ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/14126594

A high observed substitution rate in the human mitochondrial DNA control region

Article *in* Nature Genetics · May 1997

DOI: 10.1038/ng0497-363 · Source: PubMed

CITATIONS
467

READS
3,660

11 authors, including:

Thomas Parsons
International Commission on Missing Persons
111 PUBLICATIONS **8,277** CITATIONS

SEE PROFILE

Kevin Sullivan
Sullivan Forensics Ltd
37 PUBLICATIONS **3,263** CITATIONS

SEE PROFILE

Dianna L. Berry
Stony Brook University
42 PUBLICATIONS **2,639** CITATIONS

SEE PROFILE

All content following this page was uploaded by Dianna L. Berry on 22 May 2014.

The user has requested enhancement of the downloaded file.

© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

A high observed substitution rate in the human mitochondrial DNA control region

Thomas J. Parsons[1], David S. Muniec[1], Kevin Sullivan[2], Nicola Woodyatt[2], Rosemary Alliston-Greiner[2], Mark R. Wilson[3], Dianna L. Berry[4], Koren A. Holland[4], Victor W. Weedn[1], Peter Gill[2] & Mitchell M. Holland[1]

The rate and pattern of sequence substitutions in the mitochondrial DNA (mtDNA) control region (CR) is of central importance to studies of human evolution and to forensic identity testing. Here, we report a direct measurement of the intergenerational substitution rate in the human CR. We compared DNA sequences of two CR hypervariable segments from close maternal relatives, from 134 independent mtDNA lineages spanning 327 generational events. Ten substitutions were observed, resulting in an empirical rate of 1/33 generations, or 2.5/site/Myr. This is roughly twenty-fold higher than estimates derived from phylogenetic analyses. This disparity cannot be accounted for simply by substitutions at mutational hot spots, suggesting additional factors that produce the discrepancy between very near-term and long-term apparent rates of sequence divergence. The data also indicate that extremely rapid segregation of CR sequence variants between generations is common in humans, with a very small mtDNA bottleneck. These results have implications for forensic applications and studies of human evolution.

Sequences from the mitochondrial DNA (mtDNA) control region (CR) are highly variable within human populations and have been a primary source of information regarding the genetic structure, age and origin of modern *Homo sapiens*[1–10]. In these and other studies of mtDNA RFLP or coding sequence variation[11–15], a standard approach is to derive a phylogenetic tree, then date branch lengths by reference to an assumed molecular clock, calibrated with divergence dates from the hominoid fossil or human archaeological record. However, this approach assumes neutral evolution and a known relationship between observed sequence divergences and the mode and rate with which substitutions accumulate. Earlier studies estimated a single mtDNA (or CR) substitution rate, with attempts to correct for multiple substitutions at the same site[1,6,11,12]. However, it is now clear that models employing increasing numbers of parameters, that accommodate various rates and classes of substitution, provide a better fit to the observed data[15–24]. These analyses point toward complex processes by which mtDNA sequences accumulate differences over time. As a result, studies based on different models of sequence evolution, or focusing on different subclasses of sites, have produced dates for the mtDNA most-recent-common-ancestor (MRCA) ranging from ~70,000 to 600,000 years ago — not including broad confidence intervals associated with particular estimates[11,22,24]. Thus, rate and pattern of mtDNA substitution remains an unsettled issue of central importance.

The high variability of mtDNA CR sequences also permits utility in forensic identity investigations[25–31]. The high copy number of mtDNA allows recovery from badly degraded, trace, or 'ancient' samples. Because mtDNA is maternally-inherited, comparisons can be made between even distant maternal relatives. However, a high substitution rate raises the possibility that maternal relatives will sometimes differ at one or more base positions. This has been seen in the case of Tsar Nicholas II[30,31] and was encountered with surprising frequency in the course of forensic casework involving sequence comparisons of family reference samples (see below). The number of forensic cases involving mtDNA is increasing rapidly (for example, in the ongoing mtDNA program to aid in 'full accounting' for >10,000 U.S. personnel missing from conflicts since World War II and in identity testing of shed hairs found at crime scenes). Thus, mtDNA substitution will have practical implications that must be addressed explicitly.

We performed an extensive collaborative study to empirically determine the frequency with which maternal relatives differ in mtDNA sequence. We compared sequences of the two CR hypervariable regions (HV-1 and HV-2, ~610 bp total) from close maternal relatives (predominantly mother:child, grandmother:grandchild, or sibling pairs), from a large number of mtDNA lineages. The result of this fine-structure analysis of mtDNA variation is an empirically-observed CR substitution rate that is approximately twenty-fold higher than rates calculated from phylogenetic studies. This observed rate has implications for both forensic identity testing and for the inference of human history from mtDNA sequences.

