THE MEDICAL IMPLICATIONS OF COMPLETE MITOCHONDRIAL DNA SEQUENCING

Author(s): Ian Logan

The Medical Implications of Complete Mitochondrial DNA Sequencing

Ian Logan

The mitochondrial DNA sequencing for the genealogical community until very recently was only of parts of the genome considered to be medically unimportant; however anyone can now have a test that will provide the complete mitochondrial DNA (mtDNA) sequence. But the results of this test may have medical implications; and in this article the factors showing this are discussed. The subjects of *mitochondrial science, mutations, mutation lists, protein coding* and *transfer RNA* are introduced to provide a background to understanding the origin of various *mitochondrial diseases*. The condition of *Leber's Hereditary Optic Neuropathy* and the diseases associated with mutations in the tRNA for the amino acid, isoleucine, are used as examples. Two complete mtDNA sequences are analysed to show how a report on their mutations can be produced; and the risk to a subject of getting a report listing potentially harmful mutations is discussed. Mention is made of various instances when complete mitochondrial sequencing may be requested under the guise of genealogy when it really is being done for medical reasons. Complete mitochondrial DNA sequencing should not be undertaken without careful consideration.

Introduction

The sequencing of mitochondrial DNA (mtDNA) has become very popular with both amateur and professional genealogists over the last few years. This subject has been made particularly interesting by Professor Bryan Sykes's book, *The Seven Daughters of Eve* (Sykes, 2001) in which he discusses the concept that all people of European descent have as their female ancestors just a few women who lived many thousands of years ago. His ideas are part of the general theory that all of mankind is descended from a "*Mitochondrial Eve*" who lived in Africa about 180,000 years ago.

However, the mtDNA sequencing that has been done until recently was only of parts of the genome considered to be medically unimportant; by which is meant that there was no proven connection between the results obtained and known illnesses. Therefore, sequencing of mtDNA was considered as harmless to the person being tested and there was no reason for anyone to have any hesitation over being tested. However, it is now possible to sequence the whole of the mtDNA genome, rather than just 4-8%; and the medical implications are beginning to give concern.

Many medical conditions are already known to be caused by changes in the mtDNA and the whole area of knowledge continues to expand rapidly. So now it is perhaps not a question of having *abnormal* mtDNA as opposed to *normal*, but rather a matter of *what harm* is being caused to a person's health by the particular mtDNA they have inherited.

It is therefore no longer possible to say that sequencing the mtDNA genome is a harmless procedure; and in this article the factors that show this are discussed. However, because the underlying science may be new to many readers who do not have a background in a science, there are a number of basic scientific details that have to be explained.

2. Mitochondrial Science

Each mature nucleated cell of the body contains several hundreds of mitochondria (Taylor 2005). These are small structures involved in energy production within the cell. A mitochondrion may be considered as a sort of *power plant,* with fuel being taken in and energy produced. In each of the mitochondria there are circular strands of genetic material made up of deoxyribonucleic acid. This is the mitochondrial DNA, or mtDNA. The structure of one of these strands was first determined in 1981 at the University of Cambridge (Anderson 1981). This structure can be called the mtDNA genome, or the mtDNA sequence. **Figure 1** shows a simplified view of mtDNA.
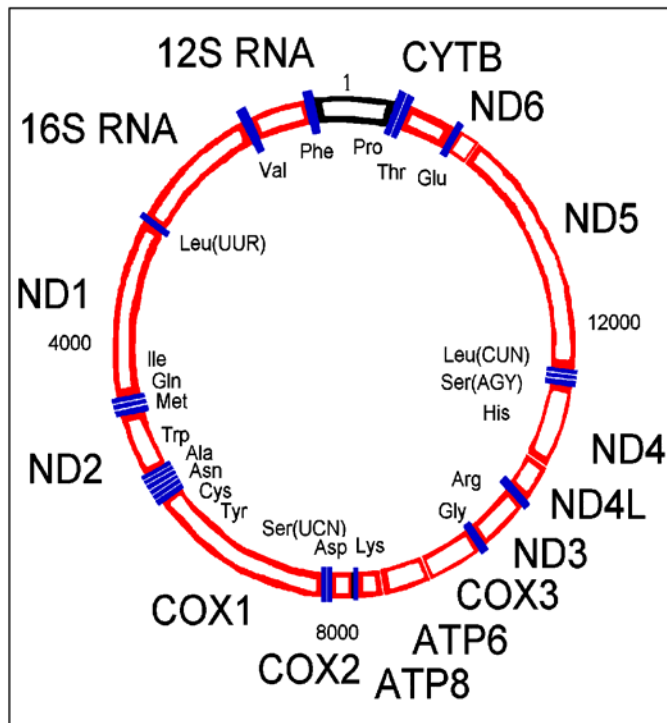
Figure 1 A Simplified View of mtDNA.
(Adapted from: http://www.mitomap.org/) The circular strand of mtDNA is shown with nucleotide base 1 at the 12 o'clock position. By convention, the numbering of the nucleotide bases goes in an anti-clockwise direction. The coding areas for the genes are shown *in red,* with their corresponding labels outside (see **Table 6** for the full names). The tRNA's are shown *in blue* with the three-letter abbreviations for the amino acids inside (see **Table 7** for the full names).

The original subject was a person of European descent and her sequence now forms the *Cambridge Reference Sequence* (CRS). In this sequence the nucleotide bases are numbered from 1 to 16,569; and subsequently, all other mtDNA sequences have been compared to the CRS. Differences between sequences are normally given as a simple list.

The purpose of the mtDNA is to carry encoded instructions for certain products in the cell. In it, there are coding regions for two ribosomal RNA genes, 13 genes for the production of proteins involved in *oxidative phosphorylation* of fatty acids and carbohydrates (Taylor 2005), and, perhaps most interestingly in the present context, the coding for 22 transfer RNA (tRNA) genes for the 20 amino acids that are the building blocks of proteins. There are also parts of the mtDNA that appear to have no function, other than perhaps to ensure the correct shape to the overall circular form of the mtDNA. The largest of these non-

functional areas contains the *hypervariable regions*, HVR-1 and HVR-2.

