

The Meaning of Your Mutations¹

Ann Turner, M.D.
DNACousins@aol.com
September 13, 2015

This report on coding region mutations was prepared expressly for participants in the *Nancy Hanks Lincoln mtDNA Study*. It is based on a core set of mutations found in all participants of the project, who can demonstrate a line of descent from matrilineal kin of Nancy Hanks Lincoln (NHL). Some participants may have additional mutations that arose in their own branch. Conversely, NHL herself may have had some mutations not found in other lines of descent, but she has no living descendants to test.

Mitochondrial DNA (mtDNA) is of special interest because of its inheritance pattern – it is inherited through the egg, so your mtDNA came from your mother, who got it from her mother, who got it from her mother...clear back to the common matrilineal ancestor of all humanity, sometimes nicknamed mitochondrial Eve. “Eve” was not the first woman, nor was she the only woman alive at the time. Her contemporaries could still have descendants, whose line zig-zagged back and forth between males and females, but Eve was the only one who had an unbroken line of daughters, through thousands of generations, clear down to the present time.

The Cambridge Reference Sequence

Your mitochondrial DNA was completely sequenced and compared to the Cambridge Reference Sequence (CRS). There is nothing special about the CRS – it was simply the first one to be sequenced in 1981, using a placenta obtained from a maternity hospital located conveniently close to Cambridge University in England (Anderson 1981). Today, mtDNA can be sequenced using a much smaller sample of cells, easily gathered from the inside of your cheeks.

As more samples from around the world were examined, it became clear that all human beings shared 99% or more of their mtDNA sequence. Yet the remaining portion, a mere handful of differences, was enough to sort people into clusters, based on similar patterns in their mtDNA.

These broad clusters are known as haplogroups, and the divisions between haplogroups occurred tens of thousands of years ago. Some haplogroups are common in Africa, others hail from Europe, and still others forge a bond between Asia and the Americas.

Additional mutations have accumulated since the founding mothers of each haplogroup lived, forming branches called subhaplogroups or subclades. A diagram showing these branches is called a phylogenetic tree. Even more distinctive *haplotypes* (the complete set

¹ Disclaimer: This report is not intended to provide medical advice. If you have any concerns, please consult with your personal physician.

of differences from the CRS) have not been classified into subclades, and they may point to an ancestor who lived more recently. Your mtDNA reflects many layers of human history.

The CRS sample was resequenced a few years ago, using more modern techniques (Andrews 1999). A very few errors were detected, and Family Tree DNA (FTDNA) compares your results to the revised edition, sometimes called the rCRS. However, the acronym CRS will be used throughout the remainder of this report.

A brief biology lesson

A little background will help you understand the significance of your own mutations. However, you may skip ahead to *Your differences* if you're eager to see them.

Mitochondria are essential cell structures, responsible for converting food into energy. Each cell has hundreds of mitochondria, and each mitochondrion has several copies of the mtDNA molecule.

The mtDNA molecule is circular, with 16,569 bases. The numbering begins at an arbitrary point in the middle of the D-loop (displacement loop, a section which spreads apart when the mtDNA molecule begins to replicate). This region is sometimes called the control region (due to its role in promoting replication), in distinction from the coding region.

The control region includes base positions 16024 through 16569 and continues around the circle to include bases 1 through 576. Because the control region is not responsible for producing proteins, mutations can accumulate without obvious adverse effects. Indeed, this stretch is also called the hypervariable region (HVR), because it provides the best hunting ground for finding differences among people. Until very recently, mtDNA testing for the consumer market was limited to the control region.

The coding region covers the remaining bases, in positions 577 to 16023. A complete list of the various functional areas is shown in Appendix 1. This region is densely packed, including genes for thirteen different proteins involved in breaking down the big molecules found in food. Each protein is composed of amino acids, which are arranged in a particular order, specified by the genetic blueprint. These particular proteins are enzymes, which facilitate chemical reactions in very small and delicate steps, so that the cell does not burst into flames as it burns food for energy.

In addition to the genes, there are two ribosomal RNAs (rRNA) and twenty-two transfer RNAs (tRNA). Ribosomes are like miniature factories, with two floors and an assembly line for constructing proteins. The two rRNAs are called 12S RNA and 16S RNA, for the size of the factory floors. The tRNAs each embrace a specific amino acid and ferry it to the factory floor, ready for the assembly workers who are following the specifications in the genetic blueprint. The genetic code uses three bases (a codon) to specify one of the twenty-some amino acids. Appendix 2 shows the codons for each amino acid.