Observed substitutions

Samples were from four sources: family reference blood samples sequenced in the course of forensic casework at the Armed Forces DNA Identification Laboratory (AFDIL) (73 mtDNA lineages, 121 generational events); blood samples from 'Oxford' British families (five mtDNA lineages, 32 generational events); DNA from CEPH pedigree cell lines (16 mtDNA lineages, 94 generational events); and DNA from Old Order Amish pedigree cell lines (40 mtDNA lineages, 80 generational events). As an internal control for sequence quality and to further establish the reproducibility of DNA sequencing in a forensic context,

[1]*The Armed Forces DNA Identification Laboratory, Office of the Armed Forces Medical Examiner, The Armed Forces Institute of Pathology, 1413 Research Blvd., Rockville, Maryland 20850, USA.* [2] *The Forensic Science Service, Priory House, Gooch Street North, Birmingham, UK.* [3]*The FBI Laboratory, Washington, DC 20535, USA.* [4]*Department of Chemistry, Gettysburg College, Gettysburg, Pennsylvania 17325, USA. Correspondence should be addressed to T.J.P.*

© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

sequence determination was replicated in independent laboratories for 63 of the 357 individuals sequenced in our study. In all cases, the sequencing results matched exactly between laboratories. Additionally, sequences were compared for 69 father:child pairs and as expected, no evidence for paternal transmission of mtDNA was observed.

Overall, 327 'generational events' were screened and within these, ten instances of substitution were detected (Table 1). Taken together, our data indicate a remarkably high substitution rate, ~one in 33 generations. Assuming a generation time of 20 years, this extrapolates to a substitution rate of 2.5/site/Myr (95% confidence interval, 1.2–4.0/site/Myr). However, the different sample types appear to display different rates. Sequences derived from blood samples detected eight substitutions within 153 generations (0.052/generation, 4.3/site/Myr), whereas sequences derived from cell lines revealed two substitutions in 174 generations (0.011/generation, 0.94/site/Myr) (the difference is significant at the P=0.03 level). Lacking a resolution to the apparent different rates observed for blood versus cell line samples (discussed below), we will adopt the pooled value (0.03/generation) for the purposes of consideration and comparison. The observed substitution rates from both the blood samples and cell lines greatly exceed evolutionary rates inferred from phylogenetic studies.

Table 1 • Summary of sequence comparisons and observed mutations

Sample source	mtDNA lineages	Generations	Mutations	Rate per generation	Rate pooled by type
AFDIL family references	73	121	7	0.0578 1/17	blood samples 0.052 1/19
'Oxford' families	5	32	1	0.029 1/32	
Amish families	40	80	1	0.013 1/80	cell lines 0.011 1/87
CEPH families	16	94	1	0.011 1/94	

Substitutions were observed throughout HV-1 and HV-2 (Fig. 1, Table 2), without obvious overall clustering. However, two substitutions in different lineages were at adjacent sites (16092 and 16093; numbering as in ref. 32). Remarkably, two independent instances of substitution were detected at position 207. In one mtDNA lineage, a daughter differed from her mother and brother at two positions, 207 and 16093 (maternal relationship in this case was unambiguously confirmed by nuclear DNA fingerprinting[33]). One observed substitution (309.1) was a cytosine insertion, occurring in the HV-2 polycytosine 'C-stretch' region. The 309.1C insertion is a common variant, shared by 49% of individuals in our database of 742 CR sequences; most individuals are heteroplasmic at low levels for length variants in this HV-2 C-stretch (AFDIL, unpublished data). Four other substitutions were observed at sites that display above-average levels of polymorphism, but five observed substitutions were at sites with slightly or substantially lower-than-average frequencies of polymorphism.

In most cases, the direction of the substitution is known. Three exceptions are substitutions that occurred in AFDIL family references for which only sequences between siblings were available. In another case (Amish lineage, presented below) the initial mutation was from a rare to a common nucleotide variant, but differential segregation in subsequent generations was to both variants. In all but one of the remaining instances where substitution direction is known, the changes were from a common to a rare nucleotide variant. Apart from the 309.1 insertion, all observed substitutions were transitions, 67% between purines and 33% between pyrimidines. Among the transitions with known direction, there is no evidence for directional bias: there were three instances of A to G transitions, two instances of G to A and one each of C to T and T to C.