The actions of mitochondria are mainly concerned with the energy processes within cells; and whilst a single mitochondrion is very small and has an insignificant influence on any particular cell, the cumulative effect of mitochondria on health and illness in a person can become clear when one considers that about 2 billion mitochondria are made every second throughout a person's life. This estimate is obtained by considering the human body as having 75 trillion cells, each with 250 mitochondria and having an average life of around 100 days.

Until recently sequencing of the mtDNA for genealogical purposes was restricted to the *hypervariable regions*; and many thousands of subjects have already been tested. But now it is possible to sequence the whole mitochondrial genome for any individual at a reasonable cost; in addition, already about 2,500 complete mtDNA sequences have been published from scientific studies. It is the data from these published complete mtDNA sequences that provide the material for this article.

3. Mitochondrial Nucleotides

The mtDNA genome is considered as a circular strand of DNA consisting of 16,569 nucleotide bases and for historical reasons the numbering of the bases begins at an arbitrary place—the start of what is termed "the hypervariable HVR-2 region."

As in other DNA, there are 4 nucleotide bases, Adenine, Guanine, Cytosine and Thymine; and from now on these will be referred to by their initial letters, i.e. A, G, C and T. The actual mtDNA is made up of a double strand, but for simplicity only one strand is normally shown. The second strand is *complementary* by which is meant that a 'T' links with an 'A', an 'A' with a 'T', a 'C' with a 'G' and a 'G' with a 'C'. The pairings of A-T and C-G are known as "*Watson-Crick base pairings*" after the discoverers of the structure of DNA.

It is a feature of mtDNA that some of the coding is read on one strand going forward in one direction, whilst other coding is read from the *complementary* strand in the opposite direction.

It is not practical to print the full sequence of 16,569 bases that makes up the Cambridge Reference Sequence (CRS) in this article, but Table 1 is included to show the first 70 nucleotide bases. The entire sequence may be found at:

http://www.mitomap.org/mitoseq.htm

Table 1
First 70 Bases of the Cambridge Reference Sequence.

The bases are numbered from 1 to 70 and form the first part of the hypervariable region HVR-2. The sequence is often split into groups of 10 bases to aid counting.

gatcacaggt ctatcaccct attaaccact cacgggagct ctccatgcat ttggtatttt cgtctgggggg
1   11   21   31   41   51   61

---

4. Mitochondrial Mutations

mtDNA sequences from persons who are maternally related to each other are normally identical; but it is found that once in about every 2,000 years in a given lineage, one of the 16569 nucleotide bases alters. This is a random process. Such an alteration is termed a *mutation.*

In most instances the change is of a "C" being replaced by a "T", or vice-versa, or a "A" being replaced by a "G", or vice-versa, and uncommonly the remaining possibilities. It is also possible for nucleotide bases to be added—these are called *insertions*, or for bases to be lost—these are called *deletions*.

Mutations cause the mtDNA sequences from persons who are not closely related to be different. And, it follows that the number and type of mutations between two persons gives a measure of their *relatedness*. Consequently, two persons who are both European, for example, will have fewer mutational differences as when compared with persons who are from very different parts of the world.

To illustrate this point, **Table 2** shows a typical list of differences from the CRS. The results come from an Italian (Achilli 2004). The list here is short as the subject is being compared against the CRS individual who is considered to come from a more northerly part of the same continent, Europe. The length of the list suggests a common matrilineal ancestor perhaps 20-30,000 years ago.

Table 3 shows the list of differences from CRS for a person from Uganda (Macaulay 2005). In this case the mutation list is very much longer; and supports the theory that the earliest African lineage, Haplogroup L0, and Europeans, as typified by the CRS, have been separate for approximately 180,000 years.

Kivisild (2005) gives the figure of 160,000 (range 138,000 - 182,000) years, based on their study of 277 complete mtDNA sequences. But in the author's view the higher figure is to be preferred, and is supported by other studies (Macauley 2005).

5. Phylogenetic trees

By studying the mutations that are found in different sequences it is possible to draw a *tree* to show how different populations have branched away from the line of mutations that leads from '*mitochondrial Eve*' to the European person whose sequence has been chosen as the CRS.

Each of these populations has been given a *label* called a *haplogroup*; and the CRS sequence belongs to Haplogroup H2, the Italian sequence, above, to Haplogroup H9 and the African sequence, above, to Haplogroup L0.

Table 4 shows the probable mutations from "*mitochondrial Eve*" (mtEve) to the CRS. All the mutation sites are to be found in the Ugandan's mutation

---

Table 2
A Typical "Mutation" List for a European

The mtDNA from an Italian subject shows differences from the CRS at the following places:

   T152C   A263G   315.C   A750G   A1438G   G3591A   A4310G   A4769G   A8860G
   T9148C   T13020C   A15326G   C16168T   T16519C

Note: The T152C shows that the 'T' at position 152 in the CRS is replaced by 'C' and the 315.C shows there is an insertion of a 'C' between 315 & 316.

list, given above, as Haplogroup L0 is the first population that branches off. However, the Ugandan subject's list of differences from CRS has twice as many mutations as appear in Table 4 because his list must include not only his mutations from mtEve, but also the reverse of the mutations between mtEve and CRS.

It is important, however, to appreciate that if the sequence chosen as the CRS had been the African sequence, as given above, European sequences would be seen as very different to the *reference sequence* - it is just a matter of history, and with hindsight perhaps the use of the hypothetical reference sequence for '*mitochondrial Eve'* would have been best from all points of view. However, in the remainder of this article, the normal convention—representing a sequence as a set of differences from CRS—will be followed. Using this convention, the word "mutation" will mean difference from CRS.

6. The Genes Encoded in the mtDNA

The mtDNA genome has 15 areas of coding for genes. There are two genes which are involved in ribosomal function (ribosomes are other small structures in a cell) and 13 genes concerned with the the biochemical process of *oxidative phosphorylation*.

