **41**:245–261.

———. 1994. The maximum likelihood point is not unique. Syst. Biol. **43**:560–564.

STEEL, M. A., and M. A. CHARLESTON. 1995. Five surprising properties of parsimoniously colored trees. Bull. Math. Biol. **57**:367–375.

STEEL, M. A., M. D. HENDY, and D. PENNY. 1993. Parsimony can be consistent! Syst. Biol. **42**:581–587.

———. 1995. Invertible models of sequence evolution. Math. Biosci. (in press).

STEEL, M. A., and D. PENNY. 1993. Distributions of tree comparison metrics—some new results. Syst. Biol. **42**:126–141.

STONEKING, M. A. 1994. In defence of "Eve"—a response to Templeton's critique. Am. Anthropol. **96**:131–141.

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

SWISHER, C. C., III, G. H. CURTIS, T. JACOB, A. G. GETTY, A. SUPRIJO, and WIDIASMORO. 1994. Age of the earliest known hominids in Java, Indonesia. Science **263**:1118–1121.

SWOFFORD, D. L. 1993. PAUP, version 3.0. Illinois Natural History Survey, Urbana.

SWOFFORD, D. L., and G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–501 *in* D. M. HILLIS and G. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

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

TEMPLETON, A. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution **37**: 221–244.

———. 1992. Human origins and analysis of mitochondrial DNA sequences. Science **255**:737.

———. 1993. The "Eve" hypothesis: a genetic critique and reanalysis. Am. Anthropol. **95**:51–72.

THORNE, A. G., and M. H. WOLPOFF. 1992, April. The multiregion evolution of humans. Sci. Am. **266**:28–33.

THORNE, A. G., M. H. WOLPOFF, and R. B. ECKHARDT. 1993. Genetic variation in Africa. Science **261**:1507–1508.

TIEMEI, C., Y. QUANG, and W. EU. 1994. Antiquity of *Homo sapiens* in China. Nature **368**:55–56.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES, and A. C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.

WADDELL, P. J., and D. PENNY. 1995. Evolutionary trees of apes and humans from DNA sequences. *In* A. LOCK and C. R. PETERS, eds. Handbook of human symbolic evolution. Clarendon Press, Oxford (in press).

WADDLE, D. M. 1994. Matrix correlation tests support a single origin for modern humans. Nature **368**:452–454.

WAKELEY, J. 1993. Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. J. Mol. Evol. **37**:613–623.

WHEWELL, W. 1967. The philosophy of the inductive sciences: founded upon their history. Vol. **2**. 2d ed. Cass, London. (Reprint of 1847 original.)

WILSON, A. C., M. STONEKING, and R. L. CANN. 1991. Ancestral geographic states and the peril of parsimony. Syst. Zool. **40**:363–365.

WOLPOFF, M. H., W. X. ZHI, and A. G. THORNE. 1984. Modern *Homo sapiens* origins: a general theory of hominid evolution involving the fossil evidence from East Asia. Pp. 411–483 *in* F. H. SMITH and F. SPENCER, eds. The origins of modern humans: a world survey of the fossil evidence. Liss, New York.

YELLEN, J. E., A. S. BROOKS, E. CORNELISSEN, M. J. MEHLMAN, and K. STEWART. 1995. A middle stone age worked bone industry from Katanda, Upper Semliki Valley, Zaire. Science **268**:553–556.

SIMON EASTEAL, reviewing editor

Received August 31, 1994

Accepted March 31, 1995