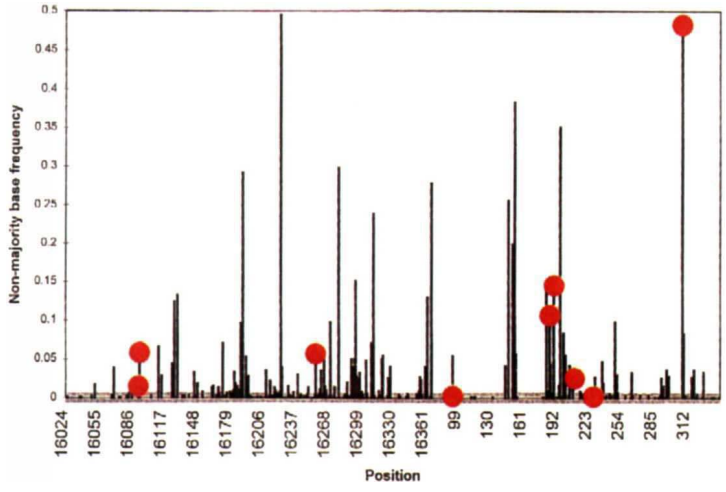

Fig. 1 Frequency of occurrence of non-majority bases versus position in the human mtDNA control region (based on 742 individual sequences from various races). Red dots indicate the positions where substitutions were observed in this study. Two independent substitutions were observed at position 207.

Observed heteroplasmy:

In our study, heteroplasmy was detected in an extended analysis of one Amish lineage where a substitution was observed (Fig 2). The initial grandmother:grandchild comparison showed the grandmother apparently fixed for 16092C, whereas the grandchild had 16092T. Subsequent analysis showed that the mother of the grandchild was heteroplasmic at position 16092, at a ratio of ~0.7T:0.3C. When five additional siblings of the grandchild were analyzed, all were apparently fixed for 16092C, having reverted to the grandmother's type. Closer scrutiny of multiple sequencing reactions revealed that the grandchild with the 16092T substitution also carried a very low percentage of C at that position. Other lineages where substitutions were observed were also scrutinized for heteroplasmy. In one case, one individual of a sibling pair was clearly heteroplasmic at a ratio of approximately 70:30 at position 16256, while the other sibling appeared fixed for the minority variant of the first sibling (the mother's sequence was unavailable). Low level heteroplasmy was also discernible in both of the siblings who differed at the 309.1C insertion. In several other lineages, the electropherograms were suggestive that heteroplasmy might be present at low levels, but firm conclusions could not be drawn from the direct sequencing data.

© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

Table 2 • Observed control region substitutions

Site	Substitution	Non-maj. base freq.	N.M. freq. av. N.M. freq.	Sample source
16092	C→T	0.012	0.36	Amish
16093	T→C	0.057	1.7	AFDIL Ref
16256	T⇄C	0.053	1.6	AFDIL Ref
94	G⇄A	0.0013	0.039	AFDIL Ref
185	G→A	0.093	2.8	AFDIL Ref
189	A→G	0.14	4.2	'Oxford'
207	G→A	0.021	0.64	AFDIL Ref
207	A→G	0.021	0.64	AFDIL Ref
234	A→G	0.0013	0.039	CEPH
309.1	In./del.	0.49	15	AFDIL Ref

Non-maj. base freq. indicates the proportion of sequences in database that differ from the standard sequence at the position of substitution. N.M. freq./av. N.M. freq. is the ratio of the non-majority base frequency at the position of substitution to the average non-majority base frequency (0.033).

Discussion

The observed substitution rate reported here is very high compared to rates inferred from evolutionary studies. A wide range of CR substitution rates have been derived from phylogenetic studies, spanning roughly 0.025–0.26 /site/Myr, including confidence intervals[3,15,16,18,20,34]. A study yielding one of the faster estimates gave the substitution rate of the CR hypervariable regions as 0.118 ± 0.031/site/Myr[6]. Assuming a generation time of 20 years, this corresponds to ~1/600 generations and an age for the mtDNA MRCA of 133,000 y.a. Thus, our observation of the substitution rate, 2.5/site/Myr, is roughly 20-fold higher than would be predicted from phylogenetic analyses. Using our empirical rate to calibrate the mtDNA molecular clock would result in an age of the mtDNA MRCA of only ~6,500 y.a., clearly incompatible with the known age of modern humans. Even acknowledging that the MRCA of mtDNA may be younger than the MRCA of modern humans[35], it remains implausible to explain the known geographic distribution of mtDNA sequence variation by human migration that occurred only in the last ~6,500 years[11,36].

While our results are at odds with those of phylogenetic studies, they are in excellent agreement with a recent report that also directly measured the CR substitution rate[37]. That study compared CR (and protein coding) sequences from multiple individuals within a single mtDNA lineage that carries a Leber hereditary optic neuropathy (LHON) mutation. Assuming no reversion mutations, a total of 81 generational events were surveyed and two CR mutations were observed, ~1/40 generations. In that case, multiple clones of PCR-amplified DNA were sequenced and the mutations were detected within low level heteroplasmic mixtures. However, at least one individual was virtually fixed for each mutation, so the same rate would have been obtained using our approach of directly sequencing the PCR product. Our results extend the observation of an unexpectedly high substitution rate to multiple independent lineages, indicating that the high rate is a general characteristic of mtDNA evolution and not an artifact of a single lineage that carries a pathogenic mitochondrial mutation.