The effective mutation rate for the coding region is lower than for the hypervariable region. Many changes would be so harmful that a woman might not even know that she had conceived. This mutation would disappear without a trace. Yet other changes seem to be relatively benign in their effect, as explained in more detail below. These polymorphisms (poly = many, morph = form), being relatively stable compared to the hypervariable regions, are useful for defining haplogroups. However, parallel and back mutations can occur – the same mutation can occur in different branches of the family tree, or one branch may revert to the ancestral value. Thus it is important to look at the whole picture, not just one location.

Your differences

Your differences from the CRS are shown in Table 1, as presented on your FTDNA personal results page, but color-coded to show the significance of various changes.

750G	1438G	2706G	<table border="1"> <tbody> <tr> <td>"private"</td> </tr> <tr> <td>X1c</td> </tr> <tr> <td>X1</td> </tr> <tr> <td>X</td> </tr> <tr> <td>not H or V</td> </tr> </tbody> </table>	"private"	X1c	X1	X	not H or V
"private"								
X1c								
X1								
X								
not H or V								
4769G	5302C	6221C						
6371T	7028T	7337A						
8860G	9055A	9615C						
9758C	11719A	12705T						
13857G	13966G	14470C						
14587G	14766T	15326G						
15654C								

Table 1

The gray boxes show the locations where you have a difference because you are *not* in haplogroup H2a2a1 (where the CRS is located). As shown in Figure 1, the CRS is just a small twig on one of the major branches of humanity, and most people (even most people in haplogroup H) will have those five polymorphisms at 750G, 1438G, 4769G, 8860G, and 15326G. You also differ from the CRS at locations leading to haplogroup H (2706G and 7028T), HV (11719A and 14766T), and R (12705T). In other words, you have the ancestral values here, and the actual mutations occurred en route to haplogroup H.

This brings you to the N node. X branches off the main line with mutations 6221C, 6371T, 13966G, and 14470C. It may not seem logical for X to be a branch of N, but these labels were created at a time when there were very few complete mtDNA sequences. Partial information was good enough to recognize clusters, but not enough to appreciate how they fit together in the whole phylogenetic tree.

X continues to branch, with X1 being defined by 5302C, 14587G, and 15654C. Your most specific assignment is X1c, with mutations 7337A and 9615C.

The color yellow is reserved for the most recent mutations, the ones that have not been observed often enough to be formally recognized as a subhaplogroup. Population geneticists sometimes call these “private” mutations, using the word in the sense of “*confined to particular persons or groups.*” Private mutations may in fact be very recent or

quite old, but regardless of their absolute age, they are the ones that narrow down the pool of matrilineal relatives to your closest cousins. Not everyone will even have a private mutation. Your private mutations are 9055A, 9758C, and 13857G.

