Human Origins and Analysis of Mitochondrial DNA Sequences

L. Vigilant *et al.* (*1*) recently presented "the strongest support yet for the placement of [their] common mtDNA [mitochondrial DNA] ancestor in Africa some 200,000 years ago." This support stems from a tree estimated by maximum parsimony from mtDNA sequence data with the use of the computer program PAUP (*2*). The African origin is inferred from this tree because (i) the most basal splits are among purely African lineages and (ii) an African origin is favored over alternatives hypothesizing a non-African origin on the basis of statistical tests that use the estimated maximum parsimony tree as the reference tree.

The single African origin hypothesis was first inferred with the use of argument (i) from a maximum parsimony tree estimated from mtDNA restriction site data (*3*). The new support of Vigilant *et al.* is critical because Maddison (*4*) has recently found 10,000 trees more parsimonious by five steps than the mtDNA restriction site "maximum parsimony tree" given by Cann *et al.* (*3*). Maddison's set of maximum parsimony trees contains cladograms with geographically mixed basal clades, thereby invalidating the original rationale for an African origin.

The phylogenetic analysis of the mtDNA sequence data is similarly flawed. Apparently, a single heuristic run with simple, sequential addition was used for the analysis of the sequence data (*1*). Such an analysis is inadequate for a data set this large, and it is critical to use random addition to avoid artifacts arising from the order of data analysis (*5*). To illustrate this inadequacy, I performed a single heuristic run on the mtDNA sequence data (kindly provided by M. Stoneking) using the random addition option of PAUP 3.0, but otherwise retaining the same parameter values used in the original analysis. I found 100 trees that are two steps more parsimonious than the tree presented by Vigilant *et al.* Figure 1 illustrates the first tree found in this search. The most basal clade in this more parsimonious tree is non-African, and non-African haplotypes tend to be the more ancient. A single random heuristic run is also an inadequate analysis, and this alternative tree is not significantly different from the tree in Vigilant *et al.* if one uses my nonparametric test (*6*). However, the existence of this more parsimonious cladogram undercuts the validity of argument (*i*).

This more parsimonious tree also invalidates the statistical analysis given in Vigilant *et al.* because that analysis is dependent on their "maximum parsimony" reference cladogram. Other serious flaws with their statistics include their estimation of the time of origin (*7*).

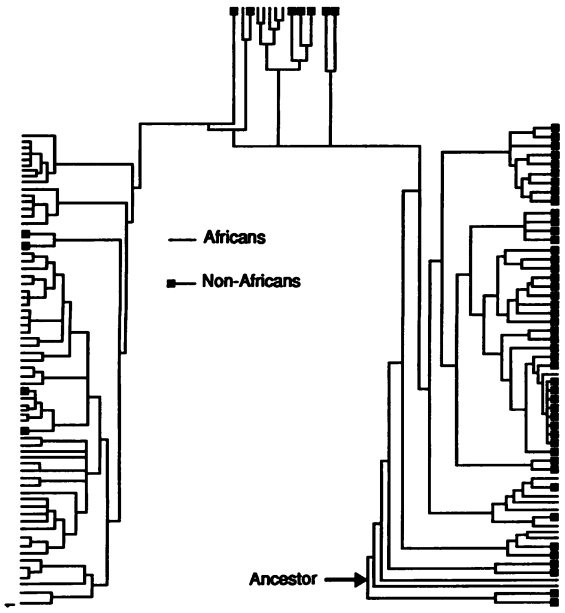

Fig. 1. An alternative cladogram inferred from the mitochondrial DNA sequence data of Vigilant *et al.* (*1*). This cladogram is more parsimonious by two steps than that given in Vigilant *et al.* The sequence of haplotype numbers in this cladogram (according to the numbering system of Vigilant *et al.*) is, starting with 1 on the extreme left, 1–18, 76, 19–26, 29, 27, 28, 30–56, 65–73, 68, 104, 57–63, 84–95, 119, 120, 96–99, 108–110, 113, 118, 114, 116, 117, 115, 121–125, 127–135, 126, 111, 112, 100, 101, 105, 107, 106, 102, 103, 77–80, 74, 75, 64, 83, 81, and 82.

