

Mitochondrial DNA and human evolution

Rebecca L. Cann*, Mark Stoneking & Allan C. Wilson

Department of Biochemistry, University of California, Berkeley, California 94720, USA

Mitochondrial DNAs from 147 people, drawn from five geographic populations have been analysed by restriction mapping. All these mitochondrial DNAs stem from one woman who is postulated to have lived about 200,000 years ago, probably in Africa. All the populations examined except the African population have multiple origins, implying that each area was colonised repeatedly.

MOLECULAR biology is now a major source of quantitative and objective information about the evolutionary history of the human species. It has provided new insights into our genetic divergence from apes¹⁻⁸ and into the way in which humans are related to one another genetically⁹⁻¹⁴. Our picture of genetic evolution within the human species is clouded, however, because it is based mainly on comparisons of genes in the nucleus. Mutations accumulate slowly in nuclear genes. In addition, nuclear genes are inherited from both parents and mix in every generation. This mixing obscures the history of individuals and

allows recombination to occur. Recombination makes it hard to trace the history of particular segments of DNA unless tightly linked sites within them are considered.

Our world-wide survey of mitochondrial DNA (mtDNA) adds to knowledge of the history of the human gene pool in three ways. First, mtDNA gives a magnified view of the diversity present in the human gene pool, because mutations accumulate in this DNA several times faster than in the nucleus¹⁵. Second, because mtDNA is inherited maternally and does not recombine¹⁶, it is a tool for relating individuals to one another. Third, there are about 10^{16} mtDNA molecules within a typical human and they are usually identical to one another¹⁷⁻¹⁹. Typical mam-

* Present address: Department of Genetics, University of Hawaii, Honolulu, Hawaii 96822.

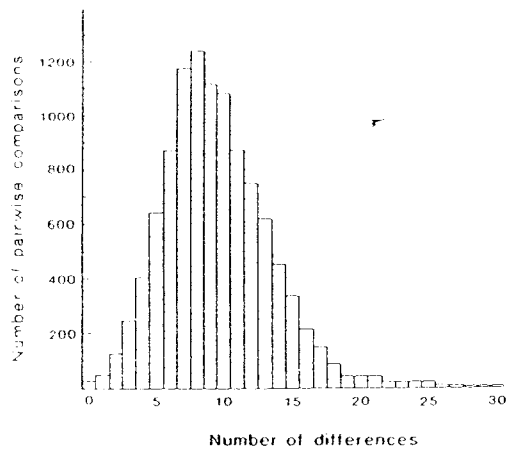


Fig. 1 Histogram showing the number of site differences between restriction maps of mtDNA for all possible pairs of 147 human beings.

malian females consequently behave as haploids, owing to a bottleneck in the genetically effective size of the population of mtDNA molecules within each oocyte²⁰. This maternal and haploid inheritance means that mtDNA is more sensitive than nuclear DNA to severe reductions in the number of individuals in a population of organisms¹⁵. A pair of breeding individuals can transmit only one type of mtDNA but carry four haploid sets of nuclear genes, all of which are transmissible to offspring. The fast evolution and peculiar mode of inheritance of mtDNA provide new perspectives on how, where and when the human gene pool arose and grew.

Restriction maps

MtDNA was highly purified from 145 placentas and two cell lines, HeLa and GM 3043, derived from a Black American and an aboriginal South African (IKung), respectively. Most placentas (98) were obtained from US hospitals, the remainder coming from Australia and New Guinea. In the sample, there were representatives of 5 geographic regions: 20 Africans (representing the sub-Saharan region), 34 Asians (originating from China, Vietnam, Laos, the Philippines, Indonesia and Tonga), 46 Caucasians (originating from Europe, North Africa, and the Middle East), 21 aboriginal Australians, and 26 aboriginal New Guineans. Only two of the 20 Africans in our sample, those bearing mtDNA types 1 and 81 (see below) were born in sub-Saharan Africa. The other 18 people in this sample are Black Americans, who bear many non-African nuclear genes probably contributed mainly by Caucasian males. Those males would not be expected to have introduced any mtDNA to the Black American population. Consistent with our view that most of

