Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1597–1601, March 1991
Evolution

Branching pattern in the evolutionary tree for human mitochondrial DNA

(polymerase chain reaction/direct sequencing/control region/distribution of pairwise differences/population growth)

ANNA DI RIENZO* AND ALLAN C. WILSON

Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Communicated by John Maynard Smith, October 29, 1990

ABSTRACT **Eighty-eight types of mitochondrial (mt) DNA were found by sequencing the most variable part of the control region from 117 Caucasians. In the tree relating those types, most of the branching events occur about two-thirds of the way from the root of the tree to the tips of the branches. Moreover, the distribution of sequence differences between all possible pairs of individuals is approximately Poisson. Other non-African populations show a similar pattern. Assuming a neutral model, these findings imply that the probability of survival of new lineages has undergone dramatic changes, probably due to population expansion. Conversely, African populations show multimodal distributions fitting with a model of constant population size.**

Owing to its fast evolution and maternal mode of inheritance, mtDNA can provide knowledge of genetic relationships among closely related individuals (1, 2). This approach has been used to suggest that all present-day mtDNA variation in humans can be traced back to a common ancestor who probably lived in an African population (3–5). We now show how the analysis of human mtDNA can give further insight into the evolutionary histories of populations. This advance has been made possible by the accumulation of sequences from the most variable part of mtDNA (4–7), by analyses suggesting that most of the variation in mtDNA is selectively neutral (8–10), and by the consequent development of theoretical methods for estimating such parameters as the size and growth rate of a population (refs. 11 and 12; M. Slatkin and R. Hudson, personal communication).

The present article applies one of these methods to the study of the demographic histories of human populations. We present sequences for 117 individuals, mainly from Sardinia and the Middle East.† These sequences allow us to analyze the branching pattern in the genealogical tree and the distribution of pairwise sequence differences between individuals. The Sardinian and Middle Eastern patterns are then compared to those obtained from other populations (refs. 4, 5, and 10; R. Ward, B. Frazer, K. Dew, and S. Pääbo, personal communication). Based on these patterns, theoretical expectations generated under the assumption of neutrality and different demographic conditions allow us to infer trends of population size through time for African and non-African populations.

MATERIALS AND METHODS

Population Samples. Sixty-nine placentas were collected in five maternity hospitals (Cagliari, Nuoro, Oristano, Ozieri, and Sassari) in Sardinia and the geographic origin of each sample was ascertained by the place of birth of the grandparents. The sampling strategy was designed to represent almost all linguistic areas except for one representing a recent foreign settlement (area 22). Sardinia was divided linguistically into 22 areas (14), which were grouped into five zones that are considered to be genetically homogeneous (15); zones A, B, C, D, and E include areas 1–2–3, 4–5–6–10, 7–8–9–11–12–13, 14–15–16, and 17–18–19–20–21, respectively. Numbers of individuals sampled were 12, 13, 15, 15, and 14 from zones A, B, C, D, and E, respectively.

DNA samples from the Middle East were kindly provided by J. Wainscoat and U. Ritte, and the donors consisted of 29 Bedouins from central and western Saudi Arabia (collected in Jeddah), 8 Israeli Arabs, and 5 Yemenite Jews (collected in Jerusalem, Bethlehem, and Nazareth). Six other Caucasian DNA samples [individuals 30, 61, 78, 94, 98, and 102 from the tree in Cann *et al.* (3)] were used only in the tree analysis.

Amplification and Sequencing of mtDNA. Total DNA was extracted from placental tissue as described elsewhere (16). Less than 100 ng of DNA was subjected to asymmetric amplification (17) of each strand in a 25-μl reaction volume with 2 units of *Thermus aquaticus* (*Taq*) DNA polymerase (Cetus); the temperature profile for 40 cycles of amplification was 92°C for 1 min, 56°C for 1 min, 72°C for 3 min. The primers were used in a molar ratio of 50:1 in each reaction; they were L15926 (5′-TCAAAGCTTACACCAGTCTTGTAAACC-3′) and H16498 (5′-CCTGAAGTAGGAACCAGATG-3′); L and H refer to the light and heavy strands, respectively, and the numbers refer to the base at the 3′ end of the primer in human mtDNA (18). This set of primers amplified an ≈620-base-pair (bp) segment that was purified by centrifugal dialysis in a Centricon-30 unit. Seven to 10 μl of the retentate was used for sequencing with the *Taq* polymerase and 2′-deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate by means of the Taquence kit (United States Biochemical), following the supplier's recommendations. The same primers were used for amplification and sequencing. Electrophoresis of reaction products through 6% polyacrylamide/7 M urea gels with wedge spacers (BRL) was carried out for 4 hr at 35 W. Gels were fixed, dried, and exposed to film for 18–72 hr.

