PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

New Approaches to Dating Suggest a Recent Age for the Human mtDNA Ancestor

Mark Stoneking, Stephen T. Sherry, Alan J. Redd and Linda Vigilant

Phil. Trans. R. Soc. Lond. B 1992 **337**, 167-175
doi: 10.1098/rstb.1992.0094

New approaches to dating suggest a recent age for the human mtDNA ancestor

MARK STONEKING, STEPHEN T. SHERRY, ALAN J. REDD AND LINDA VIGILANT

Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A.

SUMMARY

The most critical and controversial feature of the African origin hypothesis of human mitochondrial DNA (mtDNA) evolution is the relatively recent age of about 200 ka inferred for the human mtDNA ancestor. If this age is wrong, and the actual age instead approaches 1 million years ago, then the controversy abates. Reliable estimates of the age of the human mtDNA ancestor and the associated standard error are therefore crucial. However, more recent estimates of the age of the human ancestor rely on comparisons between human and chimpanzee mtDNAs that may not be reliable and for which standard errors are difficult to calculate. We present here two approaches for deriving an intraspecific calibration of the rate of human mtDNA sequence evolution that allow standard errors to be readily calculated. The estimates resulting from these two approaches for the age of the human mtDNA ancestor (and approximate 95% confidence intervals) are 133 (63–356) and 137 (63–416) ka ago. These results provide the strongest evidence yet for a relatively recent origin of the human mtDNA ancestor.

1. INTRODUCTION

The idea that all mitochondrial DNA (mtDNA) variation in modern populations traces back to a single ancestor who lived in Africa some 200 ka ago (Cann *et al.* 1987) has stimulated much interest, research, and debate concerning the origins of modern humans. This African origin hypothesis was reinforced by a recent study (Vigilant *et al.* 1991) that addressed many of the perceived weaknesses of the original study by Cann *et al.* (1987)†. Controversy focuses not so much on the geographic origin of the human mtDNA ancestor, but rather on when she lived: if the analyses suggesting a 200 ka date are correct, then the implication is that non-African populations older than 200 ka could not have contributed mtDNA types to modern populations. The non-African populations would have been replaced, without much (if any) interbreeding, by anatomically-modern populations spreading from Africa in the past 100 ka (Cann *et al.* 1987; Stoneking & Cann 1989).

If, however, the age of the human ancestor has been estimated incorrectly and is in fact nearer to 1 Ma ago, then the mtDNA results would instead be indicating the first spread of hominids (presumably *Homo erectus*) out of Africa some 1 Ma ago. The replacement controversy then disappears (for mtDNA), as few anthropologists disagree with the contention that these were in fact the first hominids to migrate from Africa, and that therefore there were no prior populations to replace.

The critical aspect of the African origin hypothesis is clearly the age of the human mtDNA ancestor, and most studies have produced estimates which centre around 200 ka ago (Cann *et al.* 1987; Stoneking & Cann 1989; Hasegawa & Horai 1991; Kocher & Wilson 1991; Vigilant *et al.* 1991). There are two parameters which need to be estimated to arrive at an age for the human mtDNA ancestor: (i) the amount of sequence evolution that has occurred since the ancestor lived; and (ii) the rate of human mtDNA sequence evolution. Unfortunately, variances of these parameters are usually unknown, making it difficult to ascertain the reliability of previous age estimates‡.

In addition, the rate of human mtDNA sequence evolution is usually determined by comparing human mtDNA to chimpanzee mtDNA. For this interspecific comparison to be accurate, human and chimpanzee mtDNA must evolve at the same rate. However, patterns of nucleotide substitution appear to differ between human and chimpanzee mtDNA (Hasegawa & Horai 1991; Kocher & Wilson 1991), which implies that the rates of substitution might differ as well.

† The statistical tests used by Vigilant *et al.* (1991) to buttress the support for an African origin of the human mtDNA ancestor have since been shown to be invalid, due to the inadequacy of parsimony analysis for these data (Hedges *et al.* 1992; Maddison *et al.* 1992; Templeton 1992). Nevertheless, although not statistically proven, an African origin remains as the best explanation for all of the human mtDNA data (Stoneking & Cann 1989; Horai & Hayasaka 1990; Hasegawa & Horai 1991; Kocher & Wilson 1991; Merriweather *et al.* 1991; Vigilant *et al.* 1991; Stoneking *et al.* 1992*a*).

‡ Although estimates of the age of the human mtDNA ancestor are often expressed as ranges, these are not 95% confidence intervals, but rather reflect the range of possible values for the parameters used to determine the age.

Phil. Trans. R. Soc. Lond. B (1992) **337,** 167–175
Printed in Great Britain

Another complication is that for the most rapidly evolving regions of the mtDNA genome (the non-coding control region), so many multiple substitutions have occurred that it is difficult to obtain an accurate measure of the true amount of sequence evolution between human and chimpanzee mtDNAs (Hasegawa & Horai 1991; Vigilant *et al.* 1991). Finally, the time of separation of human and chimpanzee mtDNAs must be stipulated, possibly introducing a further source of error into the rate calibration.

We present here two approaches to an intraspecific calibration of the rate of human mtDNA sequence evolution. Intraspecific calibration (i.e. relying solely on mtDNA sequence evolution within human populations) avoids the aforementioned shortcomings of chimpanzee–human comparisons, providing independent estimates of the age of the human mtDNA ancestor. The two approaches to intraspecific calibration also produce standard errors of the parameters, enabling construction of approximate 95% confidence intervals for the age estimates. The results from these two approaches are nearly identical, and provide some of the strongest evidence yet for a relatively recent age of human mtDNA ancestor.

