

A Suggested Genome for “Mitochondrial Eve”

Ian Logan

Abstract

The “Out of Africa” theory has become widely accepted in the 20 years since the original work was done on 147 mitochondrial DNA samples. Now there are more than 3,700 genomes available for study. The human phylogenetic tree is now very complicated. But no sequence has so far been published that includes all the mutations that have arisen in the last 200,000 years. The present article describes 52 mutations that have occurred on the lineage of the Cambridge Reference Sequence (CRS), thereby suggesting a mitochondrial genome for “Mitochondrial Eve.”

Introduction

It is now 20 years since Rebecca Cann, Mark Stoneking and Allan Wilson (Cann, 1987) presented their famous article “Mitochondrial DNA and human evolution” and wrote, “We infer from the tree of minimum length that Africa is the likely source of the human mitochondrial gene pool.” In this way they started what later became known as the “Out of Africa” theory. Initially there was a lot of scepticism concerning the validity of the idea, but now the “Out of Africa” theory is widely accepted (Watson, 1997; Quintana-Murci, 1999; Macaulay, 2005).

The original work was based on the partial sequencing of the mitochondrial DNA (mtDNA) of 147 subjects from Africa, Asia, Australia, Europe and New Guinea. As recently as 7 years ago there were still very few published complete mitochondrial genomes, but today over 3,700 genomes from all over the world have been made available for study. These data have allowed the determination of a phylogenetic tree for mtDNA genomes, and this tree and every genome has been found to be consistent with the “Out of Africa” theory.

As part of the “Out of Africa” theory it is necessary to consider that there was a single matrilineal ancestor for the whole of mankind who lived in Africa around 180,000 to 200,000 years ago. The name “Mitochondrial Eve” has been given to this woman (Richards, 2001).

Mitochondrial DNA

Mitochondria are found in all the nucleated cells of the body and are concerned with the production and trans-

fer of energy within cells and the production of RNA that is involved in the process of making proteins. Inside every mitochondrion there are circular rings of deoxyribonucleic acid (the mtDNA) and each ring is made up of about 16,569 nucleotide bases. These bases are four in number: Adenine, Cytosine, Guanine and Thymine, and for simplicity they are normally represented by their initial letters—A, C, G, T. The actual sequence of mtDNA in the human was first determined in Cambridge in 1981 (Anderson, 1981), and later slightly revised (Andrews, 1999). This initial sequence forms the basis of the Cambridge Reference Sequence (CRS, consider “sequence” being interchangeable with “genome”), against which it is usual to compare all other human mtDNA genomes. Although the mtDNA is circular it is usually considered for discussion purposes as a simple list of bases numbered from 1 to 16,569.

The human mitochondrial genome coding region contains genes for 13 enzymes, two RNA ribosome components (rRNAs) and 22 different transfer RNAs (tRNAs), and some small regions concerned with the replication of the mitochondria. There is also a large non-coding region known as the “Control Region” or the “Hypervariable Regions.” For this discussion the control region will simply be termed “HVR1,” for the part at locations 16024-16569, and “HVR2,” for the part at 1-576.

Mutations

A nucleotide base in the mtDNA may on a rare occasion undergo a mutation; that is, a nucleotide base at a particular place can change. For example: a location that is initially occupied with an Adenine may be filled by a Guanine - such a change at location 263 will here be represented by “A263G.” It is also possible for there to be insertion of extra bases or deletion of bases—these types of mutation are much less common, but arguably more interesting.

Although the non-coding regions make up less than a tenth of the mtDNA, over the past 200,000 years these

Address for correspondence: ianlogan22@btinternet.com

Received: July 12, 2007; accepted: August 30, 2007.

regions have shown a disproportionate number of mutations. This is thought to have occurred because there is no selective pressure on mutations in the non-coding regions, by which is meant that mutations in the non-coding regions are taken to be functionless and harmless to the cell and the person whose mitochondria show such a mutation. However, in the coding regions, mutations will be selected against and will not persist if their effects are harmful to the cell and the person. Mutations that are found in the coding regions and have persisted, usually do not affect the actual amino-acid sequence of the gene product, and in the tRNAs, appear not to compromise the incorporation of the corresponding amino acid during protein synthesis. The effect of a mutation in the coding regions for the ribosomal proteins is still largely unknown.

The Phylogenetic Tree

When the mutations found in human mtDNA genomes are studied it is possible to draw a tree based on the common occurrence of the mutations. This tree is known as a phylogenetic tree. The first phylogenetic tree was presented in "Mitochondrial DNA and human evolution" (Cann, 1987).