Phylogenetic Analysis. Character-state trees that minimize the number of phylogenetically inferred substitutions were constructed by means of the computer algorithm of the PAUP package (19). At least 100 equally parsimonious trees were analyzed by computing the consensus tree and the frequency of each grouping; this allows us to identify the features common to all trees of the same length. Distance trees were built with the neighbor-joining computer program (20). The extent of substitution inferred to have occurred during the evolution of any two sequences is termed sequence divergence.

RESULTS

mtDNA Variation in Sardinia and the Middle East. The most variable part of human mtDNA is the control region, the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M58058–M58144).

Table 1. Variable sites in the control regions of 88 mtDNA types

MtDNA type	Place in tree	1111111111111111111111111111122222222222222222222222222222222233333333333 1244788900122333444455666666678899990001112233344456666777777888899900012333346 9846018136325089056937123456906946890171274634813756789012356168956712430246927 GACCTCCCTGTGCGTAACCTCACCCCCTCCTATACCCTACAAATCCCCCCCCACCCACCCTATATAAGCTTCGACTTTCG	Zone in Sardinia
1 Ref.	32		A, B, C, D
2	31	.....G.....AT.....	E
3	48	.....A.....G.....T.....A	E
4	13	.....C.....C.....	D
5	69	.....C.....T.....	E
6	43	.....A.....T.....T.....	E
7	47	.....T.....T.....	E
8	51	.....C.....	B, C, E
9	30	.....G.....	E
10	61	.....G.....C.....C.....C.....	E
11	41	.....T.....T.....C.....	E
12	40	....C....A.....	C
13	42	.....A.....T.....G.....T.....	C, E
14	29	..T.....T.....C.....	C
15	77	.....C.....G.....T.T..C.....T.....	D, E
16	55	.....C.....C.....	D
17	28	.....T.....	B, D, E
18	79	...T...GC.....	A
19	27	A.....	E
20	74	.....C.....C.....T.T.....	A, C
21	60	....C.....C.....C.....	D
22	2	.....A..T....T.C.....TGC.....T.G.....C...T.....	D
23	10	.....A.....C.....C.....	C
24	39	.....A.....C.....T.....	D
25	37	.....C.....	D
26	82	...T....C.....T.....T.....	E
27	14	.....C.....T....G.....C.....	B
28	26	.....T.....	C
29	5	.....A.....C.....T.....C.....C.....C...	D
30	59	.....A.....C...T.....C.....	B
31	25	.....A.....	C
32	81	...T....C.....T.....T.....C.....	B
33	73	.....C.....T.T.....	B, D
34	24	.G.....	B
35	58	.....C.....C.....	B
36	68	.....C....A.....T.....T....C.....	B
37	9	.....-..T..C.....C.....T.	B
38	12	.....C.....	A
39	36	.....T.T..C.....	A
40	34	....C.....	A
41	8	.....C.....C.....T.....	A
42	23	.....T.....	A
43	35	.....G.....T.....C.....	A
44	38	.....A.....	A
45	78	...T....C.....C.....	A
46	16	....C.....CT.....T.....T.....T.....A	A
47	11	.....C.....G...T.....	
48	88	...T....C.CA.....T.....T.....	
49	76	.....C.....G.....T..C.....T.....	
50	86	.....C..A.....T.C.....T.....T.....	
51	65	.....C.....T.....T.....	
52	80	...T...C.....T.....G.G.....	
53	7	.....T.....C.....C.....C.....	
54	54	.....T.....C.....G.....C.....C..	
55	1	.....A..T....T.C.....TGC.....T.G.....C.....	
56	71	....C...C....A.....T.T.....	
57	64	.....C.....C..	
58	62	.....C.....	
59	3	.....A.....C.....T.....C.....C.....	
60	63	.....C.....C.....C..	
61	67	.....C.....T.....T..C..	
62	6	.....A.....C.....C...G.....	
63	4	.....A.....C.....T.....CT.....C.....C...	
64	87	...T....C.CA.....T.....TT.....T.....	
65	46	.....C.....T.....T.....T.....G.....	
66	15	.....C.....T.....T.....	
67	45	.....T.....T.....A.....	
68	18	.....CT.....T.....T...T.....G.....A	
69	44	.....T.....C....C....	
70	33	....C.....T.G.....	
71	84	...T....C..A.....T.....	
72	83	...T....C..A.....T.....T.....	
73	50	...T.....C.....	
74	66	.....C.....G.....T..C..	
75	22	.....C.....T.....A.....	
76	85	...T....C..A.....T.....T.....	
77	17	.....CT.....T.....T.....T.....G.....A	
78	75	.....C.....G.....T..C.....T.....A	
79	70	.....C.....T.T.....	
80	21	.....T...T.....	
81	72	.....C.....C..C.....T.....T.T.....	
82	20	.....G.....	
83	19	.....G.....	
84	49	.....T.....C.....	
85	53	....C.....C.....C.....	
86	57	.....T.....C.....C.....	
87	56	.....C.....C.....C.....	
88	52	....C.....C.....C..	