What could account for the disparity between the observed substitution rate and those derived from phylogenetic analyses? One initially attractive explanation is that we have observed substitutions predominantly at mutational 'hot spots,' while phylogenetic estimates reflect rates averaged over all sites. Phylogenetic studies indicate that over extended time periods different sites manifest widely different substitution rates[16,18,19,21]. Because our study involves a very restricted period of evolutionary time, it is reasonable to suspect that we would observe substitutions at sites where they occur most rapidly. This explanation could readily be applied to the highly polymorphic insertion at position 309.1 and two additional substitution sites have a non-majority base frequency that is greater than twice the average value. However, five observed substitutions were at sites with levels of polymorphism below the average of 0.033. Moreover, the distribution of polymorphism at sites where substitutions were observed is similar to that exhibited by the pool of all polymorphic sites in our random, multi-racial database of 742 CR sequences (Fig. 1). For example, 21% of the polymorphic sites observed in the database have a non-majority base in only a single individual and 20% of observed substitutions occurred at such sites. Additionally, 68% of all polymorphic sites have a non-majority base frequency of less than 0.1, with 60% of the observed substitutions occurring at such sites. Furthermore, the nucleotide at position 207, where two independent mutations were observed, is not unusually variable, with a non-majority variant in only 2% of the population. Thus, the 'hot spot' hypothesis, in the absence of additional elements, does not seem a sufficient explanation for the high observed substitution rate.

The above discussion assumes a general correlation between the substitution rate and level of polymorphism among sites, as has been established[22]. However, at any particular site chance historical events could obscure a direct relationship between the substitution rate and the level of polymorphism. This was addressed for HV-1 sequences using a phylogenetic approach, to infer the actual number of times substitutions have occurred in history: 29 'fast' sites, with five or more independent substitutions, were identified[21]. Of the three observed substitutions that occurred within HV-1, only one occurs at one of these 'fast' sites. Again, it does not seem that the high observed substitution rate can be explained mainly by having sampled a restricted class of sites with unusually high rates of substitution.

There appears to be a difference in the rate and pattern of substitutions observed directly between generations, as compared to those manifested by sequences that diverged over longer periods of time. Various processes might account for this, singly or in combination. It may be that new substitutions, while common between generations, are usually eliminated through random genetic drift before reaching an appreciable frequency in the population. Another possibility is that of an inherent mechanistic bias in the frequencies of particular mutational transitions, so that some substitutions revert relatively rapidly to the more stable original state (however, it seems this would have to occur variably in a site-dependent fashion, as our overall data indicate roughly equal proportions of transitions in both directions). Finally, it is possible that some CR substitutions are slightly deleterious and over time are selectively removed from the population. This would result in a low population frequency, despite a relatively high substitution rate. Departure from neutrality has been observed in mtDNA protein-coding genes[38,39] and in RFLP variation over the entire mtDNA genome[38,40]. While the CR is certainly under less selectional constraint than coding genes, the region has crucial regulatory functions and internal sub-regions display quite different levels of variation both within and between species[41]. Some portions of the CR are thus not as free to evolve as others and it is quite plausible that selectional constraint, while clearly present, need not be absolute between one generation and the next. In this light, it is interesting to note that the observed substitution of the nucleotide at position 234 occurs within conserved sequence block CSB-1. In a wide survey of mammalian species, spanning ~200 Myr of sequence divergence, the nucleotide at position 234 appears uniquely as adenine[41]. Such conservation suggests selectional constraint at this site and 234G may be a substitution that can exist only transiently in the population (in our database, at a frequency of 0.0013).

The apparent difference in observed substitution rate between blood samples and cell lines is difficult to interpret. We suspect

© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

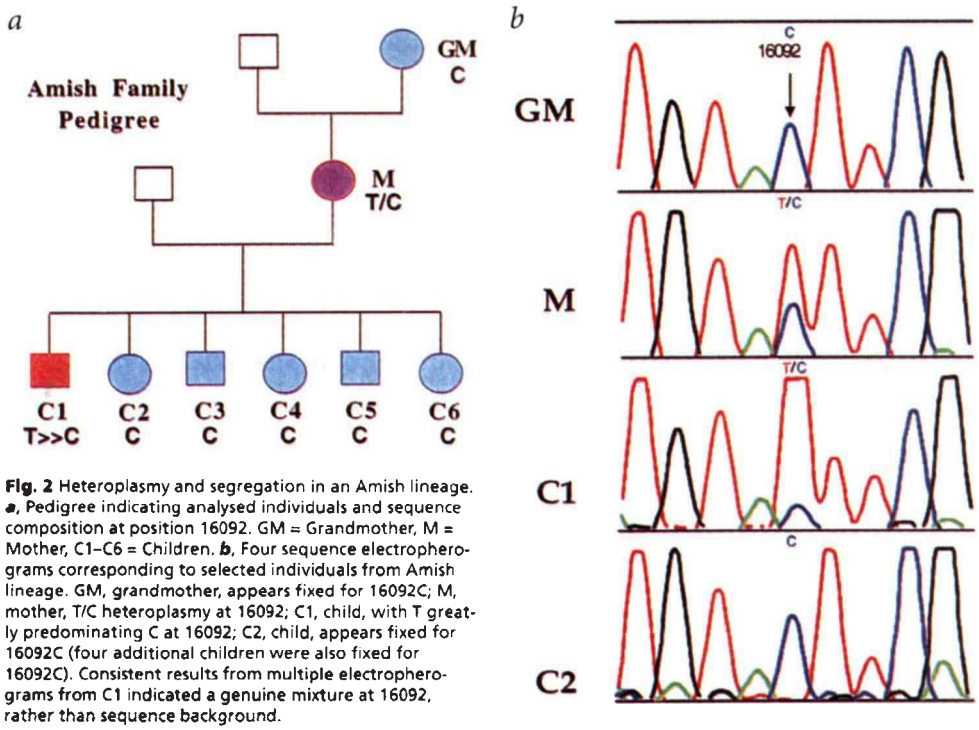

Fig. 2 Heteroplasmy and segregation in an Amish lineage. *a*, Pedigree indicating analysed individuals and sequence composition at position 16092. GM = Grandmother, M = Mother, C1–C6 = Children. *b*, Four sequence electropherograms corresponding to selected individuals from Amish lineage. GM, grandmother, appears fixed for 16092C; M, mother, T/C heteroplasmy at 16092; C1, child, with T greatly predominating C at 16092; C2, child, appears fixed for 16092C (four additional children were also fixed for 16092C). Consistent results from multiple electropherograms from C1 indicated a genuine mixture at 16092, rather than sequence background.

that this may be due simply to chance, as the difference is only marginally significant (P=0.03). If there were a real difference between these mtDNA sources, one might expect that the cell lines would display an artificially higher rate of substitution due to extended replications in culture. However, the opposite trend is observed. One conceivable explanation is that cell lines display fewer differences within lineages because of enhanced selection in culture, for example, favouring the original CR sequences (that may replicate more efficiently). If this were the case, the data from the cell lines would underestimate the actual substitution rate.

In six of the ten observed substitutions, an essentially complete shift in the predominant mtDNA type occurred across a single sexual generation. In three other cases, substitutions were observed between siblings: at the position in question, one sibling pair showed a fixed difference with no evidence of heteroplasmy, another pair had one member with a 70:30 heteroplasmic mixture and the third pair showed low level heteroplasmy of opposite ratio in both members (in the HV-2 C-stretch where heteroplasmy is common). In addition, the Amish lineage (Fig. 2) showed five offspring of a heteroplasmic mother that were apparently fixed for her minority mtDNA variant and a sixth that was nearly fixed for the other variant. These observations indicate an extremely rapid segregation of mtDNA variants across generations.

A bottleneck model has been well established in mtDNA studies to account for rapid, differential segregation of mtDNA variants[42]. There is conflicting information regarding the size of the intergenerational bottleneck, both between species[43,44] and within humans[45,46]. An elegant series of experiments has recently been reported that provides a detailed picture of mtDNA segregation in artificially heteroplasmic mice[47]. The effective bottleneck was estimated to be 200 mtDNA molecules, corresponding to a mean time to fixation of new mutations of fifteen generations. This is clearly inconsistent with the pattern in humans we report here. The difference may reflect inherent differences between the species in the mechanism of segregation, or it may be related to the manner in which the segregating variants originated (through cytoplasmic transfer of homogenomic mitochondria in the mouse study and, presumably, through intra-organellar mutation in humans). Another recent study of heteroplasmic segregation in human mtDNA lineages indicates that, in fact, the size of the bottleneck can vary significantly from one meiotic generation to the next[46]. Our observations indicate that the effective bottleneck in humans is commonly very small indeed.