It is important to recognize the critical need for performing rigorous phylogenetic and statistical analyses of molecular data in making evolutionary inferences. A single heuristic run of the computer program PAUP with simple addition is inadequate for a phylogenetic analysis of large data sets.

ALAN R. TEMPLETON
Department of Biology,
Washington University,
St. Louis, MO 63130

REFERENCES AND NOTES

1. L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A. C. Wilson, *Science* **253**, 1503 (1991).
2. D. L. Swofford, *PAUP 3.0 User's Manual (Draft 2/9/91)* (Illinois Natural History Survey, Champaign, 1991).
3. R. L. Cann, M. Stoneking, A. C. Wilson, *Nature* **325**, 31 (1987).
4. D. R. Maddison, *Syst. Zool.* **40**, 355 (1991).
5. T. Crease, thesis, Washington University, St. Louis (1986); D. R. Maddison, *Syst. Zool.* **40**, 315 (1991).
6. A. R. Templeton, in *Statistical Analysis of DNA Sequence Data*, B. S. Weir, Ed. (Dekker, New York, 1983), pp. 151–179. The nonparametric test for DNA sequence data given here is the same as the "winning sites test" when all informative differences are due to single mutational transitions. Hence, this is the same test used in (*1*).
7. ———, *Am. Anthropol.*, in press.

15 November 1991; accepted 14 January 1992

A recent analysis of human mitochondrial DNA sequences from widely distributed populations (*1*) resulted in a phylogenetic tree that supported an African origin for human mitochondrial DNA. This finding, with the use of the method of maximum parsimony, was shown to be significant with two statistical tests. We have reanalyzed these data with the same method and another method (neighbor-joining), and our results do not show statistical resolution for the geographic origin of human mitochondrial DNA.

For both of our phylogenetic analyses, we used the data set of the original study (*1, 2*). Our maximum parsimony analysis resulted in a large number of equally parsimonious trees of 523 steps (*3*), five steps shorter than in the original analysis. As would be expected in a parsimony analysis, when the number of sequences (136 humans) is larger than the number of characters (117 informative sites), there is a large (and in this case unknown) number of maximum parsimony (MP) trees (*4*). Because individual MP trees are not necessarily generated randomly from the total set of MP trees, any subset is likely to be biased by the order in which the sequences in the analysis are added (*5*). To avoid this bias we performed five separate analyses, each with sequences added randomly, 10^4 MP trees saved, and a majority-rule consensus tree generated. Each of the five majority-rule trees was considerably dif-