Dots indicate matching with the reference sequence (mtDNA type *1*; ref. 18). The base number of each site in the control region (19, 28, etc., written vertically above the base) is assigned by aligning base 16024 in the reference sequence (18) to base 1. With the exception of types *3*, *7*, and *8*, which were shared by individuals from different populations (Table 2), types *1–46* were found in Sardinia, types *47–83* were found in the Middle East, and types *84–88* were found in the Caucasians from Cann *et al.* (3). Place in tree refers to the numerical position of each type in Fig. 1, position 1 being at the bottom of the tree. The zones in Sardinia where types *1–46* occur are described in the text.

Table 2. Geographic distribution of mtDNA types

mtDNA type	No. of individuals	
	Sardinians	Middle Easterners
1	15	—
3	1	1
7	1, 1*	—
8	3	1
13	3	—
15	2	—
17	3	—
20	2	—
33	2	—
48	—	2
54	—	2
66	—	2

The types are numbered as in Table 1. To test whether the low degree of type sharing between individuals from different populations is significantly different from random expectation, we calculated the probability of picking at random two individuals from the same population as the square of the relative size of the population sample and the probability of picking two individuals belonging to different populations as twice the product of the relative size of each population sample. These probabilities multiplied by the number of pairs of individuals with identical types were compared to the observations. The difference between the expected and observed number of pairs is highly significant ($P < 0.0001$, χ^2 test).
*Individual of Greek descent.

major noncoding region, which regulates transcription and replication. Variable sites are unevenly distributed in this 1122-bp region (6, 7); indeed, a segment spanning the first 400 bp contains most of the polymorphisms (64%). This segment was amplified and sequenced directly in 69 Sardinians, 42 Middle Easterners, and 6 other Caucasians from southern Europe. Eighty-eight different types were identified and differences among them are evident at 79 variable sites (Table 1).

Twelve of these types were shared by more than one individual (Table 2). The geographic pattern of sharing agrees with expectations based on the assumption that Mediterranean and Middle Eastern populations stem from an ancestral population that lived in prehistoric times and that the Sardinian population originated mainly by migration from the south European mainland 9000–6000 years ago (21, 22). Neither within Sardinia nor within the Middle East did we find any correlation between shared types and geographic locations; thus, these populations have no geographic structuring at our level of resolution.‡

The low extent of sharing of types between Sardinia and the Middle East attests to the high resolving power of the sequencing approach and is consistent with the hypothesis of their divergence from a common ancestral population in prehistorical times. During that period, mtDNA mutations have accumulated so that few types in Sardinia remain identical with their counterparts in the Middle East. The existence of two types that are shared between these geographic areas could obviously reflect later (i.e., historical) migrations (15, 22), but recourse to the latter explanation is not necessary.

Nonrandom Branching Pattern in the mtDNA Tree. A tree relating the 88 mtDNA types is shown in Fig. 1. This tree is accounted for by 148 events; they are all base substitutions except for a single base deletion. As would be expected of a tree

‡Our results differ in resolving power from those published on Sardinian nuclear genes, mtDNA restriction maps, surnames, and dialects (22–28), yet we see no conflict between the implications of the different data sets.