Table 1 MtDNA divergence within and between 5 human populations

| Population | % sequence divergence | | | | |
|----------------|-----------------------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 |
| 1. African | 0.47 | 0.04 | 0.04 | 0.05 | 0.06 |
| 2. Asian | 0.45 | 0.35 | 0.01 | 0.02 | 0.04 |
| 3. Australian | 0.40 | 0.31 | 0.25 | 0.03 | 0.04 |
| 4. Caucasian | 0.40 | 0.31 | 0.27 | 0.23 | 0.05 |
| 5. New Guinean | 0.42 | 0.34 | 0.29 | 0.29 | 0.25 |

The divergence is calculated by a published method²⁶. Values of the mean pairwise divergence between individuals within populations (δ_x) appear on the diagonal. Values below the diagonal (δ_{xy}) are the mean pairwise divergences between individuals belonging to two different populations, X and Y. Values above the diagonal (δ) are interpopulation divergences, corrected for variation within those populations with equation (1).

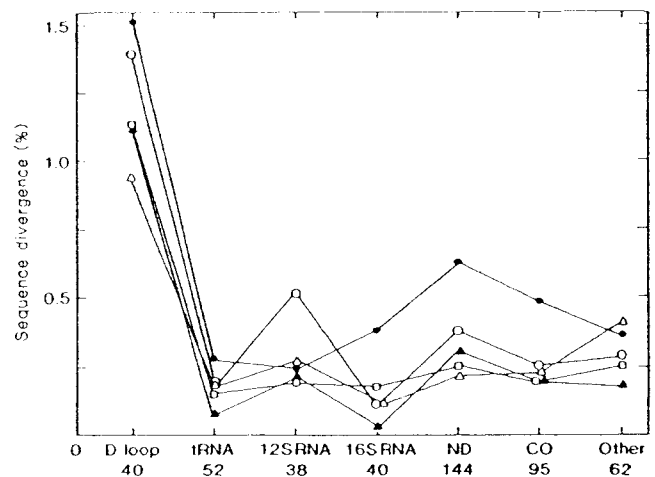


Fig. 2 Sequence divergence within 5 geographic areas for each of 7 functional regions in human mtDNA. Sequence divergence (δ_x) was estimated from comparisons of restriction maps²⁶. Symbols for the 5 races are: ●, Africa; ○, Asia; △, Australia; □, Caucasian; ▲, New Guinea. Along the horizontal axis are the numbers of restriction sites in each functional region (D loop, transfer RNA genes, 12S and 16S ribosomal RNA genes, NADH dehydrogenase subunits 1-5, cytochrome oxidase subunits and other protein-coding regions).

these 18 people are a reliable source of African mtDNA, we found that 12 of them bear restriction site markers known²¹ to occur exclusively or predominantly in native sub-Saharan Africans (but not in Europeans, Asians or American Indians nor, indeed, in all such Africans). The mtDNA types in these 12 people are 2-7, 37-41 and 82 (see below). Methods used to purify mtDNA and more detailed ethnographic information on the first four groups are as described^{17,22}; the New Guineans are mainly from the Eastern Highlands of Papua New Guinea²³.

Each purified mtDNA was subjected to high resolution mapping^{22,24} with 12 restriction enzymes (*HpaI*, *AvaiI*, *FnuDII*, *HhaI*, *HpaII*, *MboI*, *TaqI*, *RsaI*, *HinfI*, *HaeIII*, *AluI* and *DdeI*). Restriction sites were mapped by comparing observed fragment patterns to those expected from the known human mtDNA sequence²⁵. In this way, we identified 467 independent sites, of which 195 were polymorphic (that is, absent in at least one individual). An average of 370 restriction sites per individual were surveyed, representing about 9% of the 16,569 base-pair human mtDNA genome.