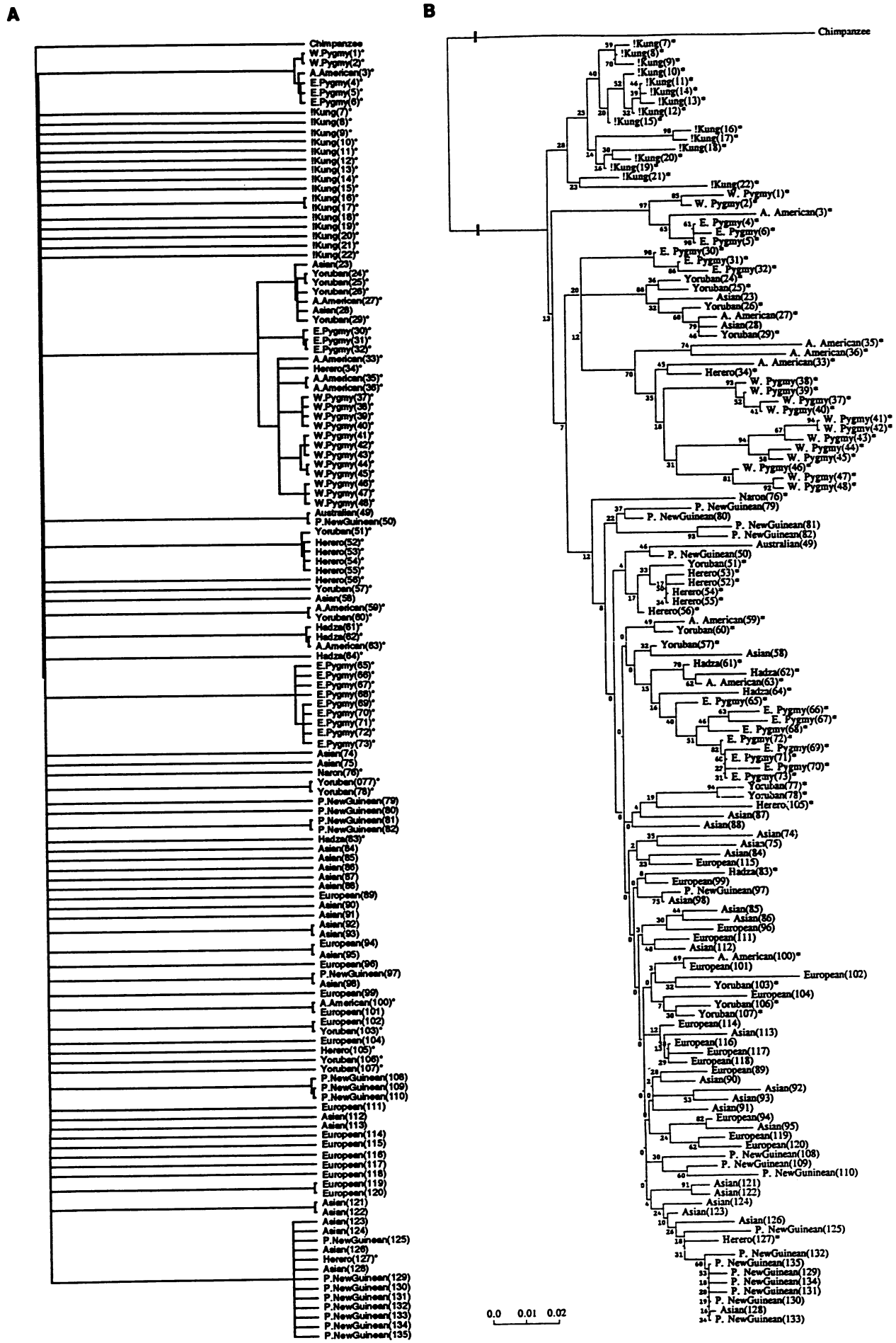
A
Chimpanzee
W.Pygmy(1)*
W.Pygmy(2)*
A.American(3)*
E.Pygmy(4)*
E.Pygmy(5)*
E.Pygmy(6)*
!Kung(7)*
!Kung(8)*
!Kung(9)*
!Kung(10)*
!Kung(11)*
!Kung(12)*
!Kung(13)*
!Kung(14)*
!Kung(15)*
!Kung(16)*
!Kung(17)*
!Kung(18)*
!Kung(19)*
!Kung(20)*
!Kung(21)*
!Kung(22)*
Asian(23)
Yoruban(24)*
Yoruban(25)*
Yoruban(26)*
A.American(27)*
Asian(28)
Yoruban(29)*
E.Pygmy(30)*
E.Pygmy(31)*
E.Pygmy(32)*
A.American(33)*
Herero(34)*
A.American(35)*
A.American(36)*
W.Pygmy(37)*
W.Pygmy(38)*
W.Pygmy(39)*
W.Pygmy(40)*
W.Pygmy(41)*
W.Pygmy(42)*
W.Pygmy(43)*
W.Pygmy(44)*
W.Pygmy(45)*
W.Pygmy(46)*
W.Pygmy(47)*
W.Pygmy(48)*
Australian(49)
P.NewGuinean(50)
Yoruban(51)*
Herero(52)*
Herero(53)*
Herero(54)*
Herero(55)*
Herero(56)*
Yoruban(57)*
Asian(58)
A.American(59)*
Yoruban(60)*
Hadza(61)*
Hadza(62)*
A.American(63)*
Hadza(64)*
E.Pygmy(65)*
E.Pygmy(66)*
E.Pygmy(67)*
E.Pygmy(68)*
E.Pygmy(69)*
E.Pygmy(70)*
E.Pygmy(71)*
E.Pygmy(72)*
E.Pygmy(73)*
Asian(74)
Asian(75)
Naron(76)*
Yoruban(077)*
Yoruban(78)*
P.NewGuinean(79)
P.NewGuinean(80)
P.NewGuinean(81)
P.NewGuinean(82)
Hadza(83)*
Asian(84)
Asian(85)
Asian(86)
Asian(87)
Asian(88)
European(89)
Asian(90)
Asian(91)
Asian(92)
Asian(93)
European(94)
Asian(95)
European(96)
P.NewGuinean(97)
Asian(98)
European(99)
A.American(100)*
European(101)
European(102)
Yoruban(103)*
European(104)
Herero(105)*
Yoruban(106)*
Yoruban(107)*
P.NewGuinean(108)
P.NewGuinean(109)
P.NewGuinean(110)
European(111)
Asian(112)
Asian(113)
European(114)
European(115)
European(116)
European(117)
European(118)
European(119)
European(120)
Asian(121)
Asian(122)
Asian(123)
Asian(124)
P.NewGuinean(125)
Asian(126)
Herero(127)*
Asian(128)
P.NewGuinean(129)
P.NewGuinean(130)
P.NewGuinean(131)
P.NewGuinean(132)
P.NewGuinean(133)
P.NewGuinean(134)
P.NewGuinean(135)
B
Chimpanzee