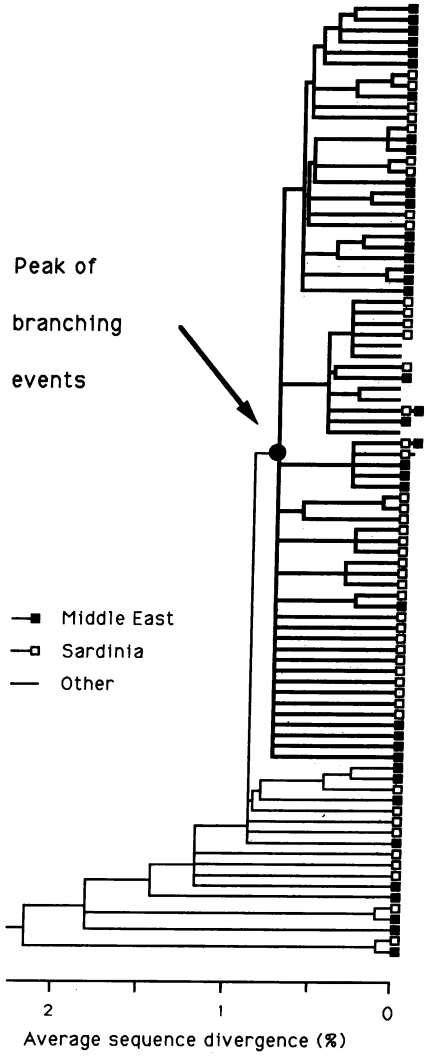

FIG. 1. Genealogical tree relating 88 Caucasian types of mtDNA. The boxes at the ends of the lineages refer to Sardinians and Middle Easterners; the other individuals are not labeled. The tree is rooted with the published sequences of common and pygmy chimpanzees (7, 29); the outgroup is not shown. The three twigs showing more than one symbol are the types shared by individuals from different geographic areas. A neighbor-joining tree rooted by the midpoint method exhibits a similar topology. Thick lines show the cluster of lineages with the highest incidence of branching. The transition-to-transversion ratio among the inferred substitutions is 20:1. The average consistency index excluding uninformative sites is 0.414, implying that each informative site has changed approximately 2.5 times, on average; the number of inferred substitutions per site ranges between 0 and 6, sites 70 and 166 showing the highest variability. There are four reasons for multiple changes at many sites: (*i*) the large number of sequences compared, (*ii*) the generally high evolutionary rate in this part of the control region, (*iii*) rate variation among sites, and (*iv*) the tremendous bias toward transitions, which are more likely than transversions to produce parallelisms and reversals (30).

for related populations, few clusters of lineages are specific to one population. Thus, the proportion of types from a given population having their nearest relative only in the same population is as low as 37%. In other words, each population stems from multiple lineages that existed in the ancestral population. Accordingly, the other Caucasians are also scattered in the tree. Moreover, the positions of these six Caucasian types with respect to one another in the restriction mapping tree (3) coincide with those in the present tree.

The most remarkable feature of this tree is the nonrandom pattern of branching; in fact, most branches in the tree

originate in a narrow interval of sequence divergence about two-thirds of the way from the root to the tips of the tree. Fig. 2 shows the frequency distribution of branch points through each interval of average sequence divergence. The peak at the 0.50–0.75 level of percent sequence divergence suggests that the probability of survival of new mtDNA lineages changed dramatically during the evolution of modern humans (11).

Frequency Distributions of Pairwise Sequence Differences. To check on this point, we looked at a distribution that does not depend on tree analysis, namely the frequency distribution of sequence differences between all possible pairs of individuals within each population. Both the Sardinian and the Middle Eastern populations exhibit distributions with a peak that is slightly to the left of center between the maximum and the minimum sequence difference (Fig. 3). Whereas the mode varies in a narrow range between 4 and 7 substitutions, the range of pairwise differences is from 0 to 15 on average. These distributions were compared to those calculated from sequences obtained by other workers from other populations. A striking similarity is evident among the distributions from non-African populations in that they are all unimodal and in reasonable agreement with a Poisson distribution, for which the mean is similar to the variance (Fig. 3). Orrego and King (13) observed such a distribution in another group of Caucasians.