Map comparisons

The 147 mtDNAs mapped were divisible into 133 distinct types. Seven of these types were found in more than one individual.

Table 2 Clusters of mtDNA types that are specific to one geographic region

| Geographic region | Number of region-specific clusters | Mean pairwise distance within clusters* | Average age of clusters† |
|-------------------|------------------------------------|---|--------------------------|
| Africa | 14 | 0.36 | 90-180 |
| Asia | 27 | 0.21 | 53-105 |
| Australia | 15 | 0.17 | 43-85 |
| Europe | 36 | 0.09 | 23-45 |
| New Guinea | 7 | 0.11 | 28-55 |

* For clusters represented by two or more individuals (and calculated for individuals, not for mtDNA types) in Fig. 3.

† Average age in thousands of years based on the assumption that mtDNA divergence occurs at the rate of 2-4% per million years^{15,30}.

‡ Assuming that Africa is the source, there is only one African cluster.

no individual contained more than one type. None of the seven shared types occurred in more than one of the five geographic regions. One type, for example, was found in two Australians. Among Caucasians, another type occurred three times and two more types occurred twice. In New Guinea, two additional types were found three times and the seventh case involved a type found in six individuals.

A histogram showing the number of restriction site differences between pairs of individuals is given in Fig. 1; the average number of differences observed between any two humans is 9.5. The distribution is approximately normal, with an excess of pairwise comparisons involving large numbers of differences.

From the number of restriction site differences, we estimated the extent of nucleotide sequence divergence²⁶ for each pair of individuals. These estimates ranged from zero to 1.3 substitutions per 100 base pairs, with an average sequence divergence of 0.32%, which agrees with that of Brown¹⁷, who examined only 21 humans.

Table 1 gives three measures of sequence divergence within and between each of the five populations examined. These measures are related to one another by equation (1):

$$\delta = \delta_{xy} - 0.5(\delta_x + \delta_y) \quad (1)$$

where δ_x is the mean pairwise divergence (in percent) between individuals within a single population (X), δ_y is the corresponding value for another population (Y), δ_{xy} is the mean pairwise divergence between individuals belonging to two different populations (X and Y), and δ is a measure of the interpopulation divergence corrected for intrapopulation divergence. Africans as a group are more variable ($\delta_x = 0.47$) than other groups. Indeed, the variation within the African population is as great as that between Africans and any other group ($\delta_{xy} = 0.40-0.45$). The within-group variation of Asians ($\delta_x = 0.35$) is also comparable to that which exists between groups. For Australians, Caucasians, and New Guineans, who show nearly identical amounts of within-group variation ($\delta_x = 0.23-0.25$), the variation between groups slightly exceeds that within groups.

When the interpopulational distances (δ_{xy}) are corrected for intrapopulation variation (Table 1), they become very small ($\delta = 0.01-0.06$). The mean value of the corrected distance among populations ($\delta = 0.04$) is less than one-seventh of the mean distance between individuals within a population (0.30). Most of the mtDNA variation in the human species is therefore shared between populations. A more detailed analysis supports this²⁷.

Functional constraints

Figure 2 shows the sequence divergence (δ_x) calculated for each population across seven functionally distinct regions of the mtDNA genome. As has been found before^{24,27,28}, the most variable region is the displacement loop ($\delta_x = 1.3$), the major noncoding portion of the mtDNA molecule, and the least variable region is the 16S ribosomal RNA gene ($\delta_x = 0.2$). In general, Africans are the most diverse and Asians the next most, across all functional regions.

Evolutionary tree

A tree relating the 133 types of human mtDNA and the reference sequence (Fig. 3) was built by the parsimony method. To interpret this tree, we make two assumptions, both of which have extensive empirical support: (1) a strictly maternal mode of mtDNA transmission (so that any variant appearing in a group of lineages must be due to a mutation occurring in the ancestral lineage and not recombination between maternal and paternal genomes) and (2) each individual is homogeneous for its multiple mtDNA genomes. We can therefore view the tree as a genealogy linking maternal lineages in modern human populations to a common ancestral female (bearing mtDNA type a).