Because our study employed direct sequencing of PCR-amplified DNA, our observations were functionally limited to instances where the mtDNA populations within individuals shifted between generations, from one clearly predominant type to another. The frequency of these 'substitutions' is not equivalent to the actual mutation rate of mitochondrial DNA, although it is directly comparable to the evolutionary rates inferred from phylogenetic studies (which are commonly referred to as 'mutation rates'). In actuality, mutations (presumably) occur on single mtDNA molecules, that must then segregate within a larger mtDNA pool at the organellar, cellular, intergenerational and developmental levels before they can appear as substitutions in the evolution of an mtDNA lineage. The actual mutation rate is masked by a number of essentially uncharacterized processes and direct observation of this rate is extremely difficult from the standpoint of detection sensitivity. In our study heteroplasmy was clearly detected in five individuals from three lineages. While this adds to a growing body of evidence that point mutation heteroplasmy in the human CR may not be particularly rare[30,31,37,46,48,49], a suitable estimate of the frequency of heteroplasmy is still lacking. This is because direct sequencing is not an appropriate method for detecting the frequency of heteroplasmy, as even fairly balanced heteroplasmic ratios can be difficult to distinguish from various types of sequencing artifacts. We therefore feel that a recently reported 'direct' mutation rate, calculated on the basis of the frequency with which heteroplasmy was observed in direct sequencing studies, is likely to be inaccurate and probably biased toward underestimation (although the resulting 22-fold range in mutation rate overlaps both with substitution rates from some phylogenetic studies and our much higher range for the observed substitution rate)[46].

Our results have implications for the use of CR sequences in forensic identity testing. mtDNA is often employed to compare questioned samples to presumed maternal references. It is now clear that the mtDNA substitution rate is sufficiently high that differences between true maternal relatives will be encountered not infrequently, providing the grounds for false exclusion. Presently, it is our policy not to report an exclusion based on a single CR sequence difference. In such a case, one can evaluate significance of the evidence through likelihood ratio calculations that incorporate the probability that a mutation has occurred

© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

within the lineage[30,31]. Our results indicate that previous phylogenetic estimates of substitution rate are not suitable for this purpose and the likelihood approach would be abetted by additional precise information on the intergenerational rate and pattern of CR sequence change. Finally, our results indicate that rapid sequence substitution on the very near term is masked by uncharacterized factors (drift or selection) when more highly diverged sequences are analysed. Further progress in interpreting mtDNA sequence comparisons in evolutionary or population genetic studies may require greater insight into these factors and the time scale on which they operate.

Methods

Samples. Organically extracted DNA from CEPH (Centre d'etude Polymorphisme Humane) reference family cell lines was obtained from the National Institute of Health and Mental Health (for AFDIL) and from the Imperial Cancer Research Fund (for the British Forensic Science Service (FSS)). Ten CEPH families were evaluated; Utah Pedigrees 1329, 1331, 1332, 1333, 1340, 1341, 1345, 13292, 13293 and 13294. Three of the pedigrees (1329, 1331 and 1333) were evaluated by both AFDIL and the British FSS. DNA was obtained from 38 'Old Order' Amish families and from two reformed Amish families, both from the state of Pennsylvania. The 'Old Order' Amish genomic DNAs were extracted organically from immortalized T or B cells and provided by the Johns Hopkins University, Immunogenetics Laboratories. Whole blood was collected from six reformed Amish family members from central Pennsylvania and DNA was extracted from prepared bloodstains using the Chelex method[50]. Whole blood was collected from U.S. Caucasians to be used as family reference sources for mtDNA casework and DNA was extracted using the Chelex method. Organically-extracted genomic DNA from British 'Oxford' families were obtained from Oxford University (for the British FSS).

Database. Levels of polymorphism at various sites were determined by reference to a combined forensic database of 742 HV-1 and HV-2 sequences from the following ethnic groups: 90 African-American (53, FBI; 40, AFDIL), 115 Afro-Caribbean (FSS), 114 Sierra Leone African (provided by C. Ginther), 90 Hispanic (provided by C. Ginther), 100 British Caucasian (FSS), 233 European American (AFDIL). Our database represents random, unrelated individuals. However, sampling error and population substructure could affect the levels of polymorphism observed at various sites (although this effect would be more likely with small, isolated populations and/or smaller sample sizes). For sites where substitutions were observed, the levels of polymorphism were very similar between our database and another large sequence database from other diverse populations[51], indicating that database bias has not significantly misled our evaluations.

Statistics. Significance of the difference between blood and cell line samples was estimated by Fisher's exact test using the SAS statistical package[52]. A bootstrap confidence interval for estimated substitution rate was calculated with a program by B. Weir.

PCR amplification and DNA sequence analysis of mtDNA. Analysis was performed using two different methods. AFDIL, the FBI and Gettysburg College amplified the two hypervariable regions separately and sequenced the PCR products using the Perkin-Elmer, Applied Biosystems Division (ABD) DyeTerminator™ *Taq* polymerase cycle sequence kit and the ABD 373A DNA Sequencer[28,53]. The British FSS amplified the entire mtDNA CR, followed by amplification of the two hypervariable regions with biotin-labelled primers. Sequencing was performed using the ABD DyePrimer™ Sequenase Kit and the ABD 373A DNA Sequencer[54]. In all cases, sequence was determined from both strands.