African populations, however, show different patterns, with two or more modes, in significant disagreement with the Poisson distribution ($P < 0.0001$). In addition, the range of sequence differences is wider in Pygmies, in agreement with the finding of the deepest human lineages within Pygmy populations (10).

DISCUSSION

In principle, there are three ways of explaining the burst of branching in the Caucasian mtDNA tree (Fig. 1) and the approximately Poisson distribution of pairwise mtDNA differences in Caucasians and other non-African populations. The first two explanations assume that the differences among mtDNA variants are the result of neutral mutations (8–10). The third explanation invokes the action of positive selection on an advantageous mtDNA type.

The first explanation is that the non-African populations underwent a short period of fast growth and geographic expansion. When a population is constant in size each female is replaced on average by one female descendant in each generation; hence, the number of mtDNA lineages remains approximately constant through time. On the contrary, in a growing population each female can be replaced by more than one female and the number of mtDNA lineages increases; thus, the variation that is constantly produced has a higher probability of being retained in the population. In other words, the stochastic loss of lineages declines in a growing population and the frequency of new branch points (per unit time) in the tree increases (11). The observed peak of branch points as well as the unimodal distribution of pairwise sequence differences could reflect a sudden increase in the probability of survival of mtDNA lineages due to a burst of population growth. In light of the hypothesis that the common ancestry of all present-day mtDNAs was African (3, 4, 7, 10), this event could be associated with the exodus of an African propagule and its quick expansion into the rest of the world.

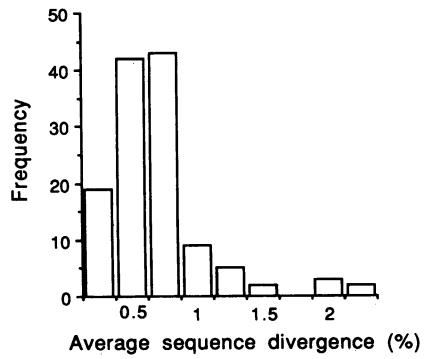

FIG. 2. Frequency distribution of branch points through each interval of average sequence divergence in the genealogical tree.

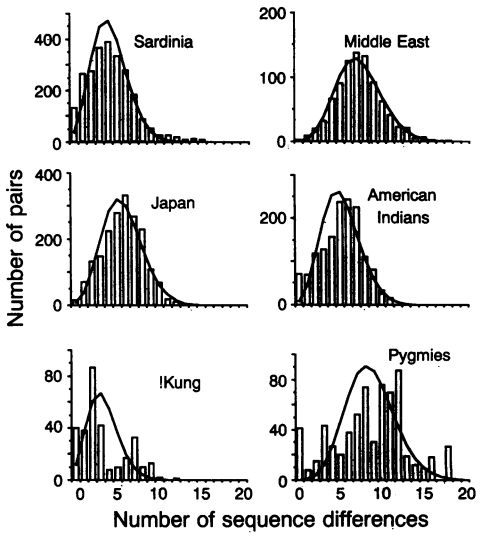

FIG. 3. Frequency distributions of sequence differences for all possible pairs of individuals in six populations. The lines show the expected Poisson distributions having the same mean as the observed distributions. The non-African population samples include 61 Japanese (5) and 52 American Indians (R. Ward, B. Frazer, K. Dew, and S. Pääbo, personal communication). The African population samples are 37 Pygmies (10), 26 !Kung (4, 10), 28 Herero (10), 17 Hadza (10), and 13 Yorubans (10). Distributions for the last three populations (not shown) also depart from a Poisson distribution ($P < 0.0001$).

The second explanation is that non-African populations have been growing exponentially at a constant slow rate throughout their history outside Africa. John Brookfield pointed out to us that this would result in a distribution of pairwise differences with a prominent peak. Brookfield's equation

$$P_t = 1 - e^{-(e^{rt}-1)/rN} \quad [1]$$

shows how P_t, the probability of observing two haploid individuals having an ancestor within the last t generations, depends on N, the population size now, and r, the population growth rate. A notable property of this equation is that there is a narrow interval of time in which most of the P values fall. P_t changes from almost 0 to almost 1 in a small range of t values around

$$t = r^{-1}(\ln rN). \quad [2]$$

For the r values that are thought to have applied to hunter-gatherer cultures during most of their history (31), namely $r \approx 0.005$ excess progeny per individual per generation, this t value is roughly 3000 generations—i.e., 60,000 years ago. Once the rate of sequence divergence for the control region is known accurately (cf. ref. 7), it will be possible to compare this t value to that estimated from the peak in the distribution of pairwise sequence differences (Fig. 3).