Many trees of minimal or near-minimal length can be made from the data; all trees that we have examined share the

following features with Fig. 3. (1) two primary branches, one composed entirely of Africans, the other including all 5 of the populations studied; and (2) each population stems from multiple lineages connected to the tree at widely dispersed positions. Since submission of this manuscript, Horai *et al.*²⁹ built a tree for our samples of African and Caucasian populations and their sample of a Japanese population by another method; their tree shares these two features.

Among the trees investigated was one consisting of five primary branches with each branch leading exclusively to one of the five populations. This tree, which we call the population-specific tree, requires 51 more point mutations than does the tree of minimum length in Fig. 3. The minimum-length tree requires fewer changes at 22 of the 93 phylogenetically-informative restriction sites than does the population-specific tree, while the latter tree required fewer changes at four sites; both trees require the same number of changes at the remaining 67 sites. The minimum-length tree is thus favoured by a score of 22 to 4. The hypothesis that the two trees are equally compatible with the data is statistically rejected, since 22:4 is significantly different from the expected 13:13. The minimum-length tree is thus significantly more parsimonious than the population-specific tree.

African origin

We infer from the tree of minimum length (Fig. 3) that Africa is a likely source of the human mitochondrial gene pool. This inference comes from the observation that one of the two primary branches leads exclusively to African mtDNAs (types 1-7, Fig. 3) while the second primary branch also leads to African mtDNAs (types 37-41, 45, 46, 70, 72, 81, 82, 111 and 113). By postulating that the common ancestral mtDNA (type a in Fig. 3) was African, we minimize the number of intercontinental migrations needed to account for the geographic distribution of mtDNA types. It follows that b is a likely common ancestor of all non-African and many African mtDNAs (types 8-134 in Fig. 3).

Multiple lineages per race

The second implication of the tree (Fig. 3)—that each non-African population has multiple origins—can be illustrated most simply with the New Guineans. Take, as an example, mtDNA type 49, a lineage whose nearest relative is not in New Guinea, but in Asia (type 50). Asian lineage 50 is closer genealogically to this New Guinea lineage than to other Asian mtDNA lineages. Six other lineages lead exclusively to New Guinean mtDNAs, each originating at a different place in the tree (types 12, 13, 26-29, 65, 95 and 127-134 in Fig. 3). This small region of New Guinea (mainly the Eastern Highlands Province) thus seems to have been colonised by at least seven maternal lineages (Tables 2 and 3).

In the same way, we calculate the minimum numbers of female lineages that colonised Australia, Asia and Europe (Tables 2 and 3). Each estimate is based on the number of region-specific clusters in the tree (Fig. 3, Tables 2 and 3). These numbers, ranging from 15 to 36 (Tables 2 and 3), will probably rise as more types of human mtDNA are discovered.

Tentative time scale

A time scale can be affixed to the tree in Fig. 3 by assuming that mtDNA sequence divergence accumulates at a constant rate in humans. One way of estimating this rate is to consider the extent of differentiation within clusters specific to New Guinea (Table 2; see also refs 23 and 30), Australia³⁰ and the New World³¹. People colonised these regions relatively recently: a minimum of 30,000 years ago for New Guinea³², 40,000 years ago for Australia³³, and 12,000 years ago for the New World³⁴. These times enable us to calculate that the mean rate of mtDNA divergence within humans lies between two and four percent per million years; a detailed account of this calculation appears