In this study, it is important to rule out sequencing error as contributing to the observed substitution rate. The overall quality of sequence determination was demonstrated by replicate analyses in different laboratories for 63 individuals, with complete concordance of results. Additionally, in all cases where substitutions were observed, amplification and sequencing were replicated and gave identical results (for AFDIL reference families where seven of the substitutions were observed, replication included new DNA extractions). Sufficient DNA was used in PCR to ensure amplification starting from vast numbers of template molecules, so it is not plausible that the reproducible results were due to some sort of PCR artifact. Thus, any undetected sequencing error that may have occurred in this study would contribute to an underestimate of the mutation rate.

Acknowledgements

We thank: N. Howell for helpful discussion and generously providing a manuscript prior to publication; S. Pääbo and two anonymous reviewers for helpful comments; S. Cordero for assistance with literature review and replicate sequence determination; J. Holland, J. Ruderman and M. McDermit for sequence analysis and/or database management; C. Ginther for database sequences; B. Sykes for Oxford DNA samples; M. Polymeropolus for CEPH DNA samples; B. Wier for bootstrap analysis; the American Registry of Pathology for funding; and many AFDIL staff members for support and/or contributing to casework sequences. The opinions and assertions expressed herein are solely those of the authors and are not to be construed as official or as the views of the United States Department of Defense or the United States Department of the Army.