In addition, M. Slatkin and R. Hudson (personal communication) have shown by computer simulation that an approximately Poisson distribution of pairwise sequence differences is expected for growing populations. Moreover, the corresponding branching pattern predicted by the computer simulation is a star-like phylogeny where most branch points are clustered into a narrow range of sequence divergence. On the basis of this theoretical expectation, it seems impossible to distinguish between explanations 1 (an episode of fast growth) and 2 (a constant growth rate) for non-African populations.

By contrast, simulations of populations with constant size showed non-Poisson distributions, usually with more than one peak (M. Slatkin and R. Hudson, personal communication). This finding agrees, on the one hand, with the expectation based on Eq. 1 that when $r = 0$

$$P_t = 1 - e^{-t/N}. \quad [3]$$

Thus a geometric decline in frequency would be expected as the pairwise difference rises, as pointed out by Avise *et al.* (12). On the other hand, M. Slatkin and R. Hudson (personal communication) observe a geometric decline only when the results of many simulations are averaged. The geometric decline is usually not evident for a given population because the tree topology varies greatly from population to population and imposes large deviations from the geometric expectation. These findings make it seem unlikely that African populations have been growing, and we suggest that they are the result of a nearly constant population size.

Hence, there are two likely scenarios of demographic growth and geographic expansion of the non-African population as opposed to the ostensibly constant size of African populations. Both these scenarios agree with the model of an African origin for modern humans (3–5, 7) and a subsequent expansion to the rest of the world. The pattern observed in African populations could also represent the relics of an expansion event whose record has been erased through time by subsequent population reduction. Indeed, according to an African origin hypothesis, the pattern typical of population growth should have persisted longest and be most extreme in those populations that are more distant from the sub-Saharan source. It is therefore noteworthy that the number of differences for the peak decreases from 7 to 4 in the order (*i*) Middle Easterners, (*ii*) Japanese and American Indians, and (*iii*) Sardinians; the relative positions of the peak in these non-African populations are in agreement with their geographic distance from sub-Saharan Africa or with their time of colonization (21, 22).

The third explanation is that the non-African populations experienced strong selection on that mtDNA type which is marked by the solid circle in Fig. 1. Although this explanation cannot be ruled out, we mention two reasons for considering it unlikely. First, the lineage that underwent explosive branching in non-African populations did not branch explosively in Africa (10). Second, neutrality tests (9) imply that the control region variation observed in all non-African populations except Sardinians could be neutral. Moreover, previous comparisons of restriction maps indicated that most surviving mutations in the mtDNA genome of human populations are neutral (8).

It is important to devise ways of distinguishing among these three explanations. The Y chromosome may prove to be valuable in this connection. If the third explanation is correct, the Y tree should not exhibit peaks coinciding temporally with those in Figs. 2 and 3. Efforts are therefore needed to find large single-copy regions of the Y chromosome that have evolved quickly enough by base substitution (and without recombining with the X chromosome) to permit genealogical resolution on the mitochondrial time scale.

We thank A. Ambrosini, J. Brookfield, A. Esposito, B. Gigliotti, H. Harpending, T. Kocher, U. Lecca, W. Maddison, F. Manca, C. Meacham, A. Novelletto, S. Pääbo, E. Prager, U. Ritte, V. Sarich, A. Sidow, M. Slatkin, L. Terrenato, K. Thomas, L. Vigilant, J. Wainscoat, R. Ward, D. Weatherall, and especially D. Wilcox for discussion, samples, and assistance. This work received support from National Science Foundation and National Institutes of Health grants to A.C.W. and a European Molecular Biology Organization fellowship to A.D.