2. RATIONALE

The rationale behind the intraspecific calibration of the rate of human mtDNA evolution was described in detail by Stoneking *et al.* (1986), and applied previously to restriction map variation (Stoneking *et al.* 1986; Stoneking & Cann 1989). In brief, the procedure relies on the identification of monophyletic clusters or 'groups' of mtDNA types specific to a defined geographic region of the world. Ideally, this region of the world should have been colonized once at a definite time that is firmly established from archaeological or biogeographical evidence, with little or no back-migration. Although no human population will exactly meet these criteria, the colonization of Papua New Guinea (PNG) comes reasonably close (Stoneking *et al.* 1986).

The first approach to an intraspecific calibration is that developed by Stoneking *et al.* (1986). This approach is based on the diversity within groups of PNG-specific mtDNA types, and hence is termed the 'within-group' approach. According to this model (figure 1), the deepest branchpoint within each group of PNG mtDNA types represents an estimate of the amount of sequence evolution that has occurred exclusively within PNG, after colonization. If we know the colonization time for PNG, then the amount of sequence evolution to the origin of the PNG group, divided by the colonization time, provides an estimate of the rate of human mtDNA sequence evolution. Our modification of the previous development of this approach is to use a method of tree analysis that also produces standard errors of each branchpoint (Nei *et al.* 1985). This enables construction of 95% confidence intervals for the crucial parameters: namely, the rate of mtDNA sequence evolution, the amount of sequence evolution since the human mtDNA ancestor, and the age of the human mtDNA ancestor.

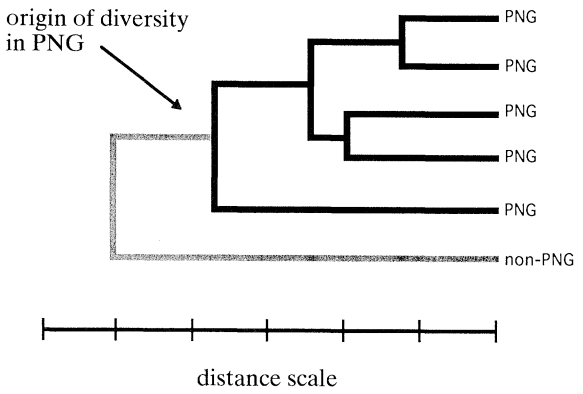

Figure 1. Rationale for using the divergence within monophyletic groups of PNG mtDNA types to estimate the rate of CR sequence evolution. A hypothetical group of PNG types is shown (black lines) with a nearest neighbour that is not from PNG (shaded lines). The arrow points to the branchpoint at which this group began diversifying within PNG; the amount of sequence evolution from this point (determined from the distance scale) can, together with the colonization time, be used to estimate the rate of CR sequence evolution.

The second approach to an intraspecific calibration is also based on the amount of sequence evolution within each group of PNG mtDNA types, but in addition it utilizes the divergence between different groups. Hence, we call this the 'between-group' approach. Figure 2 illustrates the rationale behind the between-group approach for a hypothetical tree of two PNG groups. These two PNG groups have a common mtDNA ancestor that lived at some unknown time prior to the colonization of PNG. At the colonization time, ancestors of these two groups migrated to PNG and the process of within-group divergence began. The observed mean intergroup divergence (d_{XY}) for these two groups is therefore the result of two processes: divergence prior to colonization, and divergence after colonization. The divergence that occurred after colonization (d_A) can be estimated by the following formula (Nei 1987, p. 276):

$$d_A = d_{XY} - (d_X + d_Y)/2 \qquad (1)$$

where d_X and d_Y are the observed mean sequence divergence within groups X and Y. If the rate of sequence evolution is constant, then d_A is equal to $2\lambda T$, where T is the time of divergence and λ is the rate of sequence evolution. Thus, by knowing the colonization time and the observed sequence divergence within and between each group, we can estimate the rate of mtDNA sequence evolution. Furthermore, standard errors of the mean divergence within and between groups can be obtained by the method of Nei & Jin (1989) and used to construct 95% confidence intervals.

These two approaches to an intraspecific calibration have several advantages over the more traditional rate estimates derived from interspecific comparisons: the rate of mtDNA evolution is not assumed to be constant across different species (but it is assumed to be constant across different human populations);

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

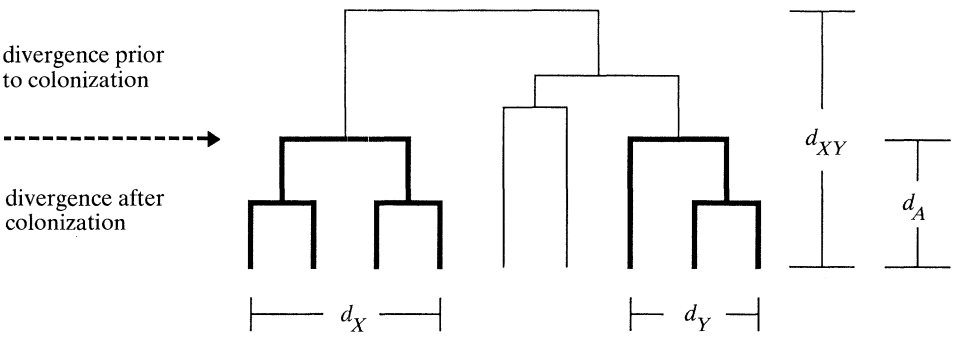

Figure 2. Hypothetical tree of two PNG groups of mtDNA types, illustrating how divergence since the colonization of PNG (d_A) can be determined from the observed mean within-group (d_X and d_Y) and between-group (d_{XY}) divergences.

species divergence times are not required (but colonization times must be known); and the period of evolutionary time is so short that the correction for multiple substitutions at the same nucleotide position is not important.