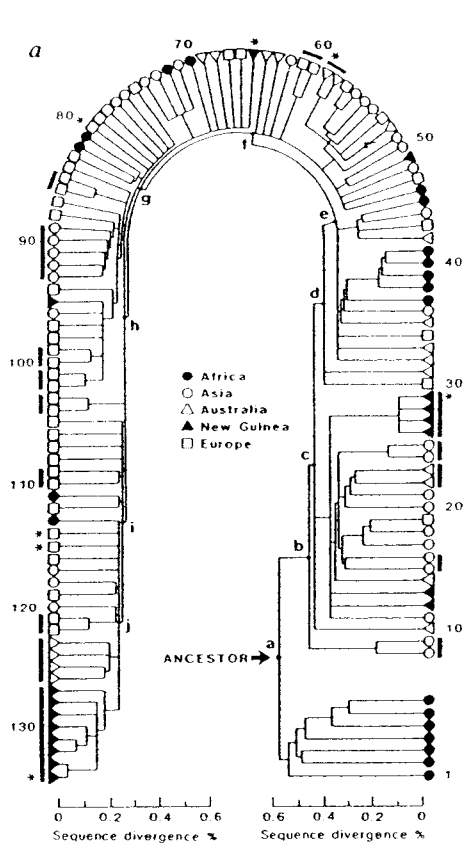


Fig. 3 *a*, Genealogical tree for 134 types of human mtDNA (133 restriction maps plus reference sequence); *b*, comprehensive list of polymorphic restriction sites used. The tree accounts for the site differences observed between restriction maps of these mtDNAs with 398 mutations. No other order of branching tested is more parsimonious than this one. This order of branching was obtained (using the computer program PAUP, designed by Dr David Swofford) by ignoring every site present in only one type of mtDNA or absent in only one type and confining attention to the remaining 93 polymorphic sites. The computer program produces an unrooted network, which we converted to a tree by placing the root (arrow) at the midpoint of the longest path connecting the two lineages (see ref. 58). The numbers refer to mtDNA types, no. 1 being from the aboriginal South African (!Kung) cell line (GM 3043), no. 45 being from the HeLa cell line and no. 110 being the published human sequence²⁵. Black bars, clusters of mtDNA types specific to a given geographic region; asterisks, mtDNA types found in more than 1 individual: type 134 was in six individuals, types 29, 65 and 80 each occurred thrice, and other types flagged with asterisks occurred twice. To place the nodes in the tree relative to the percent divergence scale, we took account of the differences observed at all 195 polymorphic sites. *b*, The numbering of sites is according to the published human sequence²⁵, with 12 restriction enzymes indicated by the following single letter code: a, *AluI*; b, *Avall*; c, *DdeI*; d, *FnuDII*; e, *HaeIII*; f, *HhaI*; g, *HinfI*; h, *HpaI*; i, *HpaII*; j, *MboI*; k, *RsaI*; l, *TaqI*. Italicized sites are present in the indicated mtDNA types and nonitalicized sites are absent in the indicated types; parentheses refer to alternative placements of inferred sites; sites separated by a slash are polymorphic for two different restriction enzymes caused by a single inferred nucleotide substitution; letters in bold face refer to noncoding regions and genes for transfer RNA, ribosomal RNA and proteins (ND, NADH dehydrogenase; CO, cytochrome oxidase; ATP, adenosine triphosphatase; CYT b, cytochrome b). For example, 8j indicates a *MboI* site beginning at nucleotide position 8 in the D loop that was not found in mtDNA type 1 but was present in type 2, etc. Note that since this site is not present in the reference sequence (type 110), the sequence beginning at position 8 is actually a semisite, differing from the *MboI* recognition sequence at one position (see ref. 23 for a detailed description of the method of mapping such inferred sites). Not all sites were scored in all individuals: 8j, 1484e and 7750c were not determined for types 1, 31, 59, 63, 68, and all of the New Guinea mtDNAs; mtDNAs 114 and 121 could not be typed with *RsaI*. The locations of some sites differ from those reported before^{23,24}, as do the individuals in which some sites occur; these revisions are based on re-examination of previously-studied mtDNAs.

elsewhere³⁰. This rate is similar to previous estimates from animals as disparate as apes, monkeys, horses, rhinoceroses, mice, rats, birds and fishes¹⁵. We therefore consider the above estimate of 2%-4% to be reasonable for humans, although additional comparative work is needed to obtain a more exact calibration.