!Kung(7)*
!Kung(8)*
!Kung(9)*
!Kung(10)*
!Kung(11)*
!Kung(14)*
!Kung(13)*
!Kung(12)*
!Kung(15)*
!Kung(16)*
!Kung(17)*
!Kung(18)*
!Kung(20)*
!Kung(19)*
!Kung(21)*
!Kung(22)*
W. Pygmy(1)*
W. Pygmy(2)*
A. American(3)*
E. Pygmy(4)*
E. Pygmy(6)*
E. Pygmy(5)*
E. Pygmy(30)*
E. Pygmy(31)*
E. Pygmy(32)*
Yoruban(24)*
Yoruban(25)*
Asian(23)
Yoruban(26)*
A. American(27)*
Asian(28)
Yoruban(29)*
A. American(35)*
A. American(36)*
A. American(33)*
Herero(34)*
W. Pygmy(38)*
W. Pygmy(39)*
W. Pygmy(37)*
W. Pygmy(40)*
W. Pygmy(41)*
W. Pygmy(42)*
W. Pygmy(43)*
W. Pygmy(44)*
W. Pygmy(45)*
W. Pygmy(46)*
W. Pygmy(47)*
W. Pygmy(48)*
Naron(76)*
P. NewGuinean(79)
P. NewGuinean(80)
P. NewGuinean(81)
P. NewGuinean(82)
Australian(49)
P. NewGuinean(50)
Yoruban(51)*
Herero(53)*
Herero(52)*
Herero(54)*
Herero(55)*
Herero(56)*
A. American(59)*
Yoruban(60)*
Yoruban(57)*
Asian(58)
Hadza(61)*
Hadza(62)*
A. American(63)*
Hadza(64)*
E. Pygmy(65)*
E. Pygmy(66)*
E. Pygmy(67)*
E. Pygmy(68)*
E. Pygmy(72)*
E. Pygmy(69)*
E. Pygmy(71)*
E. Pygmy(70)*
E. Pygmy(73)*
Yoruban(77)*
Yoruban(78)*
Herero(105)*
Asian(87)
Asian(88)
Asian(74)
Asian(75)
Asian(84)
European(115)
Hadza(83)*
European(99)
P. NewGuinean(97)
Asian(98)
Asian(85)
Asian(86)
European(96)
European(111)
Asian(112)
A. American(100)*
European(101)
European(102)
Yoruban(103)*
European(104)
Yoruban(106)*
Yoruban(107)*
European(114)
Asian(113)
European(116)
European(117)
European(118)
European(89)
Asian(90)
Asian(92)
Asian(93)
Asian(91)
European(94)
Asian(95)
European(119)
European(120)
P. NewGuinean(108)
P. NewGuinean(109)
P. NewGuinean(110)
Asian(121)
Asian(122)
Asian(124)
Asian(123)
Asian(126)
P. NewGuinean(125)
Herero(127)*
P. NewGuinean(132)
P. NewGuinean(135)
P. NewGuinean(129)
P. NewGuinean(134)
P. NewGuinean(131)
P. NewGuinean(130)
Asian(128)
P. NewGuinean(133)
0.0 0.01 0.02