In this paper we detail the application of these two approaches to intraspecific calibration to an extensive dataset that includes many PNG mtDNA sequences. We derive estimates of the rate of mtDNA sequence evolution, the age of the human mtDNA ancestor, and associated 95% confidence intervals. After mentioning some caveats to this procedure, we compare our findings to previous estimates and discuss the resulting implications for the origin of modern humans.

3. MATERIALS AND METHODS

Sequences of the two hypervariable segments of the mtDNA control region (CR) were obtained for 30 individuals from PNG and 20 individuals from Indonesia. Additional sequences (including 20 from PNG) were from Vigilant *et al.* (1991). The 50 PNG mtDNAs come from various highland and coastal localities and were pre-selected from 119 PNG mtDNAs to be representative of the mtDNA diversity in this sample that had previously been characterized by high-resolution mapping with restriction enzymes (Stoneking *et al.* 1990). The Indonesian samples come from the Moluccas and the Lesser Sunda islands, and were selected from 148 samples that were previously screened for a deletion of one of two copies of a nine base-pair (b.p.) repeated sequence located between the COII and lysine tRNA genes (Stoneking *et al.* 1992*b*). All of the Indonesian mtDNAs chosen for sequencing have the deletion.

Amplification of CR hypervariable segments by the polymerase chain reaction (PCR), preparation of single-stranded templates by asymmetric PCR (Gyllensten & Erlich 1988), and dideoxy sequencing were done essentially as described previously (Vigilant *et al.* 1989). Estimates of sequence divergence and associated standard errors (Nei & Jin 1989) were corrected for multiple substitutions by the Jukes–Cantor method (Jukes & Cantor (1969); Kimura's two-parameter method (Kimura 1980) gave essentially identical results). These divergence values, as well as neighbour-joining (NJ) trees (Saitou & Nei 1987), and unweighted pair group method of averaging (UPGMA) trees with associated standard errors of the branch points (Nei *et al.* 1985), were computed with programs kindly provided by M. Nei.

4. RESULTS

(*a*) *Sequences*

The CR hypervariable segments span 764 nucleotides; in actuality, an average of 706 nucleotides of DNA sequence was determined from each individual. When combined with the data of Vigilant *et al.* (1991), there were 205 polymorphic nucleotide sites. For the 239 individuals comprising the combined dataset, there were 165 different mtDNA types, including 41 mtDNA types among the 50 PNG sequences and 16 mtDNA types among the 20 Indonesian sequences. Only one type occurred in more than one population, and that type was shared by six individuals from PNG, one Indonesian, and one Philippino.

(*b*) *Comparing sequences to restriction maps*

As a means of assessing the information obtained from the CR sequences, we compared the restriction maps that would be inferred from 50 PNG sequences to the restriction maps that were actually determined for these individuals (Stoneking *et al.* 1990). One discrepancy was noted that could be explained by assuming that the restriction site polymorphism in question did not actually map to the sequenced portion of the control region. In every other instance complete concordance was noted between the restriction maps inferred from the CR sequences and the previously determined restriction maps.

We also compared estimates of mean sequence divergence from CR sequences and from restriction maps for each of the 1225 pairs of PNG individuals (figure 3). A clear linear trend is evident, with a moderate correlation coefficient between the two measures of divergence ($r = 0.57$). The slope of the regression line (not shown) indicates that CR sequence divergence values are on average about 13 times bigger than restriction map divergence values (i.e. the CR hypervariable segments are evolving some 13 times

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

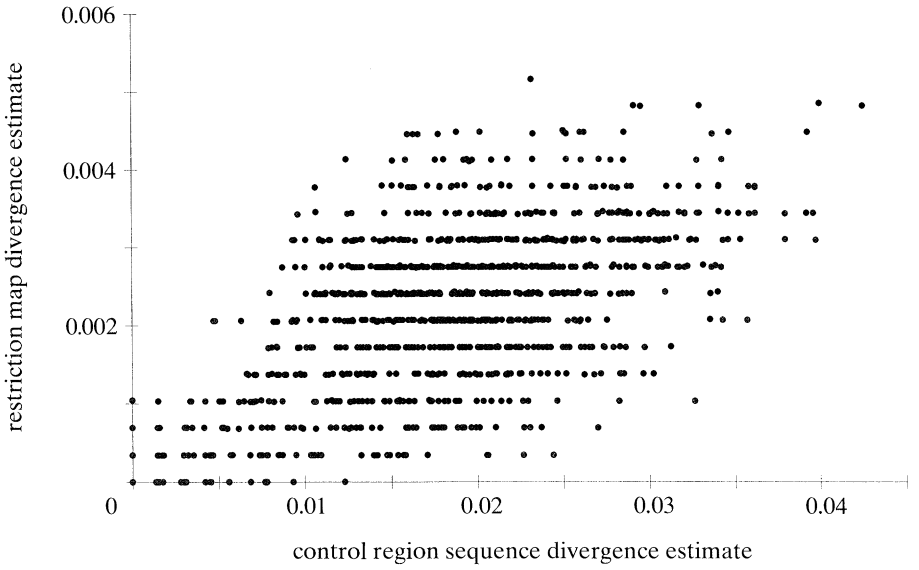

Figure 3. Relationship between the estimates of pairwise sequence divergence based on restriction maps and those based on CR sequences for the 50 PNG mtDNAs. Each point indicates the Jukes–Cantor divergence for one pair of individuals. The horizontal 'stripe' effect arises because the divergences based on restriction maps can only take on certain discrete values (i.e. one restriction site difference = 0.034% divergence, two site differences = 0.068% divergence, etc.).

faster than the average rate for the entire human mtDNA genome).