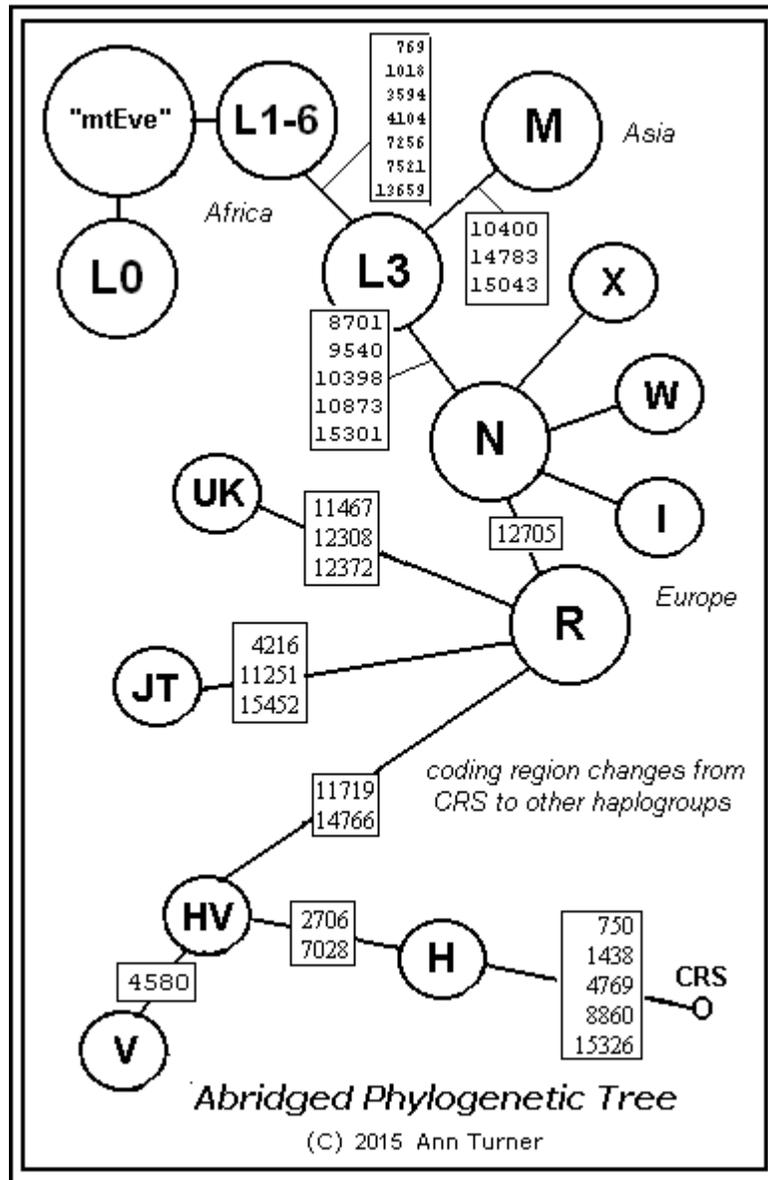


Figure 1

The CRS is in a rare subhaplogroup of H, with a number of mutations that occurred after the clan mother for haplogroup H lived. Anyone who is not in the same subhaplogroup as the CRS will show the five differences listed between H and the CRS. Similarly, anyone who is not in H will show differences at 2706 and 7028, and anyone who is not in H or V will show differences at 11719 and 14766. Haplogroups J and T are closely related, as shown by the mutations they share; likewise U and K spring from a common origin. People in haplogroups N, X, W and I will all note a difference at 12705. L3 is the source for the major European and Asian haplogroups, N and M. Mutations for subhaplogroups, and mutations from L1-L6 back to mtEve, are not shown, nor are the numerous branches springing from M and the L subhaplogroups that remained in Africa.

Ranking your mutations

Table 2 shows your polymorphisms arranged in a different order, roughly ranked from low to high according to the degree of interest they hold for you personally. At the lowest level, millions if not billions of people will show the same “mutations.” Polymorphisms at a higher level merit your special attention. Some people will only see mutations at the first two levels.

The table is full of technical details about each of your mutations. If you’re not especially interested in “looking under the hood,” just check column one for the highest code and skip ahead to the next section, *Information about your most distinctive mutations*.

The mutations are coded as follows:

- 1) Polymorphisms where the CRS has the rare value. Most people in the world, even most people within haplogroup H, will show these differences. Your differences may be called mutations, but the mutations actually occurred in the CRS – you have the ancestral versions.
- 2) Polymorphisms defining haplogroups and subhaplogroups, as shown in Table 1. Since these mutations have persisted in many people for thousands of years, there is little reason to suspect they have any great medical significance.
- 3) Polymorphisms in non-coding positions, just a few spots located here and there between the functional areas.
- 4) Polymorphisms in ribosomal RNA, with no known adverse effects. The factory floor has some minor remodeling that does not affect the assembly line.
- 5) Polymorphisms in transfer RNA, with no known adverse effects. The tRNA can still grab the right amino acid and deliver it to the factory.
- 6) Polymorphisms pertaining to amino acids
 - a) Synonymous – Several different three-base codons may be used for the same amino acid. It is like substituting the word “nice” for “pleasant” when describing today’s weather. The words convey the same basic meaning. See Appendix 2.
 - b) Conservative – The amino acid is different, but it has similar properties, such as size or electric charge. It is like saying “The weather will be warm today.” Most people would agree that warm weather is nice and pleasant, but the meaning is not precisely identical.
 - c) Non-conservative – The amino acid is different, with different properties. It is like saying “The weather will be a little breezy today.” The protein function may be affected in subtle ways; however, no disease has been associated with the polymorphism. To carry the analogy further, the weather is still suitable for a picnic, but you might need to arrange the paper napkins so they won’t blow away.
- 7) Polymorphisms linked to a disease. Discretion should be used when sharing these mutations publicly.

These categories are overlapping and not mutually exclusive. For example, the mutation at 14766, defining superhaplogroup HV, is a non-conservative mutation, yet large numbers of people with and without the mutation survive and thrive.