Since then the different branches of the phylogenetic tree have been considered as denoting "haplogroups" and given labels, from A-Z. For example, the CRS genome is now considered to be in Haplogroup H.

Over the past 20 years, the phylogenetic tree has been greatly expanded and is now very complicated. The gradual change in its development can be followed by looking through the published papers from Maca-Meyer (2001), Herrnstadt (2002), Mishmar (2003), Kivisild (2006), Torroni (2006), Ruiz-Pesini (2007). However, despite trees of ever increasing size being produced, there does not appear to any single tree that includes all the mutations that have occurred since Mitochondrial Eve along the line leading to the CRS.

For the purposes of this paper a simplified phylogenetic tree is shown in Figure 1.

All mtDNA genomes now show an average of about 50 mutations which have occurred in the 200,000 years since Mitochondrial Eve. This paper discusses the mutations that have occurred in the CRS lineage over this period of time. In so doing a genome for Mitochondrial Eve is suggested here for the first time.

Methods

The main source for mtDNA genomes is "The National Centre of Biotechnology Information" (NCBI) where over 3,700 complete human mitochondrial genomes are now available for study at the "Entrez Nucleotide" website: <http://www.ncbi.nih.gov/entrez/query.fcgi>

Each mtDNA genome can be viewed by entering the appropriate accession number; for example "EU157923" gives the latest genome to be made available (as of September 2007).

Information on the structure of parts of the phylogenetic tree has been taken from various papers, in particular Herrnstadt (2002), Mishmar (2003), Kivisild (2006), Torroni (2006), Ruiz-Pesini (2007). However in most of the papers the example trees have been drawn using only the mutations found in the coding region. For the purpose of this paper the author has used his own computer programs to determine all the mutations present in the 3,700 genomes that are presently available and has thereby been able to build a phylogenetic tree based on mutations from both the coding and non-coding regions of the mitochondrial DNA.

Results and Discussion

A suggested mutation list back to Mitochondrial Eve

The Cambridge Reference Sequence (CRS) is the genome of a British woman belonging to Haplogroup H2. By definition the CRS does not contain any differences or mutations from itself, but as one looks back in time towards Mitochondrial Eve it is possible to identify about 50 mutations that have occurred.

For convenience, the mutations are considered here in two parts: firstly the mutations that have occurred in approximately the last 60,000 years, and secondly, the mutations which occurred in the previous 140,000 years.

Mutations occurring in the last 60,000 years

20 mutations appear to have occurred in approximately the last 60,000 years. These mutations are now well accepted and are detailed on many of the published phylogenetic trees. These mutations can be considered as being those mutations that have occurred since Homo Sapiens first left Africa. It follows that all African genomes, i.e. all genomes in the "L" Haplogroups can be expected to show all of these mutations as differences from the CRS. Of course, the occasional mutation might not be present in a particular genome because of a further mutation returning the CRS value.

In the following discussion, we will start at CRS and work our way back in time. This approach is, of course, opposite from the progression that actually occurred. Where specific mutations are mentioned, the leading letter is that present in CRS and the letter following the location is the ancestral state. This approach has the advantage that this is the way that each mutation would be stated in a list of differences from CRS.

The most recent eight mutations on the line leading to CRS, and the first we encounter going back in time are:

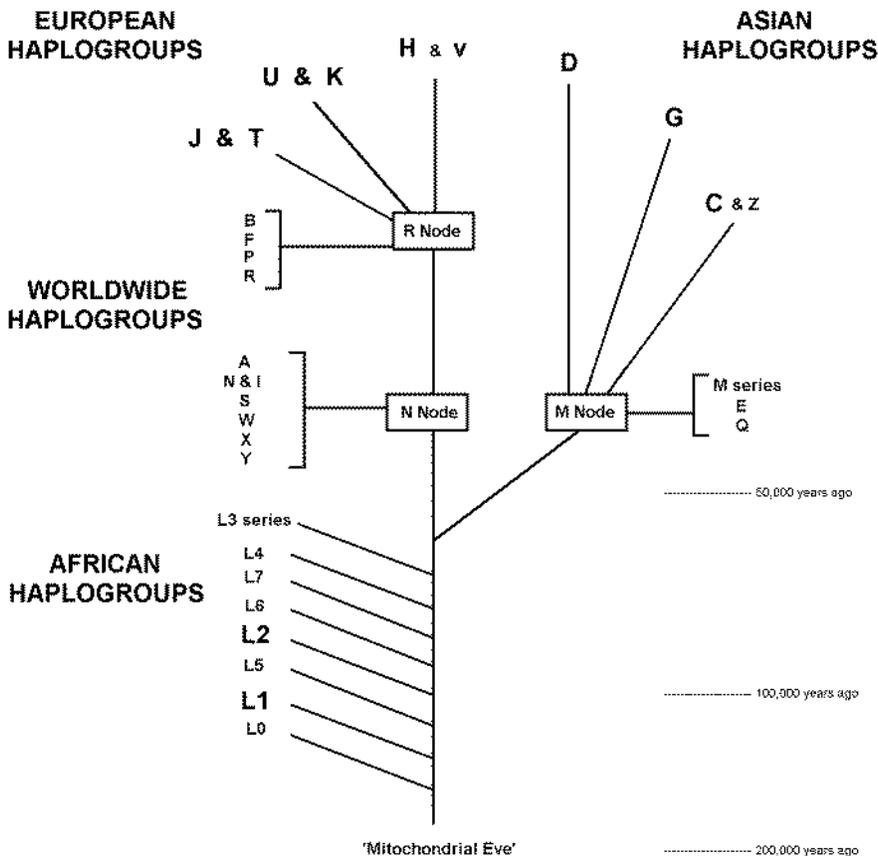


Figure 1. A Simplified Phylogenetic Tree. The largest haplogroups are L1, L2, J, T, U, K, H, D, G, and C

A750G 315.1C A4769G A1438G
 A15326G A8860G 309.1C A263G

C7028T A2706G C14766T G11719A A73G

These mutations all occur within Haplogroup H. The two insertions 309.1C and 315.1C mean that the CRS is two bases shorter than many other genomes.

The two mutations:

C16223T C12705T

The area of the genome from 303-315 is largely made up of C's, and is termed a Poly-C area. This area is very variable and can even be different between relatives.

come between the "N" node and the "R" node (see Figure 1), effectively separating the major European haplogroups from the remainder of the tree.

There is a further variable area, at 514-523, which in the CRS has five repetitions of CA. The author suggests that this particular pattern has been preserved in the line leading to CRS ever since Mitochondrial Eve, as five repetitions are typically found in the majority of African and Asian genomes.

It is interesting to note that despite being in the HVR1 area, the mutation at 16223 appears very stable and is therefore extremely useful in determining if a genome belongs above the "N" node, or elsewhere in the tree.

Separating Haplogroups H and V from the "R" node are 5 mutations (see Figure 1):

G15301A T10873C A10398G T9540C A8701G

and they come between the major forking to the Asian

The presence of so many mutations in just this small area of the phylogenetic tree indicates a significant *bottleneck* in the spread of mankind and implies that the population at this period of time outside Africa was extremely low. Presumably, there were people living during this period who had some of these mutations, but not all five. However, none of their descendants have survived, except for that line having all five.

The 20 mutations that have occurred in the line leading to CRS during the 60,000 years since humankind left Africa are reviewed in Table 1, together with comments about their function and location.

Mutations occurring in the previous 140,000 years

It is fairly easy to give a firm list of mutations that have occurred over the last 60,000 years because there are many genomes available for the European and Asian haplogroups. However, for the period of 140,000 years closer to Mitochondrial Eve, it is not possible to be so confident as there are far fewer published genomes. Indeed the list suggested here may need to be revised as new genomes are published.

32 mutations appear to have occurred in the approximate period of 140,000 years back closer to Mitochondrial Eve.

The most recent mutations we encounter are:

G1018A G769A

between the branches leading to Haplogroups L4 and L7 and below the series of L3 haplogroups.

The next five mutations come between the branch leading to Haplogroup L6 and the branches leading to Haplogroups L4 and L7:

C16278T C13650T C7256T C3594T T152C

Between the branches leading to Haplogroups L2 and L6 are two mutations:

G7521A A4104G

Between the branches to Haplogroup L5 and Haplogroup L2 there are 12 mutations:

T16519C T16311C T16189C C16187T
A15301G C13506T A13105G T10810C
G10688A C8655T T825A G247A

Once again, the high number of mutations suggests there was a significant *bottleneck* in human evolution at the time, perhaps around 120,000 years ago, which might have lasted for many thousands of years.

Note that this is the second time a mutation at location 15301 has occurred, which means that genomes beyond this point have the CRS value.