Other measures of diversity also indicate that more variation is detected with CR sequences than with high-resolution restriction maps. Among the 50 PNG individuals, CR sequences identified 41 different mtDNA types whereas restriction maps identified only 35 different types. The average probability of identity (that is, the probability of two individuals selected at random having identical mtDNA types) is 3.8% for CR sequences and 4.2% for restriction maps. For individuals with identical CR sequences, the probability that they also have identical restriction maps is 53.8%; for individuals with identical restriction maps, the probability that they also have identical CR sequences is 45.8%.

(*c*) *Phylogenetic analysis*

An abbreviated UPGMA tree relating the 165 CR sequences and rooted with a chimpanzee CR sequence (Foran *et al.* 1988) is shown in figure 4. Of the 41 PNG types, 35 (85%) fall into one of three groups. An NJ tree for these data was also constructed§; the NJ tree differs in overall branching pattern from the UPGMA tree, but nevertheless retains the three PNG groups (data not shown).

Various characteristics of these three groups (and the other PNG types) are given in table 1. Group 1 differs dramatically from Groups 2 and 3 in three different characteristics (table 1). The frequency of individuals coming from coastal localities, speaking AN languages, and having the 9 b.p. deletion are all significantly elevated in Group 1. The 9 b.p. deletion is particularly noteworthy, as all of the individuals in Group 1 have the deletion, but all other PNG individuals lack the deletion. The distribution of the 9 b.p. deletion in Pacific populations (Hertzberg *et al.* 1989; Stoneking & Wilson 1989, Stoneking *et al.* 1992*b*) suggests that it was associated with the primary Austronesian/proto-Polynesian migration that probably began some 3500 to 5000 years ago (Bellwood 1989).

Groups 2 and 3 are similar to each other in residence, language, and 9 b.p. deletion composition. Furthermore, Groups 2 and 3 are each monophyletic groups of PNG mtDNA types, whereas Group 1 is not monophyletic but also includes non-PNG mtDNA types from Asia and Indonesia with the 9 b.p. deletion (including the one case of an identical mtDNA type shared by individuals from PNG, Indonesia, and the Philippines). This is consistent with Group 1 representing a recent migration to PNG, while Groups 2 and 3 represent a more ancient migration, presumably the earliest human colonization of PNG.

(*d*) *Intraspecific calibration: general considerations*

As outlined previously, the amount of divergence within and between Groups 2 and 3 can be used to estimate the rate of CR sequence evolution (Group 1 is too recent a migration to be used for this purpose). Group 3 consists of two extremely divergent clusters that are designated 3a and 3b (figure 4). For the purpose of calibrating the rate of CR sequence evolution, we consider 3a and 3b to represent separate migrations‖. If Group 3 is not split up, a faster rate of CR sequence evolution (and hence an even younger

§ UPGMA and NJ trees are not subject to the problems associated with parsimony analysis of these data (Hedges *et al.* 1992; Templeton 1992; Maddison *et al.* 1992, Stoneking *et al.* 1992*a*).

‖ This interpretation is supported by the restriction map analysis, since Groups 3a and 3b comprise different clans (Stoneking *et al.* 1990).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