The two ribosomal genes code for *ribosomal RNA*. In this article the function of the genes will not be discussed further as there does not appear to be any definite medical condition associated with any of the many known mutations in these areas of coding. Presumably these mutations do not affect overall ribosomal function to any significant degree. The mutations are, however, useful when constructing *phylogenetic trees*. **Table 5** gives the details of these genes.

However, the 13 genes that encode for proteins involved in *oxidative phosphorylation* are very important as they are involved with many medical conditions. These genes are listed in **Table 6.**

These genes all code for protein using a triplet-codon system, which means that a group of 3 consecutive nucleotides identify an amino acid for the protein.

To illustrate this, 18 nucleotide bases from near the start of the gene ND1 (beginning at 3331) are shown with the corresponding amino acid coded by each triplet of bases:

CTC	ATT	GTA	CCC	ATT	CTA	....
Leu	Ile	Val	Pro	Ile	Leu	....

Table 7 shows the full details of the triplet-codon system.

But what is of interest here is the way in which mutations can affect the coding. A mutation may lead to the gene encoding a different amino acid at a particular place, with the potential for impairing the protein's function. However, it is also possible that the effect of a mutation can be *neutral,* as in the case where a mutated codon codes for the same amino acid.

To show this here are three examples of mutations occurring in the coding areas for genes:

First, a mutation that does effect a change:

The mutation A15326G alters the 1st nucleotide base of the triplet-codon for the 193rd. amino acid of the gene CYT-B. At locations 15326-15328 in the CRS is ACA, which is a triplet code for threonine, and the mutation A15326G changes the triplet code to GCA and the amino acid to alanine. This is usually shown as A15326G …. Thr > Ala to indicate the amino acid changes.

Secondly, a *neutral* mutation:

G15301A alters the third nucleotide base of the triplet-codon for the 184th amino acid of the gene for CYT-B. Locations 15299-301 in the CRS has TTG, a triplet-codon for leucine, and the mutation G15301A changes this to TTT, which is also a triplet-codon for the amino acid leucine. This can be written as G15301 .. Leu >Leu to indicate the amino acid does not change.

And thirdly, a mutation of medical interest:

The mutation T14484C is associated with a form of blindness (discussed below) and has an occurrence of no more than 0.01%. T14484 alters the first base of the triplet code for the 64th amino acid of the protein gene for ND6. But note this gene is read from the complementary strand.

Locations 14484, 14483 & 14482 contain TAC, which when complemented on the other strand becomes ATG, a triplet-codon for methionine. The mutation T14484C changes the triplet-codon to CAC, and the complement of this into GTG, which is a triplet code for valine. This can be written as T14484C …. Met > Val.

More examples of the effects of mutations in the coding areas for the genes will be given later.

Table 3
A More Complicated "Mutation" List

The mtDNA from an Ugandan subject, from Haplogroup L0, shows differences from the CRS at the following places:

G143A	T146C	T152C	G185A	A189G	G247A	A263G	315.C	A750G
G769A	825A	A978G	G1018A	C1048T	A1438G	G1719A	A2245G	A2706G
G2758A	C2772T	C2789T	T2885C	C3107T	C3516A	C3594T	C3852T	A4104G
C4194T	C4312T	A4562G	T4586C	A4769G	C4964T	C5321T	T5442C	C5603T
T6185C	A6359G	C7028T	A7146G	T7148C	C7256T	G7521A	C8468T	C8655T
A8701G	A8860G	C9042T	A9347G	T9540C	T9581C	C9620T	C9818T	G10143A
A10398G	G10589A	C10664T	G10688A	T10790C	T10810C	T10873C	T10915C	T11287C
T11299C	A11641G	G11719A	G11914A	G12007A	C12705T	A13105G	A13276G	A13470G
C13506T	C13650T	C13680T	G13708A	G13928A	C14109T	C14620T	C14766T	C15136T
A15326G	G15431A	T15852C	G16129A	C16169T	T16172C	C16173T	C16187T	T16189C
C16223T	A16230G	16239T	C16278T	T16311C	C16327T	T16368C	T16519C	

Table 4
Mutation Sites from "Mitochondrial Eve" to the CRS

(Omitting hypervariable areas)

Line from Branch
Point to Branch Point*   Mutations Along this Line at these Locations

Line from Branch Point to Branch Point*	Mutations Along this Line at these Locations						
From mtEve to BP-to-L1	4312	6185	9755	10589	11914	12007	
From BP-to-L1 to BP-to-L5	2758	2885	7146	8468			
From BP-to-L5 to BP-to-L2	825	8655	10688	10810	13105	13506	15301
From BP-to-L2 to BP-to-L4	3594	4104	7256	7521	13650		
From BP-to-L4 to L3	769	1018					
From L3 to N	8701	9540	10398	10873	15301		
From N to R	12705						
From R to HV	11719	14766					
From HV to CRS	750	2706	1438	4769	7028	8860	15326

* "Branch Point" or "BP" means a point on the line in the mtDNA phylogenetic tree between mtEve and CRS, where a branch leads to another named haplogroup. These branch points are shown in Figure 1 in the article by Kivisild (2005). A detailed discussion of the reasons for including each of the mutations given above will be the subject of a forthcoming article in JoGG.

Note: 15301 has changed twice.

Table 5 The Names and Locations of the Ribosomal Genes Found in mtDNA.
Adapted from: http://www.mitomap.org/

Name:	Location
12S ribosomal RNA	648 – 1601
16S ribosomal RNA	1671 – 3229

Table 6
The Names and Locations of the *Oxidative Phoshorylation* Enzyme Genes Found in mtDNA.
(Adapted from: http://www.mitomap.org/)

Name: Location:

NADH dehydrogenase subunit 1 (ND1)   3307 - 4263
NADH dehydrogenase subunit 2 (ND2)   4470 - 5513
NADH dehydrogenase subunit 3 (ND3)   10059 - 10404
NADH dehydrogenase subunit 4L (ND4L)   10470 - 10766
NADH dehydrogenase subunit 4 (ND4)   10760 - 12137
NADH dehydrogenase subunit 5 (ND5)   12337 - 14148
NADH dehydrogenase subunit 6 (ND6)   14149 - 14673 (Complement)
cytochrome c oxidase subunit I (COX1)   5904 - 7445
cytochrome c oxidase subunit II (COX2)   7586 - 8269
cytochrome c oxidase III (COX3)   9207 - 9987
ATP synthase F0 subunit 6 (ATP6)   8527 - 9207
ATP synthase F0 subunit 8 (ATP8)   8366 - 8572
cytochrome b (CYTB)   14747 - 15886

Note: Because of post-transcription processing of the last few bases in the RNA copy of the DNA, the number of locations is not always a multiple of three (see also Note 2 to **Table 7**).