Code	Position	You#	CRS	Function	Amino Acid
1	750	G	A	12S_rRNA	
1	1438	G	A	12S_rRNA	
1	4769	G	A	ND2	synonymous
1	8860	G	A	ATPase6	Thr->Ala
1	15326	G	A	Cytb	Thr->Ala
2	2706	G	A	16S_rRNA	
2	5302	C	T	ND2	Ile->Thr
2	6221	C	T	COI	synonymous
2	6371	T	C	COI	synonymous
2	7028	T	C	COI	synonymous
2	7337	A	G	COI	synonymous
2	9055	A	G	ATPase6	Ala->Thr
2	9758	C	T	COIII	synonymous
2	11719	A	G	ND4	synonymous
2	12705	T	C	ND5	synonymous
2	13857	G	A	ND5	synonymous
2	13966	G	A	ND5	Thr->Ala
2	14470	C	T	ND6	synonymous
2	14587	G	A	ND6	synonymous
2	14766	T	C	Cytb	Ile->Thr
6a	9615	C	T	COIII	synonymous
6c	15654	C	T	Cytb	Met -> Thr

Table 2

The bases C and T are rather similar to each other in chemical structure, and likewise for the bases G and A. Thus most substitutions are C <-> T and G <-> A (called “transitions”). Other combinations (called “transversions”) occur more rarely, and they are less subject to parallel and back mutations. Your mtDNA shows no transversions.

You may also occasionally encounter news items about some condition being associated with a certain haplogroup. These are often preliminary reports, which do not hold up when the study is repeated in a different population. For instance, one study set in northern Italy found that haplogroup J was associated with longevity. Yet a later study in southern Italy did not replicate the finding. Many medical researchers do not take haplogroup structure into account, and they may report a “mutation” that is actually common throughout the world. As Herrnstadt wrote in her article, *An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders*,

“As we note here, however, such associations have usually been observed only in single studies and it is difficult to draw broad conclusions on the basis of the available evidence. At a minimum, we suggest that, a haplogroup-group association must be detected in multiple subpopulations or in a large, carefully controlled population survey.”

Information about your most distinctive mutations

Figure 2 shows how your haplotype might be placed on a future version of the phylogenetic tree maintained by Mannis van Oven. He uses GenBank, a centralized

collection of sequence data deposited by authors of technical articles (and more recently by private individuals who have obtained their complete sequence).²

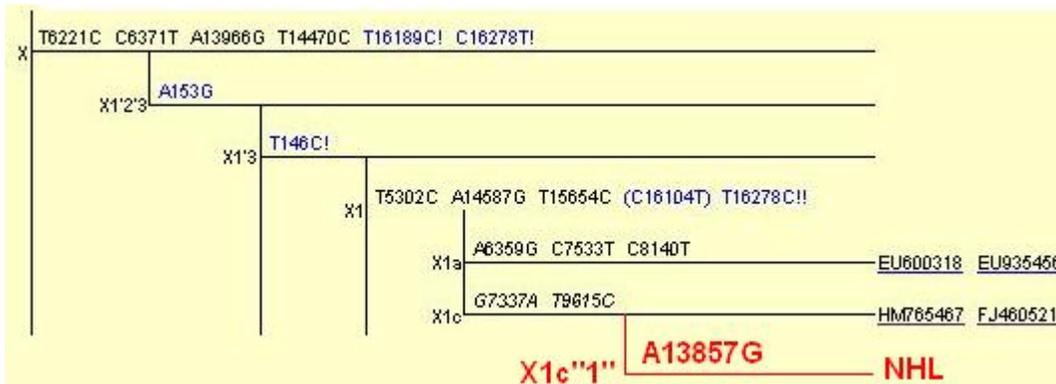


Figure 2
Modified from Build 16
<http://phylotree.org>

There are a total of five X1c sequences at GenBank (out of ~ 30,000 records), three without any geographic information. Two of the sequences share one of your “private” mutations, A13857G, which could become the motif for a new subclade. I have called it X1c“1” with quotation marks to emphasize the informal and provisional nature of the label. One of the sequences (from Tunisia) has the GenBank ID FJ460521. The other sequence (from Spain) has the GenBank ID KM245146. This does not necessarily mean that *your* ancestors ever lived in those places. The clan mother could have lived elsewhere, with descendants migrating here and there, but an origin in the Western Mediterranean is possible. Additional locations will undoubtedly be added as more and more sequences are added to GenBank.