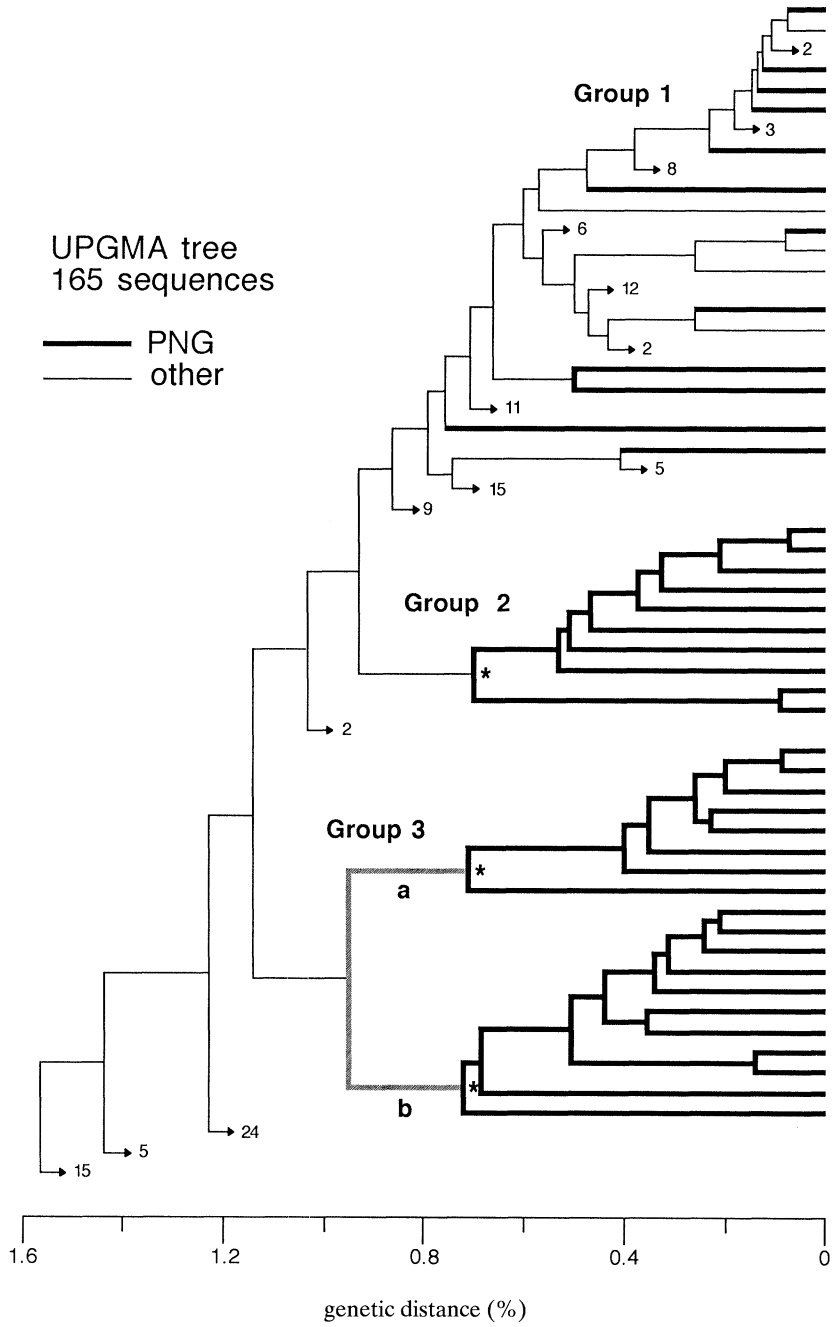

Figure 4. Abbreviated UPGMA tree for 165 human mtDNA CR sequences, rooted with a chimpanzee CR sequence (Foran *et al.* 1988). Branchpoints are positioned with respect to the scale of genetic (Jukes–Cantor) distance at the bottom of the tree. Thick lines indicate the 41 PNG mtDNA types while thin lines indicate non-PNG types. Thin lines terminating in arrows followed by a number indicate that branching structures involving the indicated number of non-PNG types occur at that location, but the actual detailed branching structure is omitted. The three PNG groups are identified (along with subgroups a and b of Group 3); asterisks indicate the deepest branchpoint and hence the origin of diversity within PNG for Groups 2, 3a, and 3b. The correspondence between PNG groups (based on CR sequences) and PNG mtDNA clans (based on restriction maps (Stoneking *et al.* 1990)) is as follows: Group 1 comprises clans 11–16; Group 2 comprises clan 10; Group 3a comprises clans 3, 4, 6 and 7; and Group 3b comprises clans 2 and 5. PNG CR sequences that do not fall within one of the three groups comprise clans 1, 8, 10, and 18; there is thus no overlap of clans between different groups.

age for the common human mtDNA ancestor) will result.

We therefore assumed that divergence within PNG commenced at the deepest points within Groups 2, 3a, and 3b (figure 4), and that these three groups were derived from the earliest migration of humans to PNG. Archaeological evidence places the earliest presence of humans in PNG at about 40 ka ago (Groube *et al.* 1986), and this date was used previously to calibrate the rate of mtDNA evolution based on restriction maps (Stoneking *et al.* 1986; Stoneking & Cann 1989). However, the earliest date for humans in Australia is about 53 ka ago (Roberts *et al.* 1990), at which time Australia and New Guinea formed one

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *Characteristics of PNG groups defined by* UPGMA *and* NJ *tree analysis*

Group	1	2	3	other	total
number of individuals	11	13	19	7	50
no. of types	6	10	19	6	41
residence (%)					
highland	0.0	84.6	63.2	85.7	58.0
coastal	100.0	15.4	36.8	14.3	42.0
language (%)[a]					
NAN	9.1	92.3	83.3	85.7	69.4
AN	90.9	7.7	16.7	14.3	30.6
9 b.p. deletion (%)					
non-deleted	0.0	100.0	100.0	100.0	78.0
deleted	100.0	0.0	0.0	0.0	22.0

[a] The language affiliation was not known for one member of group 3. NAN, non-Austronesia; AN, Austronesian.

land mass. To obtain the slowest rate of CR sequence evolution consistent with the data, and thus the oldest ages for the human mtDNA ancestor, a maximum time of 60 ka was assumed for the initial colonization of PNG. Dates less than 60 ka will result in faster rates, and hence even younger dates for the human mtDNA ancestor.

(*e*) *Intraspecific calibration: within-group approach*

For each of the three groups, the distance to the deepest node and the associated standard error are presented in table 2. These are converted into estimates of the rate of CR sequence evolution by assuming a colonization time of 60 ka, resulting in an average rate of CR sequence evolution of $11.81 \pm 3.11\%$ per Ma. The amount of sequence evolution corresponding to the human mtDNA ancestor in the UPGMA tree (figure 4) is $1.56 \pm 0.21\%$ (for the NJ tree the average amount of sequence evolution to the human mtDNA ancestor was nearly identical, 1.58%). For the within-group approach, then, the best estimate for the age of the human mtDNA ancestor is therefore 133 ka, whereas an approximate 95% confidence interval (based on 95% confidence intervals for the rate of CR sequence evolution and the amount of sequence evolution corresponding to the human mtDNA ancestor) is 63–356 ka ago.

Table 3. *Estimates of within- and between-group sequence divergence*

(Average divergence (%) within each group is on the diagonal, uncorrected estimates of divergence between each pair of groups (i.e. d_{XY}) are below the diagonal, and corrected estimates of divergence between each pair of group (i.e. d_A) are above the diagonal.)

group	2	3a	3b
2	0.84 ± 0.18	1.34 ± 0.44	1.78 ± 0.42
3a	2.17 ± 0.48	0.81 ± 0.18	1.00 ± 0.03
3b	2.72 ± 0.47	1.93 ± 0.38	1.03 ± 0.19

(*f*) *Intraspecific calibration: between-group approach*

Estimates of d_{XY} and d_A (calculated according to equation (1)) are given in table 3. The average value of d_A is $1.37 \pm 0.40\%$, which (assuming a colonization time of 60 ka) leads to an average rate of $\lambda = 11.42 \pm 3.33\%$. This rate is virtually identical to that estimated above by the within-group approach, and results in an age of the human mtDNA ancestor of 137 ka with an approximate 95% confidence interval of 63–416 ka.