Table 7
The Triplet-Codon System Used in Genes to Code for Amino Acids

Amino Acid	Triplet-codons			Amino Acid	Triplet-codons		
Alanine(Ala)	GC*			Leucine(Leu)	TTA	TTG	CT*
Arginine(Arg)	AGA	AGG	CG*	Lysine(Lys)	AAA	AAG	
Asparagine(Asn)	AAC	AAT		Methionine(Met)	ATA	ATG	ATT
Aspartic Acid(Asp)	GAC	GAT		Phenylalanine(Phe)	TTC	TTT	
Cysteine(Cys)	TGC	TGT		Proline(Pro)	CC*		
Glutamic acid(Glu)	GAA	GAG		Serine(Ser)	AGC	AGT	TC*
Glutamine(Gln)	CAA	CAG		Threonine(Thr)	AC*		
Glycine(Gly)	GG*			Tryptophan(Trp)	TGG		
Histidine(His)	CAC	CAT		Tyrosine(Tyr)	TAC	TAT	
Isoleucine(Ile)	ATC			Valine(Val)	GT*		

And the beginning and end of coding areas usually have one of the triplet-codons:
START   ATA   ATG   ATT   STOP   TAA   TAG   TGA

Note (1): An asterisk (*) indicates any of A, C, G, or T.
Note (2): For simplicity, 'T' is used here as this is the nucleotide in mtDNA, although a uracil (U) is used in the RNA that is transcribed. Also note that mitochondria use a slightly different genetic code from nuclear DNA—ATA and ATT code for methionine and START instead of the normal isoleucine. The STOP codons are sometimes produced only in post-transcriptional processing, so some mtDNA genes will not have a normal STOP code at the end location of the gene.

7. The *transfer RNA* coding areas

In a cell a protein is assembled from its constituent amino acids according to the order specified by the gene for that protein. It is therefore necessary for the cell to have an adequate supply of each of the different amino acids available to be used. However, there also needs to be a method of handling the amino acid molecules; and the system that has evolved to do this involves the use of small pieces of ribonucleic acid (RNA). There is a slightly differently coded piece of RNA for each amino acid and these pieces of RNA are called the *transfer RNAs* (tRNA).

The structure of each of the *transfer RNAs* is very similar one to another and when the pieces are free within the cell they all take up a *cloverleaf* shape. The *stem* and *side leaves* of a *transfer RNA* binds to its appropriate amino acid whilst the *end* of the middle *leaf* is termed the *anti-codon,* as it is this area that binds with a gene's triplet-codon. For example, the *anti-codon* for isoleucine is GAT, which once complemented and reversed becomes ATC, which is a triplet-codon for isoleucine.

Table 8 gives three examples of the sequences that code for tRNAs. Each is aligned to show the structural similarities. The anti-codon areas are shown in bold and underlined. For interest a sequence from another mammalian species is also shown. In the mtDNA there are 22 areas that code for the tRNAs; and the details are given in **Table 9.**

Drawn in **Figure 2** is the *cloverleaf* shape for just one tRNA, that for isoleucine, using the sequence from the DNA instead of the corresponding RNA. Note that,

except for just one instance, the *stems* of the *leaf* show *Watson-Crick* pairings, ie., with A-T and C-G pairings. It is thought that these pairings in the *stems* add to the stability of the structure. The sequence coding for this tRNA is at 4263-4331 in the mtDNA. The three bases at 4292-4294 are termed the *anti-codon*. The 'GAT', once complemented and reversed, gives the triplet-codon, 'ATC,' which codes for isoleucine (as shown in **Table 7**).
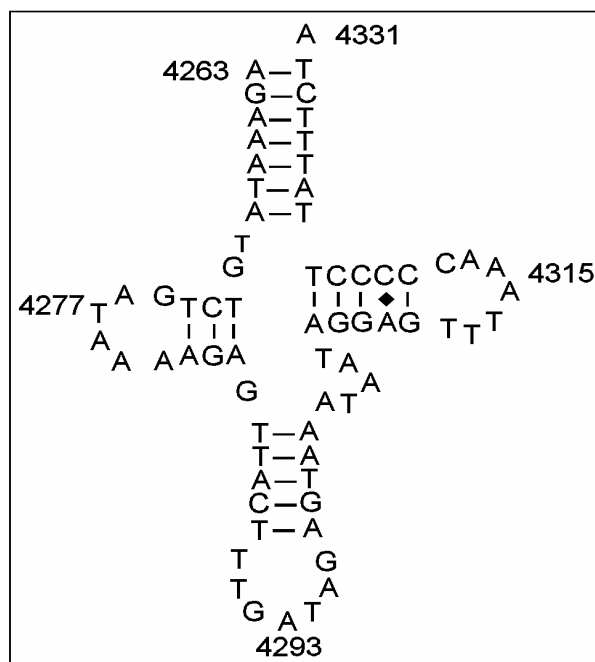

Figure 2 The tRNA for isoleucine drawn in a 2-dimensional '*cloverleaf'* shape.