Between the branches leading to Haplogroups L1 and L5 there are a further four mutations:

Table 1. The Mutation List Covering the Last 60,000 Years

| <u>Mutation</u> | <u>Function</u> | <u>Position in the “Phylogenetic Tree”</u> |
|-----------------|--------------------|---|
| A750G | 12S-rRNA* | mutation used to define Haplogroup H2a |
| 315.1C | HVR2 | mutation within Haplogroup H2 |
| A4769G | ND2 (Met > Met) | mutation within Haplogroup H2 |
| A1438G | 12S-rRNA* | mutation defines Haplogroup H2 |
| A15326G | CytB (Thr > Ala) | mutation within Haplogroup H |
| A8860G | ATP6 (Thr > Ala) * | mutation within Haplogroup H |
| 309.1C | HVR2 | mutation within Haplogroup H |
| A263G | HVR2 | mutation within Haplogroup H |
| C7028T | COX1 (Ala > Ala) | mutation used to define Haplogroup H |
| A2706G | 16S-rRNA | mutation used to define Haplogroup H |
| C14766T | CytB (Ile > Thr) * | mutation used to define Haplogroups H and V |
| G11719A | ND4 (Gly > Gly) | mutation used to define Haplogroup pre-HV |
| A73G | HVR2 | mutation used to define Haplogroup pre-HV |
| C16223T | HVR1 | between the N and R nodes |
| C12705T | ND5 (Ile > Ile) | between the N and R nodes |
| G15301A | CytB (Leu > Leu) | between “L” haplogroups and N node |
| T10873C | ND4 (Pro > Pro) | between “L” haplogroups and N node |
| A10398G | ND4 (Thr > Ala) * | between “L” haplogroups and N node |
| T9540C | COX2 (Leu > Leu) | between “L” haplogroups and N node |
| A8701G | ATP6 (Thr > Ala) * | between “L” haplogroups and N node |

Note: All mutations in the HVR1 and HVR2 areas are considered functionless, but the effects of the mutations marked * are unknown.

C8468T A7146G T2885C G2758A

The last seven mutations come between Mitochondrial Eve and the branch to Haplogroup L0:

A16230G G12007A G11914A G9755A
T6185C C4312T C1048T

These differences from the CRS are all found in Haplogroup L0, but as they also occur in the chimpanzee (*Pan troglodytes*) mtDNA, this suggests that they are mutations on the main ancestral line. There are many other mutations found in Haplogroup L0 which are not to be found in chimpanzee mtDNA and it is presumed these are not therefore on the direct line of human evolution.

The 32 mutations from the period 200,000-60,000 years before present are reviewed in Table 2, together with comment about their location and function.

Conclusion

A suggested mitochondrial genome for Mitochondrial Eve is therefore the Cambridge Reference Sequence (CRS) modified by the 50 mutations given below—accepting 15301 as having mutated back to its CRS value.

| | | | |
|---------|---------|---------|---------|
| A73G | T152C | G247A | A263G |
| 309.1C | 315.1C | A750G | G769A |
| T825A | G1018A | C1048T | A1438G |
| A2706G | G2758A | T2885C | C3594T |
| A4104G | C4312T | A4769G | T6185C |
| C7028T | A7146G | C7256T | G7521A |
| C8468T | C8655T | A8701G | A8860G |
| T9540C | G9755A | A10398G | G10688A |
| T10810C | T10873C | G11719A | G11914A |
| G12007A | C12705T | A13105G | C13506T |
| C13650T | C14766T | A15326G | C16187T |
| T16189C | C16223T | A16230G | C16278T |
| T16311C | T16519C | | |

The actual sequence is available in a supplementary text data file which accompanies this paper.¹

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) [Sequence and organization of the human mitochondrial genome.](#) *Nature*, 290:457-465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) [Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.](#) *Nature Genet*, 23:147.

Cann RL, Stoneking M, Wilson AC (1987) [Mitochondrial DNA and human evolution.](#) *Nature*, 325:31-36.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell, N (2002) [Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups.](#) *Am J Hum Genet*, 70:1152-1171.

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ (2006) [The role of selection in the evolution of human mitochondrial genomes.](#) *Genetics*, 172:373-387.

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) [Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa.](#) *Nature Genet*, 23:437-441.

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) [Major genomic mitochondrial lineages delineate early human expansions.](#) *BMC Genet*, 2:13.

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M (2005) [Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes.](#) *Science*, 13:1034-1036.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) [Natural selection shaped regional mtDNA variation in humans.](#) *Proc Natl Acad Sci*, 100:171-176.