Fig. 1. Phylogenetic trees showing the relationships of human mitochondrial DNA sequences (692 sites) from data of Vigilant *et al.* (*1*). Africans are identified by asterisks. A. American denotes African American. (**A**) A strict consensus tree of 50,000 maximum parsimony trees, each constructed by the maximum parsimony method. (**B**) A neighbor-joining tree showing bootstrap *P* values (0 to 100%) for each node. Nucleotide substitutions per site are indicated on scale at lower left.

ferent from one another, which confirms that a large number of MP trees exist and that different subsets are biased. Although the two to ten most basal nodes in the five majority-rule trees lead exclusively to Africans, the branching order of even those lineages differs among the five trees. To determine the groups supported in all MP trees, we obtained a strict consensus tree (*6*) of the 5×10^4 MP trees (Fig. 1A). Although this number of trees represents only a small fraction of the total set of MP trees, the poor resolution of relationships (Fig. 1A) indicates that parsimony analysis is unable to resolve the deep branches of the tree. Additional MP trees would not alter that conclusion.

Our neighbor-joining reanalysis (*7*) resulted in a single tree showing some geographic cohesiveness among the Africans (Fig. 1B). Most notably, all 16 !Kung form a group, in contrast with the original tree (*1*) where they were placed as 13 independent deep branches. This difference is important because it was the deep branching of the !Kung that provided statistical support for an African origin. Although the two deepest branches of our neighbor-joining tree lead exclusively to Africans (!Kung and Pygmies), those bifurcations are not statistically supported (bootstrap, $P = 0.13$ and $P = 0.07$, respectively). Only six nodes in the tree, all defining small clusters (two to six individuals), are statistically significant (bootstrap, $P \geq 0.95$).

The reason that this reanalysis differs so greatly from the original study (*1*) is that the tree on which the first conclusions were drawn was not representative of the total set of MP trees. Thus, the two statistical tests made in the original analysis are not valid. Those tests cannot be performed on the trees presented in Fig. 1 because their branching order is not statistically resolved. Although an African origin for humans is supported by other kinds of data and other molecular data (*8*), and is suggested by the mtDNA sequence data (Fig. 1B), the available sequence data are insufficient to statistically resolve the geographic origin of human mitochondrial DNA.

Templeton concludes that the original phylogenetic analysis (*1*) was inadequate for the same reasons described here. However, we note that the 100 trees he found are four steps longer than the 50,000 trees we have analyzed (*6*); hence, the tree he presents (his figure 1) is not an MP tree. Furthermore, the African origin hypothesis was not derived solely from the phylogenetic analysis; patterns of mtDNA variation within different human populations also have been used to support an African origin (*1, 9*).

What data are needed to resolve the evolutionary history of our species if this data set, perhaps the largest available, is insufficient? The absence of a strong association between mtDNA sequence and geography, especially among the non-Africans (Fig. 1B), suggests that the same multiple mtDNA types have been maintained in widely separated populations since those populations diverged, thus confounding an evolutionary interpretation of the data. DNA sequence data from multiple nuclear genes, in combination with the mtDNA sequence data, likely will be needed to overcome the effect of individual gene phylogenies. We then may be able to gain a better perspective of human origins and evolution.