Table 8
tRNA Sequences From the CRS

Alanine: (complemented and reversed from CRS)
AAGGGCT TA GCTT AATTA   AAGT G GCTGA TT**TGC**GT TCAGT TGAT   GCAGA GTGGGGT TTTGC AGTCCTT A

Valine:
CAGAGTG TA GCTT AACACA AAGC A CCCAA CT**TAC**AC TTAGG AGAT   TTCAA CTTAAC   TTGAC CGCTCTG A

Isoleucine:
AGAAATA TG TCTG ATAA   AAGA G TTACT TT**GAT**AG AGTAA ATAAT AGGAG TTTAAAC CCCCT TATTTCT A

And, for interest as it illustrates a point that will be mentioned later, the following is the sequence for the tRNA of isoleucine from a Hill Wallaroo, a species of Kangaroo! (Janke 1997)

AGAAATA TG TCTG ACAA   AAGA A TTATC TT**GAT**AG GATAA ATTAT AGGGG TGCAAGC CCCCT TATTTCT A

Note (1): The anti-codon areas are shown in bold and underlined.
Note (2): The table shows each tRNA sequence grouped into 15 functional parts.
Note (3): See **Figure 2** to see the tRNA for isoleucine drawn as a two-dimensional "cloverleaf"

Table 9
The Locations of the Transfer RNAs
(Adapted from: http://www.mitomap.org/)

1. Phenylalanine	577 - 647		12. Serine(UCN)	7446 - 7514 (Complement)
2. Valine	1602 - 1670		13. Aspartic acid	7518 - 7585
3. Leucine(UUR)	3230 - 3304		14. Lysine	8295 - 8364
4. Isoleucine	4263 - 4331		15. Glycine	9991 - 10058
5. Glutamine	4329 - 4400 (Complement)		16. Arginine	10405 - 10469
6. Methionine	4402 - 4469		17. Histidine	12138 - 12206
7. Trytophan	5512 - 5579		18. Serine(AGY)	12207 - 12265
8. Alanine	5587 - 5655 (Complement)		19. Leucine(CUN)	12266 - 12336
9. Asparginine	5657 - 5729 (Complement)		20. Glutamic acid	14674 - 14742 (Complement)
10. Cysteine	5761 - 5826 (Complement)		21. Threonine	15888 - 15953
11. Tyrosine	5826 - 5891 (Complement)		22. Proline	15956 - 16023 (Complement)

Note: The two areas of coding for Leucine and Serine match against different anti-codons.

8. Mitochondrial Diseases

It is not intended here to discuss mitochondrial diseases in great detail, but there are several points to be made initially.

Firstly, the term itself can be used to include both the diseases that arise from the malfunctions in the mitochondria that are the result of errors in the nuclear chromosomal inheritance and also for diseases that are caused by errors in mtDNA inheritance. In this article only diseases that result from changes in the mtDNA genome itself will be considered.

Secondly, there is the fascinating problem of *homoplasmy* and *heteroplasmy* - by which is meant that a particular subject with all of their mtDNA strands identical one to another is termed *homoplasmic*; whereas, a person is said to be *heteroplasmic* if they have both a population of mtDNA strands that are one type and also a population of strands showing a mutation.

And thirdly, there are also very rare occasions when a severe illness can arise from major rearrangements, or deletions, of the mtDNA, but these will not be discussed further.

Disease associated with coding for genes:

Quite a large number of conditions are associated with mutations in the coding for genes, but for the most part each condition is very uncommon. However, the condition of *Leber's Hereditary Optic Neuropathy*

Table 10
A List of Mutations and Associated Disease for the tRNA Coding Area for Isoleucine

A4267G	Myopathy	(Muscle)
A4269G	Infantile Cardiomyopathy	(Heart)
T4274C	Ophthalmoplegia	(Eye)
G4284A	Varied familial presentation	(Various)
T4285C	Ophthalmoplegia	(Eye)
T4290C	Progressive Encephalopathy	(Brain)
T4291C	Hypomagnesaemic Metabolic Syndrome	(Blood pressure)
A4295G	Cardiomyopathy	(Heart)
G4298A	Ophthalomplegia	(Eye)
A4300G	Cardiomyopathy	(Heart)
G4309A	Ophthalmoplegia	(Eye)
A4317G	Infantile cardiomyopathy	(Heart)
C4320T	Encephalocardiomyopathy	(Brain & Heart)

(LHON) is common enough to warrant discussion. LHON is a well recognised condition and there are about 2,000 cases in the UK (among a population of about 60 million).

LHON is a condition that causes severely impaired vision and it was first described by *Theodore Leber* in about 1870. He, of course, had no understanding of the mode of inheritance and, indeed, this was still unknown until the 1980's. Now the condition is known to be associated with a number of mutations in the mtDNA.

In this condition the typical presentation is for an otherwise fit young man to lose a large part of his vision over just a few months; and once lost the vision does not recover. But although most common in young men, the condition can affect persons of all ages, both sexes and of all races. And, like with so many other mitochondrial diseases the condition can be mild or severe, fluctuate in the degree of its severity with time, and appear to skip generations.

It is thought the optic nerves (the nerves from the brain to the eyes) are affected in this condition because they consume large amounts of energy. And, the very slight impairment of function in the production of *energy* in the mitochondria of the cells in the optic nerves becomes significant. Some other problems, such as tremor, have been reported in sufferers of LHON, but it is more usual for a person with LHON to be normal except for the severe visual impairment.

Some mutations that have been shown to be associated with LHON are:

> 3460A (ND1), 10663C (ND4L), 11778A (ND4) or 14484C (ND6), among others.

These mutations can be found in family pedigrees and can also occur *sporadically*. The fact that several mutations can be involved with the condition suggests that the disease results from a general disruption to the *oxidative phosphorylation* process rather than from something specific to a particular mutation.

Disease associated with the coding for the tRNA's:

There are many diseases associated with mutations in the 22 tRNA coding areas and new discoveries continue to be made on a regular basis.