Richards M, Macaulay V (2001) [The mitochondrial gene tree comes of age.](#) *Am J Hum Genet*, 68:1315-1320.

Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, Kreuziger J, Baldi P, Wallace DC (2007) [An enhanced MITOMAP with a global mtDNA mutational phylogeny.](#) *Nucleic Acids Res*, 35(Database issue):D823-8

Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) [Harvesting the fruit of the human mtDNA tree.](#) *Trends Genet*, 22:339-345.

Watson E, Forster P, Richards M, Bandelt HJ (1997) [Mitochondrial footprints of human expansions in Africa.](#) *Am J Hum Genet*, 61:691-704.

¹ Editor's Note: The referenced supplementary text file contains one long string of bases without any internal reference points. For a tabular version comparing CRS with mtEve, with CRS sequence locations shown every ten bases, see: <http://www.jogg.info/ref/indexref.htm>.

Table 2. The Mutation List from 60,000- 200,000 Years Ago

| <u>Mutation</u> | <u>Function</u> | <u>Position in the Phylogenetic Tree</u> |
|-----------------|------------------------------|---|
| G1018A | 12S-rRNA | between Haplogroup L4 and the L3 series |
| G769A | 12S-rRNA | between Haplogroup L4 and the L3 series |
| C16278T | HVR1 | between Haplogroup L6 and Haplogroups L4 and L7 |
| C13650T | ND5 (Pro > Pro) | between Haplogroup L6 and Haplogroups L4 and L7 |
| C7256T | COX1 (Asn > Asn) | between Haplogroup L6 and Haplogroups L4 and L7 |
| C3594T | ND1 (Val > Val) | between Haplogroup L6 and Haplogroups L4 and L7 |
| T152C | HVR2 | between Haplogroup L6 and Haplogroups L4 and L7 |
| G7521A | tRNA Asp | between Haplogroup L2 and Haplogroup L6 |
| A4104G | ND1 (Leu > Leu) | between Haplogroup L2 and Haplogroup L6 |
| T16519C | HVR1 | between Haplogroup L5 and Haplogroup L2 |
| T16311C | HVR1 | between Haplogroup L5 and Haplogroup L2 |
| T16189C | HVR1 | between Haplogroup L5 and Haplogroup L2 |
| C16187T | HVR1 | between Haplogroup L5 and Haplogroup L2 |
| A15301G | CytB (Leu > Leu) (CRS value) | between Haplogroup L5 and Haplogroup L2 |
| C13506T | ND5 (Tyr > Tyr) | between Haplogroup L5 and Haplogroup L2 |
| A13105G | ND5 (Ile > Val) | between Haplogroup L5 and Haplogroup L2 |
| T10810C | ND4 (Leu > Leu) | between Haplogroup L5 and Haplogroup L2 |
| G10688A | ND4 (Val > Val) | between Haplogroup L5 and Haplogroup L2 |
| C8655T | ATP6 (Ile > Ile) | between Haplogroup L5 and Haplogroup L2 |
| T825A | 12S-rRNA | between Haplogroup L5 and Haplogroup L2 |
| G247A | HVR2 | between Haplogroup L5 and Haplogroup L2 |
| C8468T | ATP8 (Leu > Leu) | between Haplogroup L1 and Haplogroup L5 |
| A7146G | COX1 (Thr > Ala) | between Haplogroup L1 and Haplogroup L5 |
| T2885C | 16S-rRNA | between Haplogroup L1 and Haplogroup L5 |
| G2758A | 16S-rRNA | between Haplogroup L1 and Haplogroup L5 |
| A16230G | HVR1 | present in Haplogroup L0 and the chimpanzee |
| G12007A | ND4 (Trp > Trp) | present in Haplogroup L0 and the chimpanzee |
| G11914A | ND4 (Thr > Thr) | present in Haplogroup L0 and the chimpanzee |
| G9755A | COX2 (Glu > Glu) | present in Haplogroup L0 and the chimpanzee |
| T6185C | COX1 (Phe > Phe) | present in Haplogroup L0 and the chimpanzee |
| C4312T | tRNA Ile | present in Haplogroup L0 and the chimpanzee |
| C1048T | 12S-rRNA | present in Haplogroup L0 and the chimpanzee |

Note: The last 7 mutations are all found in Haplogroup L0. But, as they also occur in the chimpanzee (Pan troglodytes) mtDNA, this suggests that they are mutations on the main ancestral line. There are other mutations found in Haplogroup L0 which are not to be found in chimpanzee mtDNA.