S. BLAIR HEDGES
SUDHIR KUMAR
KOICHIRO TAMURA
Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park, PA 16802
MARK STONEKING
Institute of Molecular Evolutionary Genetics and Department of Anthropology, Pennsylvania State University, University Park, PA 16802

REFERENCES AND NOTES

1. L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A. C. Wilson, *Science* **253**, 1503 (1991).
2. The data set consists of 136 different mtDNA sequences, each with 1137 sites. The original analysis and this analysis were performed only with sites 1–358 and 604–937; most other sites were missing information. Of the 692 sites used, 219 were variable and 117 were informative for the parsimony analyses.
3. PAUP [D. L. Swofford, PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0, Computer program (Illinois Natural History Survey, Champaign, 1990)] was used with the following options: heuristic search, simple addition sequence, hold = 100 trees, tree bisection-reconnection branch swapping, and maxtrees = 1000; and 1000 MP trees of length 523 were obtained out of a presumably large and unknown number. This MP tree length is five steps shorter than that obtained in the original study, probably because of more efficient search options. This difference, and the slight differences in the numbers of variable and informative sites used in the two studies, are not responsible for the major differences in the conclusions of these studies.
4. These are 10^{267} possible bifurcating trees for this data set; the number of MP trees is unknown, but almost certainly is much larger than 1 billion.
5. Different "islands" of MP trees may exist [M. D. Hendy, M. A. Steel, D. Penny, I. M. Henderson, in *Classification and Related Methods of Data Analysis*, H. H. Bock, Ed. (Elsevier, Amsterdam, The Netherlands, 1988), pp. 355–362].
6. In order to obtain representative samples of MP trees from the total (unknown) number, we used the random addition sequence of PAUP with maxtrees = 10^4 and obtained strict, semistrict, and majority-rule consensus trees of those 10^4 MP trees, each of length 522 (one step shorter because of the increase in "maxtrees"). This was repeated five times with different random numbers (for the additional sequence), and a strict consensus tree was made of the five separate strict consensus trees. The total number of different MP trees in this sampling probably is fewer than 5×10^4 because of possible overlap between the five subsets, although the differences in the majority-rule consensus trees suggest that there is little, if any, overlap. A strict consensus tree is used because there is no a priori reason to favor one MP tree over another (the length of this strict tree, 545 steps, is much longer than the length of each individual tree). A semistrict consensus tree showing only uncontested groupings was nearly identical to the strict tree.
7. The neighbor-joining method [N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987)] was used with the proportion distance (*p*); a very similar tree was obtained with the Jukes-Cantor distance. Statistical significance of the groups on the tree was determined by the bootstrap method [J. Felsenstein, *Evolution* **39**, 783 (1985)] with 2000 replications (S. B. Hedges, *Mol. Biol. Evol.*, in press).
8. C. B. Stringer and P. Andrews, *Science* **239**, 1263 (1988); M. Nei and G. Livshits, *Hum. Heredity* **39**, 276 (1989); in *Population Biology of Genes and Molecules*, N. Takahata and J. F. Crow, Eds. (Baifukan, Tokyo, 1990), pp. 251–265; S. Horai, K. Hayasaka, *Am. J. Hum. Genet.* **46**, 828 (1990).
9. R. L. Cann, M. Stoneking, A. C. Wilson, *Nature* **325**, 31 (1987); M. Stoneking and R. L. Cann, in *The Human Revolution*, P. Melbars and C. Stringer, Eds. (Edinburgh Univ. Press, Edinburgh, Scotland, 1989), pp. 17–30.
10. We thank L. Maxson, M. Nei, and R. Zauhar for the use of their facilities, A. Rzhetsky for assistance, and M. Nei for helpful comments. Supported by grants from the National Science Foundation (BSR 8918926 to L.M. and S.B.H. and BSR 9096248 to M.N.) and the National Institutes of Health (GM 20293-20 to M.N.).

26 November 1991; accepted 14 January 1992