But whereas, most of the diseases caused by mutations in the gene coding area are uncommon, the diseases associated with mutations in the tRNA coding areas are very well known. For example, there are associations between mtDNA mutations and some forms of *deafness, eye disease, heart disease, myopathy, strokes and*

hypertension. And, quite probably, mtDNA mutations may also be involved in conditions such as *fatigue* and *migraine.*

Again however, the point has to be made that the *mutated mtDNA* itself does not cause the disease in the way that *cystic fibrosis* is caused by a faulty protein because of a mutation in the DNA of a chromosome, but rather the *mutated mtDNA* may lead to a reduction in the intracellular availability of a particular amino acid and it is this biochemical imbalance in the cells of a particular tissue that results in illness.

It is beyond the scope of this article to go through all the conditions associated with the different tRNA's, but it will be helpful to look at a representative tRNA to see what diseases are associated with the mutations in the coding area for that tRNA. The following discussion takes as its example the tRNA for isoleucine.

As shown above the tRNA coding for the amino acid isoleucine is located in the mtDNA from 4263 to 4331. In this area there are mutations which are *pathological* (harmful), some that are possibly neutral in their effects, and one that on theoretical grounds could actually be beneficial. For reasons that are not known it would appear that mutations in the mtDNA coding for the tRNA for isoleucine mainly cause trouble in the heart and the eyes. **Table 10** gives a list of mutations and associated diseases for the tRNA coding area for isoleucine (taken from the Mitomap database (October 2005).

It is appropriate here to discuss the mutations the isoleucine tRNA found in the published sequences of this tRNA and discuss why these mutations might, or might not, be pathological. The mutations that have been found in the sequences are:

> A4295G, A4310G, C4312T, A4315T and the insertion 4315.A

A4295G is considered to be pathological as seen in **Table 10** because it is very close to the anti-codon loop and presumably disturbs its functioning.

A4310G can be presumed to be of *low* importance as it re-establishes a C-G pairing. The form 4310G is considered to be the older form and can be seen to be present in the *wallaroo* sequence, shown above.

C4312T can also be presumed to be of *low* importance as 4312T is thought to be the older form, and 98% of mankind in effect show the mutation 4312C. However, A4315T and 4315.A, unfortunately, must both be considered to be of *possible* harmful significance as each alters the shape of a loop of the tRNA.

Table 11
First Report

Origin: Italy   Haplogroup: H9   Identity: Not revealed

Mutation list:	Genome Area	Frequency:	Effect:	Importance:
T152C	HVR-2	High	Nil	Very low
A263G	HVR-2	High	Nil	Very low
315.C	HVR-2	High	Nil	Very low
A750G	12S rRNA	High	?	Low
A1438G	12S rRNA	High	?	Low
G3591A	ND1	Very low	Leu > Leu	Low
A4310G	tRNA-Ile	Very low	Form C-G pair	Low
A4769G	ND2	High	Met > Met	Low
A8860G	ATP6	High	Thr > Ala	Low
T9148C	ATP6	Low	Leu > Leu	Low
T13020C	ND5	Low	Gly > Gly	Low
A15326G	CYT-B	High	Thr > Ala	Low
C16168T	HVR-1	Low	Nil	Very low
T16519C	HVR-1	Common	Nil	Very low

9. Analysis of Selected Complete mtDNA Sequences

The above discussion in parts 1-8 has been rather detailed but the reader should now have some idea as to how a "mutation" list obtained from a given sequence can be interpreted with each mutation being rated as to the possibility of its causing harm.

Note the importance of a mutation will be *low* whenever its frequency in the world is *high*, or the effect on the tRNA, or gene, is considered *unimportant*—either because it stabilises the structure, is a *reversion* back to a common mammalian form, or does not change the amino acid. To show this, possible medical consequence reports for two sequences are shown in **Tables 11 and 12**.

For the present it is unlikely that genealogical testing companies will wish to produce such reports, but the reports here are produced to show how such reports might appear.

The first sequence is from a person in haplogroup H9 (Achilli 2004) and has the mutations shown in **Table 11**. The sequence includes the mutation at A4310G, - which above was shown to be of *low* importance. All of the other mutations are similarly of *low*, or *very* low importance, supporting the overall conclusion of the report is that *none* of the mutations suggest a significant medical problem.

However, compare this with a second report from a person in Haplogroup K1 (Coble 2004) and has the mutations shown in **Table 12**. This sequence includes the mutation A4295G, which above was shown to be associated with *cardiomyopathy*. There are also three other mutations that may *possibly* be important. The overall conclusion from this report is that the mutations suggest a *significant medical risk* to the subject.

10. So is There a Risk From Complete mtDNA Sequencing ?

Yes, and the risk from having a complete mtDNA sequencing begins when a person first thinks about ordering the test - as for the first time it is possible to have a test performed for genealogical purposes that can actually affect the person's well being.

Before agreeing to mtDNA sequencing it is important to consider just how one might be affected by the results. Fortunately, for about 90-95% of people the results will not show any mutations of medical significance. But this means that about 5-10% of people can expect results that show a *significant* mutation—for example, as is shown in **Table 12** where the subject has the *significant* mutation at A4295G. Because of the random nature of mutations, it is impossible to predict in advance who might receive such results.

The figure of 5-10% is the present author's estimate from studying the mutations found in all of the 2,500 published complete mtDNA sequences. It is possible that some of the sequences have been published by the researchers because of the mutations that the sequences contain—this would make the 5-10% figure too high.