5. DISCUSSION

(*a*) *Caveats*

We have presented two different approaches to an intraspecific calibration of the rate of human mtDNA CR sequence evolution; these two approaches yield virtually identical results. Both strategies are based on the divergence of PNG-specific groups of mtDNA types, and thus are not subject to the criticisms that have been levelled at calibrations based on comparing human and chimpanzee mtDNAs. However, the intraspecific calibration does require other assumptions that need to be evaluated.

First, the intraspecific calibration relies on an accurate date for the colonization of PNG. In order to arrive at the slowest possible rate consistent with the data, we have used a date somewhat older than archaeological evidence would suggest, namely 60 ka. However, this date has steadily increased (as now archaeological evidence has come to light), from 30 ka

Table 2. *Calibration of the rate of* CR *sequence evolution and estimates of the age of the common human mtDNA ancestor*

group	distance (%) to the group ancestor	rate (% per Ma) of CR sequence evolution[a]	human mtDNA ancestor[b]	
			age/ka	95% CI/ka
2	0.69 ± 0.18	11.55 ± 3.00	135	65–357
3a	0.71 ± 0.20	11.83 ± 3.27	132	62–374
3b	0.72 ± 0.18	12.03 ± 3.07	130	63–336
average	0.71 ± 0.19	11.81 ± 3.11	133	63–356

[a] Assuming that PNG was colonized 60 ka ago.
[b] Based on an average distance of $1.56 \pm 0.21\%$ to the common ancestor.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 4. *Comparison of estimates from Vigilant et al. (1991) and the present study for the amount of* CR *sequence divergence since the common ancestor, the rate of* CR *sequence divergence, and the age of the human mtDNA ancestor*

study	divergence (%) from ancestor	rate of divergence (%)	age of the ancestor/ka
Vigilant *et al.* (1991)[a]	2.87	11.5–17.3	166–249
this study[b]	3.12 ± 0.42	23.62 ± 6.22	63–356
this study[c]	3.12 ± 0.42	22.84 ± 6.66	63–416

[a] The ranges for the rate of divergence and age of the ancestor are derived from using a range of 4–6 Ma for the separation between chimpanzees and humans; these are not confidence intervals.
[b] Based on the average distance to the root of each PNG group.
[c] Based on the corrected divergence between PNG groups.

for the first intraspecific calibration (Stoneking *et al.* 1986), to 40 ka for the next calibration (Stoneking & Cann 1989), to 60 ka in the present study. More recent dates are not a problem, as they result in faster rates and hence younger estimates of the age of the human mtDNA ancestor. Older dates, however, will result in slower rates and more ancient estimates for the age of the human mtDNA ancestor.

Given the above trend, can we be confident that the date for the first entry of humans into Sahul (the combined Australia and New Guinea land masses) will not be pushed back even further with future knowledge? It is, of course, impossible to say if earlier dates will be forthcoming from studies of the archaeological and paleontological record in Sahul. However, it does not seem likely that the date will be pushed back much further, as the earliest appearance of anatomically modern humans in the fossil record occurs about 100 ka ago in Africa and the Near East (Stringer & Andrews 1988; Stringer *et al.* 1989). Even if the first entry of humans into Sahul did occur as much as 80 ka ago, the corresponding age of the human ancestor would increase from 133 to just 176 ka, with a maximum age of 464 ka (from the 95% confidence interval). Thus, increasing the age for the colonization of Sahul will not substantially alter our conclusions.

A second assumption is that migration was essentially one-way, so that there was no back-migration from PNG to Indonesia or southeast Asia. However, if undetected back-migration has occurred, then the resulting estimate of the age for the human mtDNA ancestor will be greater than the true age. For example, suppose that the nearest non-PNG neighbour of PNG Group 2 (figure 4) is a back-migrant. If this is the case, then the amount of sequence change observed in the tree for Group 2 will be less than the amount of sequence change that has actually accumulated since colonization. We therefore would underestimate the true amount of sequence evolution, and hence undetected back-migration would lead to an estimated rate that is slower than the true rate.

A third assumption is that the rate of mtDNA sequence evolution is the same in all human populations. One way to check this assumption is to test the pattern of mtDNA variation within each population for departure from neutral expectations. We applied the test developed by Tajima (1989) to the PNG data and found that the mtDNA variation in PNG is consistent with neutrality (data not shown). Vigilant (1990) tested the remaining populations and found three cases in which neutrality was rejected. Two of these probably represent departures from random sampling, as they involve agglomerations of 'Europeans' and 'Asians' that do not reflect known substructure within these continental groups (Horai & Hayasaka 1990; DiRienzo & Wilson 1991). The third involves the Herero of Botswana, in which there is historical evidence for a recent bottleneck (Vigilant 1990). We thus find no evidence that selection has distorted the rate of mtDNA evolution within human populations.

Finally, it is assumed that the PNG groups identified by phylogenetic analysis truly are monophyletic clusters, and that further work will not find mtDNA types from outside PNG that fall within these groups. This is a critical assumption, because if the PNG groups are not truly monophyletic, then we have overestimated the amount of sequence evolution that occurred within PNG after colonization. The amount of sequence evolution that we determine to occur within PNG would actually have occurred prior to the colonization of PNG. The result will be an overestimate of the rate of CR sequence evolution, and a corresponding underestimate of the true age of the human mtDNA ancestor.