As Fig. 3 shows, the common ancestral mtDNA (type a) links mtDNA types that have diverged by an average of nearly 0.57%. Assuming a rate of 2%-4% per million years, this implies that the common ancestor of all surviving mtDNA types existed 140,000-290,000 years ago. Similarly, ancestral types b-j may have existed 62,000-225,000 years ago (Table 3).

When did the migrations from Africa take place? The oldest of the clusters of mtDNA types to contain no African members stems from ancestor c and included types 11-29 (Fig. 3). The

apparent age of this cluster (calculated in Table 3) is 90,000-180,000 years. Its founders may have left Africa at about that time. However, it is equally possible that the exodus occurred as recently as 23-105 thousand years ago (Table 2). The mtDNA results cannot tell us exactly when these migrations took place.

Other mtDNA studies

Two previous studies of human mtDNA have included African individuals^{21,28}; both support an African origin for the human mtDNA gene pool. Johnson *et al.*²¹ surveyed ~40 restriction sites in each of 200 mtDNAs from Africa, Asia, Europe and the New World, and found 35 mtDNA types. This much smaller number of mtDNA types probably reflects the inability of their methods to distinguish between mtDNAs that differ by less than 0.3% and may account for the greater clustering of mtDNA

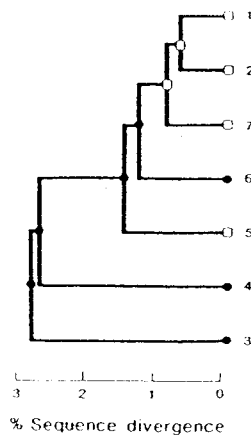


Fig. 4 Genealogical tree relating the nucleotide sequences of D loops from seven human mtDNAs. This tree, which requires fewer mutations than any other branching order, was constructed by the PAUP computer program from the 900-bp sequences determined by Greenberg *et al.*²⁸ and was rooted by the midpoint method. Symbols: ●, African origin (Black American); □, Caucasian.

types by geographic origin that they observed. (By contrast, our methods distinguish between mtDNAs that differ by 0.03%.) Although Johnson *et al.* favoured an Asian origin, they too found that Africans possess the greatest amount of mtDNA variability and that a midpoint rooting of their tree leads to an African origin.

Greenberg *et al.*²⁸ sequenced the large noncoding region, which includes the displacement loop (D loop), from four Caucasians and three Black Americans. A parsimony tree for these seven D loop sequences, rooted by the midpoint method, appears in Fig. 4. This tree indicates (1) a high evolutionary rate for the D loop (at least five times faster than other mtDNA regions), (2) a greater diversity among Black American D loop sequences, and (3) that the common ancestor was African.

Nuclear DNA studies

Estimates of genetic distance based on comparative studies of nuclear genes and their products differ in kind from mtDNA estimates. The latter are based on the actual number of mutational differences between mtDNA genomes, while the former rely on differences in the frequencies of molecular variants measured between and within populations. Gene frequencies can be influenced by recombination, genetic drift, selection, and migration, so the direct relationship found between time and mutational distance for mtDNA would not be expected for genetic distances based on nuclear DNA. But studies based on polymorphic blood groups, red cell enzymes, and serum proteins show that (1) differences between racial groups are smaller than those within such groups and (2) the largest gene frequency

differences are between Africans and other populations, suggesting an African origin for the human nuclear gene pool^{11,12,35}. More recent studies of restriction site polymorphisms in nuclear DNA^{14,36-42} support these conclusions.