Table 12
Second Report

Origin: America   Haplogroup: K1   Identity: Not revealed

Mutation list:	Genome Area	Frequency:	Effect:	Importance:
A73G	HVR-2	High	Nil	Very low
A263G	HVR-2	High	Nil	Very low
315.C	HVR-2	High	Nil	Very low
C497T	HVR-2	Common	Nil	Very low
523.C	HVR-2	Common	Nil	Very low
523.A	HVR-2	Common	Nil	Very low
A750G	12S rRNA	High	?	Low
T1189C	12S rRNA	Common	?	Low
A1438G	12S rRNA	High	?	Low
A1811G	12S rRNA	Common	?	Low
A2706G	12S rRNA	Common	?	Low
A3480G	ND1	Common	Lys > Lys	Low
A4295G	tRNA-Ile	Very Low	Critical	Significant
A4769G	ND2	High	Met > Met	Low
A5711G	tRNA-Asn	Low	Adds C-G	Possible
G6260A	COX1	Low	Glu > Glu	Low
C7028T	COX1	Common	Ala > Ala	Low
A8860G	ATP6	High	Thr > Ala	Low
G9055A	ATP6	Common	Ala > Thr	Low
T9698C	COX3	Common	Leu > Leu	Low
A10398G	ND3	Common	Thr > Ala	Low
A10550G	ND4	Common	Met > Met	Low
G10586A	ND4	Low	Ser > Ser	Low
T11299C	ND4	Common	Thr > Thr	Low
A11467G	ND4	Common	Leu > Leu	Low
T11485C	ND4	Low	Gly > Gly	Low
G11719A	ND4	Common	Gly > Gly	Low
C11840T	ND4	Low	Leu > Leu	Low
A12308G	tRNA-Leu	Common	Harmless	Low
G12372A	ND5	Common	Leu > Leu	Low
T13740C	ND5	Low	Ile > Ile	Low
T13886C	ND5	Low	Leu > Pro	Possible
C14167T	ND6	Common	Glu > Glu	Low
T14502C	ND6	Low	Ile > Val	Possible
C14766T	CYT-B	Common	Ile > Thr	Low
T14798C	CYT-B	Common	Phe > Leu	Low
A15326G	CYT-B	High	Thr > Ala	Low
G15884A	CYT-B	Common	Ala > Thr	Low
T16224C	HVR-1	Common	Nil	Very low
T16311C	HVR-1	Common	Nil	Very low
T16519C	HVR-1	Common	Nil	Very low

But against this, mutations associated with medical conditions continue to be identified—this would raise the figure. The author suggests the figure of 5-10% when mtDNA is sequenced for genealogical purposes

Besides the person themselves worrying, should they be telling other people that they are going to be tested? And, if so, who should be told? Close relatives and doctors, perhaps. But probably not at this early stage an employer or health insurance company. It is difficult to judge how important *pre-testing anxiety* might become, but for

some people, fear of adverse results will quite properly stop them from undergoing the testing.

However, if mtDNA sequencing is undertaken it becomes important to determine just how significant the results might be. The results for the Haplogroup H9 person shown earlier, demonstrate that test results can suggest there are no adverse medical mutations, and in such a case there would not appear to be any reason for concern. For the great majority of tests the outcome will be similar.

But what about a case such as that for the Haplogroup K1 person (shown earlier)? Here the results suggest that the subject has a mutation linked to a significant heart condition.

In such a situation it would seem important to convey things in their proper proportions and give the person the most complete understanding of the situation as is possible. In most instances it would be sufficient to explain that most mtDNA mutations do not cause the disease with which they are associated, but rather explain why the particular condition might arise.

Just at present it would seem that most genealogical testing companies are not involving themselves in analysing their results for any medical implications; and it is uncertain how such service might be provided. But what is clear is that counselling people who have been been found to have important mutations will not be easy, and perhaps will lead to as many unanswerable questions being posed, as questions for which answers can be given.

But it is not just the person who receives test results indicating possible medical problems who is affected— other family members will also have obvious concerns. A mother who receives adverse test results will certainly worry about the health of her children; and a younger person will have concerns about their siblings and mother.

Medically significant mtDNA mutations should not be ignored, in the author's view, and their possible significance should be discussed with the family's medical practitioners.

Overall, the discovery of potentially harmful mtDNA mutations may have many unexpected and undesirable consequences.

11. Sequencing for reasons other than for genealogy

Having discussed sequencing for genealogical purposes it would now seem appropriate to discuss briefly three occasions when a person may choose to have complete mtDNA sequencing because it is now available at reasonable cost. The test may be the result of the person's own decision, or perhaps it might be performed on the advice of their doctor. In these cases the person is less likely to be worried about finding a medically significant mutation—as this is what they are seeking.

The first instance is when a person suspects they might have a particular mtDNA mutation because a maternal relative has an identified mitochondrial condition. In this situation the person must be very careful to understand that should the test show up the mutation then they may be at some risk of developing the same condition. But, a negative result is to be viewed with some suspicion as a single test of a buccal (cheek cell) smear (the usual method of sampling for genealogical testing) is not a very satisfactory method of testing for any condition that may be *heteroplasmic*.

The second occasion is when a person with a medical condition may wish to undergo testing to prove the mitochondrial origin for their illness. A particular example of this might be a person with failing vision in whom *Leber's Hereditary Optic Neuropathy* (LHON) is a possibility. It should be remembered that a negative result cannot be assumed to be reliable because of the possibility of *heteroplasmy*. However, the commonest mutation for LHON, T14484C, is usually *homoplasmic*; and if T14484C is shown then considering the rarity of the mutation it will be likely that the condition is present and as a result the cause of their failing vision. This might also be the case when considering conditions such as *cardiomyopathy*, a more general *myopathy*, or *ophthalmoplegia* (paralysis of one or more muscles that move the eyes).

The third reason that some may choose to have such a test is perhaps the most interesting and the most speculative, since it is the availability of the test itself that may lead to it being ordered by people suffering from a number of medical conditions.

If one considers that most mitochondrial diseases are very variable in their effects, it is reasonable to say that sufferers of a number of non-specific illnesses may well have a significantly increased chance of having an hitherto unexpected mitochondrial disease.

So what conditions might these be? What is a non-specific illness?

Here are to be considered conditions that do not have a well understood basis to their causation. Most diseases have *symptoms* and observable *physical signs* - so there is usually no doubt about a person suffering from a chromosomal disorder, a bacterial or viral infection, the after-effects of an injury, degeneration conditions, a cancer, and auto-immune disorders. But what about *symptoms* of *fatigue, weakness, headache, loss of*

appetite, and conditions lacking any definitely observable *signs.* In these situations, whilst there is the possibility of a mitochondrial basis, it is likely that there is little to be gained by sequencing the mtDNA.