Behar has estimated the age of X1c at about 14,400 years (plus or minus 5300 years),³ based on the amount of variation that has accumulated within the haplogroup. The broad range reflects the uncertainty about whether the current samples are truly representative. The two sequences with the possible motif 13857G differ from each other and from you at several other positions, implying that the split (if genuine) occurred many thousands of years ago.

A good way of keeping track of future developments is to periodically use Google Scholar, <http://scholar.google.com>, with its full-text index and links to sources. The Advanced Search options allow you to limit hits to certain dates. A possible search strategy would be

mtDNA haplogroup X1c

² <http://www.ncbi.nlm.nih.gov/Genbank/>

³ Table S5 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3322232/bin/mmc1.pdf>

Publications Consulted

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
Sequence and organization of the human mitochondrial genome
Nature. 1981 Apr 9;290(5806):457-65

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N.
Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.
Nat Genet. 1999 Oct;23(2):147.

Behar DM et al, A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. Cell. Am J Hum Genet 2012 Apr; 90:675–684.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3322232/>

Coble, MDW, The Identification of Single Nucleotide Polymorphisms in the Entire Mitochondrial Genome to Increase the Forensic Discrimination of Common HV1/HV2 Types in the Caucasian Population, Dissertation, George Washington University, 2004

Finnila S, Lehtonen MS, Majamaa .
Phylogenetic network for European mtDNA.
Am J Hum Genet. 2001 Jun;68(6):1475-84

Florentz C, Sissler M.
Disease-related versus polymorphic mutations in human mitochondrial tRNAs. Where is the difference?
EMBO Rep. 2001 Jun;2(6):481-6

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N.
Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups.
Am J Hum Genet. 2002 May;70(5):1152-71.

Herrnstadt C, Howell N
An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders
Mitochondrion. 2004 Sep;4(5-6):791-8

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis KK, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman MW, Cavalli-Sforza LL, Oefner PJ
The role of selection in the evolution of human mitochondrial genomes.
Genetics. 2006 Jan;172(1):373-87

Logan I

The Medical Implications of Complete Mitochondrial DNA Sequencing.
Journal of Genetic Genealogy. 2005 Fall; 1(2): 40-53.

<http://www.jogg.info/12/Logan.pdf>

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM.
Major genomic mitochondrial lineages delineate early human expansions.
BMC Genet. 2001;2:13. Epub 2001 Aug 13.

Moilanen JS, Majamaa K.

Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA.

Mol Biol Evol. 2003 Aug;20(8):1195-210

Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP.

Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia.

Am J Hum Genet. 2004 Dec;75(6):966-78.

Pereira L et al. Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity.
Am J Hum Genet. 2011 Apr 8;88(4):433-9.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3071914/>

Ruiz-Pesini E, Wallace DC.

Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA.

Hum Mutat. 2006 Nov;27(11):1072-81.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB.

Correcting for purifying selection: an improved human mitochondrial molecular clock.

Am J Hum Genet. 2009 Jun;84(6):740-59.

Taylor RW, Turnbull DM.

Mitochondrial DNA mutations in human disease.

Nat Rev Genet. 2005 May;6(5):389-402

van Oven M, Kayser M

Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation

Hum Mutat 29:E386-E394. Epub date October 13, 2008

<http://phylotree.org>

Vilmi T, Moilanen JS, Finnila S, Majamaa K.
Sequence variation in the tRNA genes of human mitochondrial DNA.
J Mol Evol. 2005 May;60(5):587-97

Websites consulted

Amino acid properties and consequences of substitutions

<http://www.russelllab.org/aas/>

Haplogroup motifs – HVR1 mutations that are commonly found in various haplogroups

<http://www.stats.gla.ac.uk/~vincent/founder2000/motif.html>

Human Mitochondrial Database

<http://www.mtodb.iqg.uu.se/>

Mitoanalyzer (note: this uses the original CRS, not the rCRS)

<http://www.cstl.nist.gov/biotech/strbase/mitoanalyzer.html>

Mitomap, especially these three pages:

Sequence

<http://mitomap.org/bin/view.pl/MITOMAP/HumanMitoSeq>

rRNA/tRNA point mutations

<http://mitomap.org/bin/view.pl/MITOMAP/MutationsRNA>

coding region point mutations

<http://mitomap.org/bin/view.pl/MITOMAP/MutationsCodingControl>

Mitomaster

<http://www.mitomap.org/MITOMASTER/WebHome>

mtDNA Community

<http://www.mtDNAcommunity.org>

National Center for Biotechnology Information (NCBI) BLAST

<http://www.ncbi.nlm.nih.gov/blast/>

Neuromuscular Disorders -- Washington University, St. Louis, MO

<http://neuro.wustl.edu/neuromuscular/mitosyn.html>

OMIM – Online Mendelian Inheritance in Man

<http://www.ncbi.nlm.nih.gov/omim?cmd=search>

Phylogenetic tree from van Oven / Kayser

<http://phylotree.org>

Appendix 1

Location of the functional parts in the coding region

Bases	Function	Abbrev
577-647	tRNA phenylalanine	tPhe
648-1601	12S ribosomal RNA	12S
1602-1670	tRNA valine	tVal
1671-3228	16S ribosomal RNA	16S
3230-3304	tRNA leucine 1	tLeu
3305-3306	non-coding nucleotides	
3307-4262	NADH dehydrogenase subunit 1	ND1
4263-4331	tRNA isoleucine	tIle
4329-4400	tRNA glutamine	tGln
4401-4401	non-coding nucleotide	
4402-4469	tRNA methionine	tMet
4470-5511	NADH dehydrogenase subunit 2	ND2
5512-5579	tRNA tryptophan	tTrp
5580-5586	non-coding nucleotides	
5587-5655	tRNA alanine	tAla
5656-5656	non-coding nucleotide	NC4
5657-5729	tRNA asparagine	tAsn
5761-5826	tRNA cysteine	tCys
5826-5891	tRNA tyrosine	tTyr
5892-5903	non-coding nucleotides	
5904-7445	Cytochrome c oxidase subunit I	CO I
7446-7516	tRNA serine 1	tSer
7517-7517	non-coding nucleotide	
7518-7585	tRNA aspartic acid	tAsx
7586-8269	Cytochrome c oxidase subunit II	CO II
8270-8294	non-coding nucleotides	
8295-8364	tRNA lysine	tLys
8365-8365	non-coding nucleotide	
8366-8572	ATP synthase F0 subunit 8	ATPase8
8527-9207	ATP synthase F0 subunit 6	ATPase6
9207-9990	Cytochrome c oxidase subunit III	CO III
9991-10058	tRNA glycine	tGly
10059-10404	NADH dehydrogenase subunit 3	ND3
10405-10469	tRNA arginine	tArg
10470-10766	NADH dehydrogenase subunit 4L	ND4L
10760-12137	NADH dehydrogenase subunit 4	ND4
12138-12206	tRNA histidine	tHis
12207-12265	tRNA serine2	tSer

12266-12336	tRNA leucine2	tLeu
12337-14148	NADH dehydrogenase subunit 5	ND5
14149-14673	NADH dehydrogenase subunit 6	ND6
14674-14742	tRNA glutamic acid	tGlu
14743-14746	non-coding nucleotides	
14747-15887	Cytochrome b	Cytb
15888-15953	tRNA threonine	tThr
15954-15954	non-coding nucleotides	
15955-16023	tRNA proline	tPro

Appendix 2

Names of the amino acids, with their three-letter abbreviations, one-letter symbols, and codons

Alanine	Ala	A	GCT, GCC, GCA, GCG
Arginine	Arg	R	CGT, CGC, CGA, CGG
Asparagine	Asn	N	AAT, AAC
Aspartic acid	Asx	D	GAT, GAC
Cysteine	Cys	C	TGT, TGC
Glutamine	Gln	Q	CAA, CAG
Glutamic acid	Glu	E	GAA, GAG
Glycine	Gly	G	GGT, GGC, GGA, GGG
Histidine	His	H	CAT, CAC
Isoleucine	Ile	I	ATT, ATC
Leucine	Leu	L	TTA, TTG; CTT, CTC, CTA, CTG
Lysine	Lys	K	AAA, AAG
Methionine	Met	M	ATA, ATG
Phenylalanine	Phe	F	TTT, TTC
Proline	Pro	P	CCT, CCC, CCA, CCG
Serine	Ser	S	TCT, TCC, TCA, TCG; AGT, AGC
Threonine	Thr	T	ACT, ACC, ACA, ACG
Tryptophan	Trp	W	TGA, TGG
Tyrosine	Tyr	Y	TAT, TAC
Valine	Val	V	GTT, GTC, GTA, GTG

Note that many of the synonymous codons differ in the third base.

There are rare instances where synonymous mutations have a functional effect, often involving splice sites in a gene's introns and exons. Mitochondrial genes have no introns, so this mechanism does not apply to mtDNA.