Relation to fossil record

Our tentative interpretation of the tree (Fig. 3) and the associated time scale (Table 3) fits with one view of the fossil record: that the transformation of archaic to anatomically modern forms of *Homo sapiens* occurred first in Africa⁴³⁻⁴⁵, about 100,000-140,000 years ago, and that all present-day humans are descendants of that African population. Archaeologists have observed that blades were in common use in Africa 80-90 thousand years ago, long before they replaced flake tools in Asia or Europe^{46,47}. But the agreement between our molecular view and the evidence from palaeoanthropology and archaeology should be treated cautiously for two reasons. First, there is much uncertainty about the ages of these remains. Second, our placement of the common ancestor of all human mtDNA diversity in Africa 140,000-280,000 years ago need not imply that the transformation to anatomically modern *Homo sapiens* occurred in Africa at this time. The mtDNA data tell us nothing of the contributions to this transformation by the genetic and cultural traits of males and females whose mtDNA became extinct.

An alternative view of human evolution rests on evidence that *Homo* has been present in Asia as well as in Africa for at least one million years⁴⁸ and holds that the transformation of archaic to anatomically modern humans occurred in parallel in different parts of the Old World^{33,49}. This hypothesis leads us to expect genetic differences of great antiquity within widely separated parts of the modern pool of mtDNAs. It is hard to reconcile the mtDNA results with this hypothesis. The greatest divergences within clusters specific to non-African parts of the World correspond to times of only 90,000-180,000 years. This might imply that the early Asian *Homo* (such as Java man and Peking man) contributed no surviving mtDNA lineages to the gene pool of our species. Consistent with this implication are features, found recently in the skeletons of the ancient Asian forms, that make it unlikely that Asian *erectus* was ancestral to *Homo sapiens*⁵⁰⁻⁵². Perhaps the non-African *erectus* population was replaced by *sapiens* migrants from Africa; incomplete fossils indicating the possible presence of early modern humans in western Asia at Zuttiyeh (75,000-150,000 years ago) and Qafzeh (50,000-70,000 years ago) might reflect these first migrations^{45,53}.

If there was hybridization between the resident archaic forms in Asia and anatomically modern forms emerging from Africa, we should expect to find extremely divergent types of mtDNA in present-day Asians, more divergent than any mtDNA found in Africa. There is no evidence for these types of mtDNA among the Asians studied^{21,54-56}. Although such archaic types of mtDNA could have been lost from the hybridizing population, the probability of mtDNA lineages becoming extinct in an expanding population is low⁵⁷. Thus we propose that *Homo*

Table 3 Ancestors, lineages and extents of divergence in the genealogical tree for 134 types of human mtDNA

| Ancestor | Total | No. of descendant lineages or clusters specific to a region | | | | | % divergence | Age* |
|----------|-------|---|------|-----------|--------|-----------|--------------|---------|
| | | Africa | Asia | Australia | Europe | N. Guinea | | |
| a | 7 | 1 | 0 | 0 | 0 | 0 | 0.57 | 143-285 |
| b | 2 | 0 | 1 | 0 | 0 | 0 | 0.45 | 112-225 |
| c | 20 | 0 | 7 | 3 | 1 | 3 | 0.43 | 108-215 |
| d | 2 | 0 | 0 | 1 | 1 | 0 | 0.39 | 98-195 |
| e | 14 | 2 | 2 | 4 | 2 | 0 | 0.34 | 85-170 |
| f | 19 | 1 | 7 | 4 | 4 | 1 | 0.30 | 75-150 |
| g | 10 | 2 | 3 | 2 | 2 | 1 | 0.28 | 70-140 |
| h | 30 | 2 | 4 | 0 | 15 | 1 | 0.27 | 68-135 |
| i | 8 | 1 | 0 | 0 | 6 | 0 | 0.26 | 65-130 |
| j | 22 | 1 | 3 | 1 | 5 | 1 | 0.25 | 62-125 |
| All | 134 | 10 | 27 | 15 | 36 | 7 | — | — |

* Assuming that the mtDNA divergence rate is 2-4% per million years^{15,30}.