In order to assess accurately the validity of the PNG groups, it is necessary to sample mtDNAs from locations ancestral to the colonization of PNG. The Indonesian samples were collected for precisely this purpose from populations that exhibit some physical Melanesian characteristics. However, the 20 Indonesian CR sequences analyzed here were pre-selected to have the 9 b.p. deletion that is associated with PNG Group 1, and indeed they all fall within Group 1. Hence, they would not be expected to either be ancestral to or break up PNG Groups 2 or 3. We intend to seek Indonesian samples that might fall within Groups 2 or 3 by using hybridization with sequence-specific oligonucleotide probes (e.g. Stoneking *et al.* 1991) to screen for nucleotide substitutions that characterize the PNG groups. CR sequences will then be determined for any Indonesian sample that might enter into one of the PNG groups. Another future direction to address this problem is to repeat the intraspecific calibration with CR sequences from

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Australia and other places that would appear to satisfy the requirements of this analysis.

(*b*) *Comparison with previous results*

The results presented herein build on the data of Vigilant *et al.* (1991) with additional CR sequences from PNG and Indonesia. In comparing the present results with the previous work, it is important to note that here amounts and rates of change are expressed in terms of changes observed from an ancestor to an existing type (i.e. as amounts and rates of CR sequence evolution), whereas in Vigilant *et al.* (1991) the corresponding quantities are expressed in terms of changes observed between two existing types (i.e. as amounts and rates of CR sequence divergence). Thus, our values need to be doubled in order to compare them to the values calculated previously. The resulting estimates of the amount of CR sequence divergence since the mtDNA ancestor, the rate of CR sequence divergence, and the age of the human mtDNA ancestor are compared with the previous estimates in table 4. The values from the two studies are in good agreement, even though quite different methods were used to calibrate the rate of CR sequence evolution or divergence.

Vigilant *et al.* (1991) did not estimate confidence intervals for the rate of CR sequence divergence or the age of the human mtDNA ancestor, because methods did not exist for calculating the variance of the transition:transversion ratio, which they used to calibrate the rate of CR sequence divergence. Recently, a method has been developed for calculating the variance of the transition:transversion ratio (M. Nei, personal communication), and the resulting 95% confidence interval for the age of the human mtDNA ancestor is approximately 110–500 ka. Hasegawa & Horai (1991) have also investigated the variance of the substitution rate, using a maximum likelihood method that was applied to subsets of the human CR sequences published by (Greenberg *et al.* 1983; Vigilant *et al.* 1989; Horai & Hayasaka 1990). Their 95% confidence interval for the age of the human mtDNA ancestor was about 180–380 ka. However, they caution that their model may be violated by the different patterns of nucleotide substitution observed in chimpanzee and human CR sequences. Finally, Stoneking & Cann (1989) predicted that a 95% confidence interval for the age of the human mtDNA ancestor would probably be on the order of 50–500 ka ago; this conjecture seems to be borne out by the various methods and datasets that have been used to estimate such confidence intervals.

(*c*) *Implications and future prospects*

There are two important implications of this work. First, the two approaches for intraspecific calibration yielded an average age for the human mtDNA ancestor of about 135 ka. This is more recent than the previous estimates (Cann *et al.* 1987; Hasegawa & Horai 1991; Vigilant *et al.* 1991), which centred around 200 ka ago. In the past it was emphasized that this date is for the human *mtDNA* ancestor, which is not necessarily the same as the actual ancestor of our species, namely anatomically modern humans. There is no requirement for the human mtDNA ancestor to be a member of our species; she might very well have been a member of some preceding species¶. However, the dates for the mtDNA ancestor from this work are coming close to the earliest dates ascertained for fossils of anatomically modern humans (Stringer & Andrews 1988; Stringer *et al.* 1989), raising the intriguing possibility that the human mtDNA ancestor might actually have been one of the first anatomically modern humans.

The second major implication of this work is that the oldest age for the human mtDNA ancestor is 356–416 ka ago. Since genetic divergence is expected to precede population divergence (Nei 1987, p. 288), the maximum age of the human mtDNA ancestor places an upper bound on when human populations began to diverge. In other words, if the human mtDNA ancestor lived not more than 416 ka ago, then human population differences cannot date back more than 416 ka (and, of course, they could be considerably younger). Therefore, non-African populations that date back more than 416 ka ago could not have contributed any mtDNA types to modern populations. This finding is clearly incompatible with the multiregional evolution hypothesis (Wolpoff 1989), which holds that human population differences should be upwards of 1 Ma old.

However, before we can completely discount the multiregional evolution hypothesis, it must be pointed out that we have analysed variation in only a single gene, namely mtDNA. Although our analysis accounts for the variance in the two components of the age of the human mtDNA ancestor (namely, the rate of CR sequence evolution and the amount of CR sequence evolution since the human mtDNA ancestor), there is a third component that introduces variance into the estimate of when human population divergence actually began. This third component is stochastic variance in the evolutionary history of different genes (Nei & Livshits 1990). Only by analysing different genes can we account for this variance; although mtDNA analyses are valuable for the insights they yield into human evolutionary history, many more genes need to be analysed in similar fashion for a statistically accurate picture of the history of human populations to emerge.

We thank K. Bhatia and A. S. M. Sofro for samples, M. Nei and T. Whittam for computer programs, R. Zauhar for computational assistance, and H. Harpending, B. Hedges and M. Nei for valuable discussion and comments. Research supported by NSF grant BNS 90-20567 to M.S.

REFERENCES

Bellwood, P.S. 1989 The colonization of the Pacific: some current hypotheses. In *The colonization of the Pacific – a genetic trail* (ed. A. V. S. Hill & S. W. Serjeantson), pp. 1–59. Oxford University Press.

¶ Indeed, based on the 200,000 year old date, she was deemed most likely to be anatomically-archaic (Stoneking & Cann 1989), not modern.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Cann, R.L., Stoneking, M. & Wilson, A.C. 1987 Mitochondrial DNA and human evolution. *Nature, Lond.* **325**, 31–36.