However, there still remains a significant number of people who have the vague *symptoms* as mentioned above, but who also have fully accepted *physical signs.* A person may well have a single nerve paralysis (a so called *mononeuritis multiplex*), history of a small stroke or encephalitic episode, heart disease with a proven reduction of cardiac output—but normal coronary arteries, or muscle weakness to the extent that a muscle biopsy may be a consideration. In all these cases the similarity to known mitochondrial diseases is clear.

So, what is the chance of there being a mitochondrial mutation as a basis to the illness suffered by a person suffering from this sort of condition? If it is considered that a person in the normal population has a risk of an *adverse mutation* of 5-10%, then in a selected group the author considers that the risk might be doubled—at which level complete mitochondrial sequencing could become a very sensible option. It is, after all, with the reduction in the cost of testing, cheaper to have complete mtDNA sequencing than to have a set of hospital-based tests.

Hitherto, the study of diseases with a mitochondrial basis has largely been developed by sequencing the mtDNA from persons showing signs of the disease. But now there is the opportunity to try a different approach: one could discover what mitochondrial mutations exist and then consider what conditions they might cause. The next few years can be expected to be very exciting.

Conclusion

In this article the medical implications of complete mitochondrial sequencing have been discussed; and, the conclusion must be that for a fit person there is a most definite risk to submitting to complete mtDNA sequencing. This risk is not high, with there being, maybe, a 5-10% chance of potentially harmful mutations being discovered. But mitochondrial diseases are important and may well be much more common than hitherto realized, especially in association with non-specific illnesses.

Electronic-Database Information

Mitomap   http://www.mitomap.org
                 The main database for mitochondrial
                 diseases and mutations

mtDB   http://www.genpat.uu.se/mtDB/
                 A searchable database of published
                 mtDNA sequences

Pubmed   http://www.ncbi.nlm.nih.gov/entrez/
                 Query.fcgi?=db=pubmed
                 An important database for medical and
                 scientific papers

NCBI   http://www.ncbi.nlm.nih.gov/entrez/
                 Viewer.fcgi?db=nucleotide
                 The main source for published mtDNA
                 sequences

Mamit-trna   http://mamit-trna.u-strasbg.fr/
                 index.html
                 A database of mammalian tRNAs

Software

MitoAnalyzer. National Institute of Standards and
                 Technology, Gaithersburg, MD, USA.
                 http://www.cstl.nist.gov/biotech/strbase/
                 mitoanalyzer.html, 2000.
                 A useful program that shows the effect of
                 mtDNA mutations

References

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogväli EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet, 75:910-918.

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, 290:457-65.

Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons T (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int J Legal Med, 118:137-146.

Janke A., Xu X., Arnason U (1997) The complete mitochondrial genome of the wallaroo (Macropus robustus) and the phylogenetic relationship among Monotremata, Marsupialia and Eutheria. Proc Natl Acad Sci USA, 94:1276-1281.

Kivisild T, Shen P, Wall D, 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

(2005) The role of selection in the evolution of human mitochondrial genomes. Genetics, in press.

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, 308:1034-1036.

Sykes B. (2001) The Seven Daughters of Eve. Bantam Press.

Taylor RW, Turnbull DM (2005) Mitochondrial DNA mutations in human disease. Nat Rev Genet, 6:389-402.

Selected Further Reading

Of particular interest:

Available in full at
http://nar.oxfordjournals.org/cgi/content/full/32/18/5430
Levinger L, Morl M, Florentz C (2004) Mitochondrial tRNA 3' end metabolism and human disease. Nucleic Acids Res, 32:5430-41.

Other papers of interest:

Gattermann N, Wulfert M, Junge B, Germing U, Haas R, Hofhaus G (2005) Ineffective hematopoiesis linked with a mitochondrial tRNA mutation (G3242A) in a patient with myelodysplastic syndrome Blood, 103:1499-1502.

Hao R, Zhao MW, Hao ZX, Yao YN, Wang ED (2005) A T-stem slip in human mitochondrial tRNALeu(CUN) governs its charging capacity. Nucleic Acids Res, 33:3606-3613.

Jacobs HT, Hutchin TP, Kappi T, Gillies G, Minkkinen K, Walker J, Thompson K, Rovio AT, Carella M, Melchionda S, Zelante L, Gasparini P, Pyykko I, Shah ZH, Zeviani M, Mueller RF (2005) Mitochondrial DNA mutations in patients with postlingual, nonsyndromic hearing impairment. Eur J Hum Genet, 13:26-33.

Kelley SO, Steinberg SV, Schimmel P (2000) Functional defects of pathogenic human mitochondrial tRNAs related to structural fragility. Nat Struct Biol, 7:862-865.

Kelley SO, Steinberg SV, Schimmel P (2001) Fragile T-stem in disease-associated human mitochondrial tRNA sensitizes structure to local and distant mutations. J Biol Chem 276:10607-10611.

Levinger L, Oestreich I, Florentz C, Morl M (2004) A pathogenesis-associated mutation in human mitochondrial tRNALeu(UUR) leads to reduced 3'-end processing and CCA addition. J Mol Biol, 337:535-544.

Moilanen JS, Finnilä S, Majamaa K (2003) Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J Mol Biol Evol, 20:2132-2142.

Ruppert V, Nolte D, Aschenbrenner T, Pankuweit S, Funck R, Maisch B (2004) Novel point mutations in the mitochondrial DNA detected in patients with dilated cardiomyopathy by screening the whole mitochondrial genome. Biochem Biophys Res Commun, 318:535-543.

Shin WS, Tanaka M, Suzuki J, Hemmi C, Toyo-oka T (2000) A novel homoplasmic mutation in mtDNA with a single evolutionary origin as a risk factor for cardiomyopathy. Am J Hum Genet, 67:1617-1620.

Wilson FH, Hariri A, Farhi A, Zhao H, Petersen KF, Toka HR, Nelson-Williams C, Raja KM, Kashgarian M, Shulman GI, Scheinman SJ, Lifton RP (2004) A cluster of metabolic defects caused by mutation in a mitochondrial tRNA. Science. 306:1190-1194.