1. Avise, J. C. (1989) *Evolution* **43**, 1192–1208.
2. Wilson, A. C., Cann, R. L., Carr, S. M., George, M., Gyllensten, U. B., Helm-Bychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D. & Stoneking, M. (1985) *Biol. J. Linn. Soc.* **26**, 375–400.
3. Cann, R. L., Stoneking, M. & Wilson, A. C. (1987) *Nature (London)* **325**, 31–36.
4. Vigilant, L., Pennington, R., Harpending, H., Kocher, T. D. & Wilson, A. C. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9350–9354.
5. Horai, S. & Hayasaka, K. (1990) *Am. J. Hum. Genet.* **46**, 828–842.
6. Greenberg, B. D., Newbold, J. E. & Sugino, A. (1983) *Gene* **21**, 33–49.
7. Kocher, T. D. & Wilson, A. C. (1991) in *Evolution of Life*, eds. Osawa, S. & Honjo, T. (Springer, Tokyo), pp. 391–413.
8. Whittam, T. S., Clark, A. G., Stoneking, M., Cann, R. L. & Wilson, A. C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9611–9615.
9. Tajima, F. (1989) *Genetics* **123**, 585–595.
10. Vigilant, L. (1990) Ph.D. thesis (University of California, Berkeley).
11. Avise, J. C., Neigel, J. E. & Arnold, J. (1984) *J. Mol. Evol.* **20**, 99–105.
12. Avise, J. C., Ball, R. M. & Arnold, J. (1988) *Mol. Biol. Evol.* **5**, 331–344.
13. Orrego, C. & King, M. C. (1990) in *PCR Protocols: A Guide to Methods and Applications*, eds. Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (Academic, New York), pp. 416–426.
14. Contini, M. (1979) *Bull. Inst. Phonet. Grenoble* **8**, 57–96.
15. Piazza, A., Griffo, R., Cappello, N., Grassini, M., Olivetti, E., Rendine, S. & Zei, G. (1991) in *Language Change and Biological Evolution*, eds. Cavalli-Sforza, L. L., Piazza, A., Ramat, P. & Wang, W. (Stanford Univ. Press, Palo Alto, CA), in press.
16. Thomas, R. H., Schaffner, W., Wilson, A. C. & Pääbo, S. (1989) *Nature (London)* **340**, 465–467.
17. Gyllensten, U. B. & Erlich, H. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7652–7656.
18. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) *Nature (London)* **290**, 457–465.
19. Swofford, D. L. (1989) PAUP, Phylogenetic Analysis Using Parsimony (Illinois Natural History Survey, Champaign, IL), Version 3.0g.
20. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
21. Spoor, C. F. & Sondaar, P. Y. (1986) *J. Hum. Evol.* **15**, 399–408.
22. Modiano, G., Terrenato, L., Scozzari, R., Santachiara-Benerecetti, S. A., Ulizzi, L., Santolamazza, C., Petrucci, R. & Santolamazza, P. (1986) *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Nat. Rend.* Series 8, **18**, 257–330.
23. Piazza, A., van Loghem, E., de Lange, G., Curtoni, E. S., Ulizzi, L. & Terrenato, L. (1976) *Am. J. Hum. Genet.* **28**, 77–86.
24. Zei, G., Guglielmino Matessi, R., Siri, E., Moroni, A. & Cavalli-Sforza, L. L. (1983) *Ann. Hum. Genet.* **47**, 329–352.
25. Wijsman, E., Zei, G., Moroni, A. & Cavalli-Sforza, L. L. (1984) *Ann. Hum. Genet.* **48**, 65–78.
26. Brega, A., Scozzari, R., Maccioni, L., Iodice, C., Wallace, D. C., Bianco, I., Cao, A. & Santachiara-Benerecetti, A. S. (1986) *Ann. Hum. Genet.* **50**, 327–338.
27. Sartoris, S., Varetto, O., Migone, N., Cappello, N., Piazza, A., Ferrara, G. B. & Ceppellini, R. (1988) *Ann. Hum. Genet.* **52**, 327–340.
28. Santachiara-Benerecetti, A. S., Scozzari, R., Semino, O., Torroni, A., Brega, A. & Wallace, D. C. (1988) *Ann. Hum. Genet.* **52**, 39–56.
29. Foran, D. R., Hixson, J. E. & Brown, W. M. (1988) *Nucleic Acids Res.* **16**, 5841–5861.
30. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. (1982) *J. Mol. Evol.* **18**, 225–239.
31. Deevey, E. S. (1960) *Sci. Am.* **203** (3), 194–204.