Received 10 January; accepted 26 February 1997

© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

1. Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D. & Wilson, A.C. Mitochondrial DNA sequence in single hairs from a southern African population. *Proc. Natl. Acad. Sci. USA* **86**, 9350–9354 (1989).
2. Horai, S. & Hayasaka, K. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* **46**, 828–842 (1990).
3. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A.C. African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507 (1991).
4. Di Rienzo, A. & Wilson, A.C. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **88**, 1597–1601 (1991).
5. Ward, R.H., Frazier, B.L., Dew-Jager, K. & Pääbo, S. Extensive mitochondrial DNA diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* **88**, 8720–8724 (1991).
6. Stoneking, M., Sherry, S.T., Redd, A.J. & Vigilant, L. New approaches to dating suggest a recent age for human mtDNA ancestor. *Phil. Trans. R. Soc. Lond. B* **337**, 167–175 (1992).
7. Horai, S. *et al.* Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol. Biol. Evol.* **10**, 23–47 (1993).
8. Mountain, J.L. *et al.* Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* **56**, 979–992 (1995).
9. Sajantila, A. *et al.* Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* **5**, 42–52 (1995).
10. Graven, L. *et al.* Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol. Biol. Evol.* **12**, 334–345 (1995).
11. Cann, R.L., Stoneking, M. & Wilson, A.C. Mitochondrial DNA and human evolution. *Nature* **325**, 31–36 (1987).
12. Merriwether, D.A. *et al.* The structure of human mitochondrial DNA variation. *J. Mol. Evol.* **33**, 543–555 (1991).
13. Ruvolo, M. A new Approach to Studying Modern Human Origins: Hypothesis Testing with Coalescence Time Distributions. *Mol. Phylo. Evol.* **5**, 202–219 (1996).
14. Chen, Y.-S., Torroni, A., Excoffier, L., Silvana Santachiara-Benerecetti, A. & Wallace, D. C. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J. Hum. Genet.* **57**, 133–149 (1995).
15. Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* **92**, 532–536 (1995).
16. Hasegawa, M. & Horai, S. Time of the deepest root for polymorphism in human mitochondrial DNA. *J. Mol. Evol.* **32**, 37–42 (1991).
17. Horai, S. *et al.* Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* **35**, 32–43 (1992).
18. Pesole, G., Sbisa, E., Preparata, G. & Saccone, C. The evolution of the mitochondrial D-loop region and the origin of modern man. *Mol. Biol. Evol.* **9**, 587–598 (1992).
19. Hassegawa, M., Di Rienzo, A., Kocher, T.D. & Wilson, A.C. Toward a more accurate estimate for the human mitochondrial DNA tree. *J. Mol. Evol.* **37**, 347–354 (1993).
20. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
21. Wakeley, J. Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J. Mol. Evol.* **37**, 613–623 (1993).
22. Wills, C. When did Eve live? An evolutionary detective story. *Evolution* **49**, 593–607 (1995).
23. Ruvolo, M., Zehr, S., von Dornum, M., Pan, D., Chang, B. & Lin, J. Mitocondrial COII sequences and modern human origins. *Mol. Biol. Evol.* **10**, 1115–1135 (1993).
24. Adachi, J. & Hasegawa, M. Tempo and mode of synonymous substitutions in mitochondrial DNA of primates. *Mol. Biol. Evol.* **13**, 200–208 (1996).
25. Ginther, C, Issel-Tarver, L. & King, M.-C. Identifying individuals by sequencing mitochondrial DNA from teeth. *Nature Genet.* **2**, 135–138, (1992).
26. Piercy, R., Sullivan, K.M., Benson, N. & Gill, P. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Leg. Med.* **106**, 85–90 (1993).
27. Holland, M.M. *et al.* Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J. Forensic Sci.* **38**, 542–553, (1993).
28. Holland, M.M. *et al.* Mitochondrial DNA sequence analysis of human remains. *Crime Lab. Digest* **22**, 3–8 (1995).
29. Wilson, M., DiZinno, J.A., Polanskey, D., Replogle, J. & Budowle, B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int. J. Leg. Med.* **108**, 68–74 (1995).
30. Gill, P. *et al.* Identification of the remains of the Romanov family by DNA analysis, *Nature Genet.* **6**, 130–135, (1994).
31. Ivanov, P.L. *et al.* Mitochondrial DNA sequence heteroplasmy in the grand duke of Russia Georgii Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genet.* **12**, 417–420 (1996).
32. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981)
33. Holland, M.M. *et al.* Identification of human remains using mitochondrial DNA sequencing: potential mother-child mutational events. in *Advances in Forensic Haemogenetics* (eds Bar, W., Fiori, A. & Rossi, U.) 399–406 (Springer Verlag, New York, 1994).
34. Nei, M. Age of the common ancestor of human mitochondrial DNA. *Mol. Biol. Evol.* **9**, 1176–1178 (1992).
35. Ayala, F. The myth of Eve: molecular biology and human origins. *Science* **270**, 1930–1936 (1995).
36. Penny, D., Steel, M., Waddel, P.J. & Hendy, M.D. Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Mol. Biol. Evol.* **12**, 863–882 (1995).
37. Howell, N., Kubacka, I. & Mackey, D.A. How rapidly does the human mitochondrial genome evolve? *Am. J. Hum. Genet.* **59**, 501–509 (1996).
38. Nachman, M.W., Brown, W.M., Stoneking, M. & Aquadro, C.F. Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* **142**, 953–963 (1996).
39. Templeton, A.R. Contingency tests of neutrality using inta/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the homonoid primates. *Genetics* **144**, 1263–1270 (1996).
40. Excoffier, L. Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *J. Mol. Evol.* **30**, 125–139 (1990).
41. Gemmel, N.J., Western, P.S., Watson, J.M. & Marshall Graves, J.A. Evolution of the mammalian mitochondrial control region — comparisons of control region sequences between monotreme and therian mammals. *Mol. Biol. Evol.* **13**, 798–808 (1996).
42. Hauswirth W.W. & Laipis, P.J. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proc. Natl. Acad. Sci. USA* **79**, 4686–4690 (1982).
43. Laipis, P., Hauswirth, W., O'Brian, T. & Michaels, G. Unequal partitioning of bovine mitochondrial genotypes among siblings. *Proc. Natl. Acad. Sci. USA* **85**, 8107–8110 (1988).
44. Solignac, M., Genermont, J., Monnerot, M. & Mounolou, J.C. Genetics of mitochondria in *Drosophila*: mtDNA in heteroplasmic strains of *D. mauritiana*. *Mol. Gen. Genet.* **197**, 183–188 (1984).
45. Poulton, J. Transmission of mtDNA: cracks in the bottleneck. *Am. J. Hum. Genet.* **57**, 224–226 (1995).
46. Bendall, K.E., Macaulay, V.A., Baker, J.R. & Sykes, B.C. Heteroplasmic point mutations in the human mtDNA control region. *Am. J. Hum. Genet.* **59**, 1276–1287 (1996).
47. Jenuth, J.P., Peterson, A.C., Fu, K. & Shoubridge, E.A. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nature Genet.* **14**, 146–151 (1996).
48. Comas, D., Pääbo, S. & Bertranpetit, J. Heteroplasmy in the control region of human mitochondrial DNA. *Genome Res.* **5**, 89–90 (1995).
49. Howell, N. *et al.* Mitochondrial gene segregation in mammals: Is the bottleneck always narrow? *Hum. Genet.* **90**, 117–120 (1992).
50. Walsh, P.S., Metzger, D.A. & Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10**, 506–513 (1991).
51. Miller, K.W.P., Dawson, J.L. & Hagelberg, E. A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *Int. J. Legal Med.* **109**, 107–113 (1996).
52. SAS Procedures Guide, release 6.07, SAS Institute, Carey, NC (1991).
53. Wilson, M.R., Stoneking, M., Holland, M. M., DiZinno, J. A. & Budowle, B. Guidelines for the use of mitochondrial DNA sequencing in forensic science, *Crime Lab. Digest* **20**, 68–77 (1993).
54. Sullivan, K.H., Hopgood, R. & Gill, P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int. J. Leg. Med.* **105**, 83–86 (1992).

View publication stats