DiRienzo, A. & Wilson, A.C. 1991 Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natn. Acad. Sci. U.S.A.* **88**, 1597–1601.

Foran, D.R., Hixson, J.E. & Brown, W.M. 1988 Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucl. Acids Res.* **16**, 5841–5861.

Greenberg, B.D., Newbold, J.E. & Sugino, A. 1983 Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Genetics* **21**, 33–49.

Groube, L.M., Chappell, J., Muke, J. & Price, D. 1986 A 40,000 year-old human occupation site at Huon Peninsula, Papua New Guinea. *Nature, Lond.* **324**, 453–455.

Gyllensten, U.B. & Erlich, H.A. 1988 Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natn. Acad. Sci. U.S.A.* **85**, 7652–7656.

Hasegawa, M. & Horai, S. 1991 Time of the deepest root for polymorphism in human mitochondrial DNA. *J. molec. Evol.* **32**, 37–42.

Hedges, S.B., Kumar, S., Tamura, K. & Stoneking, M. 1992 Human origins and the analysis of mitochondrial DNA sequences. *Science, Wash.* **255**, 737–739.

Hertzberg, M., Mickleson, K.N.P., Serjeantson, S.W., Prior, J.F. & Trent, R.J. 1989 An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *American J. hum. Genet.* **44**, 504–510.

Horai, S. & Hayasaka, K. 1990 Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. hum. Genet.* **46**, 828–842.

Jukes, T.H. & Cantor, C.R. 1969 Evolution of protein molecules. In *Mammalian protein metabolism* (ed. H. N. Munro), pp. 21–132. New York: Academic Press.

Kimura, M. 1980 A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. molec. Evol.* **16**, 111–120.

Kocher, T.D. & Wilson, A.C. 1991 Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. In *Evolution of life: fossils, molecules, and culture* (ed. S. Osawa & T. Honjo), pp. 391–413. Tokyo: Springer-Verlag.

Maddison, D.R., Ruvolo, M. & Swofford, D.L. 1992 Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. *Syst. Biol.* (In the press.)

Merriweather, D.A., Clark, A.G., Ballinger, S.W., Schurr, T.G., Soodyall, H., Jenkins, T., Sherry, S.T. & Wallace, D.W. 1991 The structure of human mitochondrial DNA variation. *J. molec. Evol.* **33**, 543–555.

Nei, M. 1987 *Molecular evolutionary genetics.* New York: Columbia University Press.

Nei, M. & Jin, L. 1989 Variances of the average numbers of nucleotide substitutions within and between populations. *Molec. Biol. Evol.* **6**, 290–300.

Nei, M. & Livshits, G. 1990 Evolutionary relationships of Europeans, Asians, and Africans at the molecular level. In *Population biology of genes and molecules* (ed. N. Takahata & J. F. Crow), pp. 251–265. Tokyo: Baifukan.

Nei, M., Stephens, J.C. & Saitou, N. 1985 Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. *Molec. Biol. Evol.* **2**, 66–85.

Roberts, R.G., Jones, R. & Smith, M.A. 1990 Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. *Nature, Lond.* **345**, 153–156.

Saitou, N. & Nei, M. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.* **4**, 406–425.

Stoneking, M., Bhatia, K. & Wilson, A.C. 1986 Rate of sequence divergence estimated from restriction maps of mitochondrial DNAs from Papua New Guinea. *Cold Spring Harb. Symp. quant. Biol.* **51**, 433–439.

Stoneking, M. & Cann, R.L. 1989 African origin of human mitochondrial DNA. In *The human revolution – behavioural and biological perspectives on the origins of modern humans* (ed. P. Mellars & C. Stringer), pp. 17–30. Edinburgh University Press.

Stoneking, M., Hedgecock, D., Higuchi, R.G., Vigilant, L. & Erlich, H.A. 1991 Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. hum. Genet.* **48**, 370–382.

Stoneking, M., Jorde, L.B., Bhatia, K. & Wilson, A.C. 1990 Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124**, 717–733.

Stoneking, M., Sherry, S.T. & Vigilant, L. 1992*a* Geographic origin of human mitochondrial DNA revisited. *Syst. Biol.* (In the press.)

Stoneking, M., Sofro, A.S.M. & Wilson, A.C. 1992*b* Implications of a mitochondrial marker for the colonization of the Pacific. In *Human evolution in the Pacific Region* (ed. C. K. Ho, G. Krantz & M. Stoneking). Pullman, Washington: Washington State University Press.

Stoneking, M. & Wilson, A.C. 1989 Mitochondrial DNA. In *The colonization of the Pacific: a genetic trail* (ed. A. V. S. Hill & S. Serjeantson), pp. 215–245. Oxford University Press.

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

Stringer, C.B., Grün, R., Schwarcz, H.P. & Goldberg, P. 1989 ESR dates for the hominid burial site of Es Skhul in Israel. *Nature, Lond.* **338**, 756–758.

Tajima, F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Templeton, A.R. 1992 Human origins and analysis of mitochondrial DNA sequences. *Science, Wash.* **255**, 737.

Vigilant, L. 1990 Control region sequence from African populations and the evolution of human mitochondrial DNA. PhD Thesis, University of California, Berkeley.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D. & Wilson, A.C. 1989 Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natn. Acad. Sci. U.S.A.* **86**, 9350–9354.

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

Wolpoff, M.H. 1989 Multiregional evolution: the fossil alternative to Eden. In *The human revolution – behavioural and biological perspectives on the origins of modern humans* (ed. P. Mellars & C. Stringer), pp. 62–108. Edinburgh University Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES