INTEGRATED DNA TECHNOLOGIES

Mitochondrial DNA (mtDNA)

Contents

Introduction: The Mitochondrion and Its Genome

Every eukaryotic cell contains at least one copy of the entire nuclear genome housed in its nucleus. In contrast, every cell contains as many as several thousand mitochondria. This organelle has been found to play a central role in numerous cellular functions such as metabolism (oxidative phosphorylation), apoptosis, and aging [1]. It has been known for many years that mitochondria are semi-autonomous, possessing their own genome and the machinery for replication, transcription, and protein synthesis [2]. The origin of mitochondria from a bacterial symbiont is commonly accepted but, since the discovery in the mid-1960's that mitochondria contain their own genome, several questions have gone unanswered. Among these questions are why eukaryotic cells would tolerate more than one genome and why the mitochondrial genome of eukaryotes has shed many but not all of its genes and has done so to a point at which it no longer contains sufficient indigenous information for replication and expression.

Mitochondria descended from free-living bacteria that became symbiotic with eukaryotic cells about 1.5 billion years ago. The original model of mitchondrial evolution held that the nucleus originated in an Archaebacterium and then the symbiosis began with a eubacterial progenitor of the modern mitochondrion [3]. The conventional "endosymbiosis theory" has been modified over the years and the revision has been labeled the "hydrogen hypothesis" [4, 5]. The hydrogen hypothesis postulates that the eukaryotic nucleus and the mitochondria were created simultaneously through fusion of a hydrogen-requiring methanogenic Archaebacterium (the host) and a hydrogen-producing alpha-proteobacterium (the symbiont). Muller and Martin base this revision on several observations made possible by the expansion of molecular biology in the 1980's and 1990's. First, the eukaryotic nucleus is a chimera of genes whose origins are clearly Archaebacterial and genes whose origins are clearly eubacterial. Second, the Archaeozoa, eukaryotes lacking mitochondria, contain mitochondria-like genes in their

nuclear genome. This suggests that the Archaeozoa once had mitochondria and lost them but not before there was lateral transfer of mitochondrial genes to the nuclear genome. Third, phylogenetic studies of Archaeozoa have shown that not all members of the family can be classified as basal eukaryotes. Some, like *Entamoeba histolytica*, are classified much further up the eukaryotic tree. Finally, mitochondrial genomes have been found to share common ancestry with hydrogenosomes in alpha-proteobacteria.

Regardless of which view of the origin of the mitochondria is correct, one thing is common to both views. The majority of the mitochondrial genes that existed in the symbiont genome of the proto-eubacterium have been transferred to the nuclear genome. In animals, mtDNA is usually small (15 to 20kb) and encodes 37 genes. Variations in the size of animal mtDNAs are due primarily to duplications rather than the presence of additional genes. The typical mitochondrial gene complement includes 13 protein subunits of the enzymes involved in oxidative phosphorylation, the two rRNAs of the mitochondrial ribosome, and the 22 tRNAs necessary for the translation of the proteins encoded [1]. A listing of mitochondria-encoded proteins and the correct gene nomenclatures are shown in Table 1.

Table 1
Animal mtDNA Genes and Gene Products

Gene Designation	Encoded Protein
COI, COII, COIII	Cytochrome oxidase subunits I, II, and III
Cytb	Cytochrome b apoenzyme
ND1-6, 4L	NADH* dehydrogenase subunits 1 to 6 and 4L
ATP6, ATP8	ATP synthase subunits 6 and 8
lrRNA	Large ribosomal subunit RNA
srRNA	Small ribosomal subunit RNA
tRNAs	18 amino acid-specific transfer RNAs
L(CUN) and L(UUR)	two leucine tRNAs
S(AGN) and S(UCN)	two serine tRNAs

*diaphorase, cytochrome b-5 reductase

In 1981 Anderson et al. published the sequence and organization of the human mitchondrial genome. This was the first mitochondrial genome to be sequenced and was 16,569bp long. The smallest mitochondrial genome sequenced to date is the 5967bp mtDNA of the parasite *Plasmodium falciparum* [6]. The largest mitochondrial genome sequenced to date is the massive 366,924bp mtDNA of the model plant *Arabidopsis thaliana* [7]. In all, GenBank currently archives more than 400 mtDNA sequences and more are added every year.

The Human Mitochondrial Genome

Publication of the human mtDNA sequence by Anderson et al. unveiled a number of surprising features [8]. The mitochondrial genome is as compact as any genome ever seen. Genes are packed in with little or no intergenic non-coding sequence and the genes themselves lack many of the traits normally expected in eukaryotic genes. Mitochondrial mRNAs lack non-translated leader and trailing sequences and more than half do not even have a stop codon. Stops are added upon polyadenylation when a terminal U or UA is converted to a UAA. The two ribosomal RNAs are the smallest known at 1,559 and 954 bases, there is no 5S RNA, and the 22 tRNAs are used to read all codons. The mitochondrial genetic code is different from the eukaryotic code; UGA is read as tryptophan rather than as STOP; AGA and AGG, normally read as arginine, are read as STOPs; AUA is methionine and not isoleucine; and the ubiquitous AUG start codon is sometimes replaced by AUA or AUU in mitochondrial genes. Subsequent studies of other mtDNAs have shown that the mitochondrial genetic code is not even universal among mitochondria. Yeast mitochondrial genomes, for example, are much larger and have not reassigned the AUA, AGA, and AGG codons. Yeast have reassigned CTN as leucine rather than threonine.

A map of the human mitochondrial genome is shown in Figure 1. The packing of the mitochondrial genome is evident. But, even though the genes are tightly packed, not all are transcribed in the same direction. Convention has designated a plus strand and a minus strand based upon gene transcription (remember, the mitochondrial genome is circular) and, while the majority of mitochondrial genes are transcribed on the plus strand, some are transcribed in the opposite direction on the minus strand. Transcription direction is indicated for all of the genes in Figure 1.

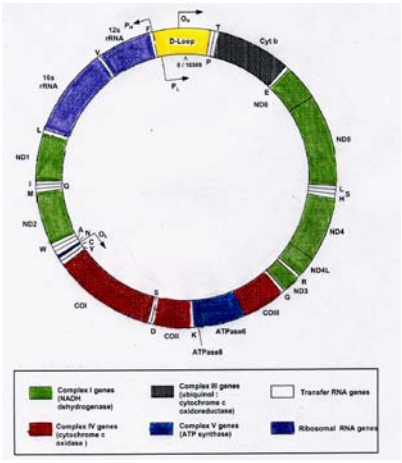

Figure 1. Map of the human mitochondrial genome. Loci are indicated by functional grouping. Gene identifiers on the outside of the map are transcribed on the plus strand and gene identifiers on the inside of the map are transcribed on the minus strand. Transfer RNA loci are designated by the single letter code of their specific amino acid. The non-coding D-loop is shown at the top of the map and nucleotide position 1 is at twelve o'clock. Figure adapted from MITOMAP: A Human Mitochondrial Genome Database. http://www.mitomap.org, 2003.

Mitochondria appear to lack an efficient DNA repair mechanism as well as protective proteins such as histones. Moreover, mitochondrial DNA is physically associated with the inner mitochondrial membrane where highly mutagenic oxygen radicals are generated [9]. As a consequence of these features, the mtDNA has a much higher mutation rate than does nuclear DNA [10]. As a result, mtDNA is involved in several hereditary human diseases. In general, organs such as heart, brain, and skeletal muscle, where aerobic demand is high and regenerative capacity is low, are the foci of mitochondrial disorders. Wallace et al. identified more than 50 deleterious mutants in human mtDNA and, of these, there are four that are the most frequent [10]. The four common mutants are associated with specific mitochondrial disorders. These are; mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS), caused by the mutation 3243A>G (nucleotide position and mutant); myoclonic epilepsy with ragged-red fibers (MERRF), caused by the mutation 8344A>G; neurogenic weakness, ataxia and retinitis pigmentosa (NARP), caused by the mutation 8998T>G; and Leber's Hereditary Optic Neuropathy (LHON), caused by the mutation 11778A>G. In addition to point mutants, there are a number of documented deletions and/or duplications in human mtDNAs that are causally associated with specific disorders or with increased risk for certain disorders.

The D-loop and Human Population Genetics

While the vast majority of the mitochondrial genome is under the scrutiny of selection because mutations in these areas are usually deleterious, there is a region in which there are no coding sequences and mutations are free to accumulate at will. This region is in the mitochondrial D-loop. The D-loop is the location of mitochondrial transcription promoters. MtDNA replication begins in the D-loop resulting in the formation of a displacement loop with a newly synthesized heavy, or H, strand of about 700nt known as 7S DNA [8]. Both strands of the mtDNA are completely transcribed from the promoters in the D-loop. In addition to the promoter sequences, there are two small regions known as the hypervariable regions I and II (HVI and HVII). Mutation rates in HVI and HVII are especially high on average and there is evidence that the rates vary within the regions as well [11].

As a result of the high average mutation rates and the lack of coding or regulatory sequences in the hypervariable regions, they have become a tremendously valuable source of presumably neutral human genetic variation. In addition, since mtDNA is maternally inherited (sperm do not have mitochondria), there is no recombination between parental genomes. Thus, in every generation, you only have one mitochondrial ancestor whereas in nuclear DNA the number of ancestors increases by a factor of 2^n, where n is the generation number. If you look back one generation, you have two nuclear DNA ancestors, your parents. If you look back two generations, you have four nuclear DNA ancestors, your grandparents. Ten generations back the number is 1,024, and so on. However, no matter how far back you go, you only have a single

mitochondrial ancestor in that generation. This direct inheritance of mtDNA led to the idea that all humans alive today had a single common mitochondrial ancestor at some point in the dim past. Since this ultimate common ancestor necessarily had to be female, the popular press seized upon the attention-getting name "**Mitochondrial Eve**."

Headlines aside, the lack of recombination and the ability of mtDNA to accumulate mutations at a high, neutral rate in the D-loop and at lower but still accelerated rates elsewhere in the mtDNA genome, led to a serious effort to use this information to estimate where and when a putative common ancestor of all *Homo sapiens* lived. The approach used is one common in evolutionary genetics. Assume that three individuals are found to have the following nucleotide at an arbitrary position; I1 = A, I2 = G, and I3 = A. Further assume that the common ancestor of all three had a G. From this it is possible to generate three possible phylogenetic trees (Figure 2). Which tree is right? Tree 1 requires that there are two G→A mutations to account for the observed pattern. This is also true of Tree 2. Tree 3, on the other hand, requires only one G→A mutation to explain the pattern. Since two mutations occurring at the same site is far less likely than one, the third tree is considered to be a more parsimonious answer. If enough such data can be assembled for various unrelated individuals the trees become more and more statistically relevant and non-redundant.

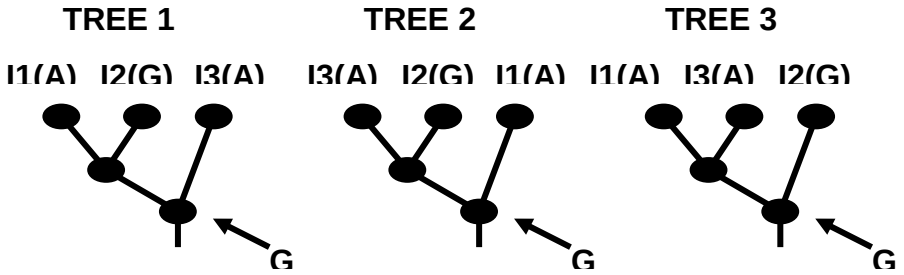

Figure 2. Three phylogenetic trees based upon DNA sequence data. Figure adapted from [12].

Mitochondrial DNA and Recent Human History

Mitochondrial DNA is unique, it is only passed on from your mother and it is passed on intact. By contrast, your nuclear DNA is a mixture of the nuclear DNA of your father and the nuclear DNA of your mother. That is, you inherit one-half of your genes from each of your parents. During meiosis, the process of sperm and egg formation, one-half of the genes in your parents' nuclear DNA are left behind. Before it is halved, however, there is a great deal of mixing of DNA that occurs. This is called recombination and it is the reason why we don't all look alike. Your mtDNA is free of this mixing. Your mtDNA is exactly the same as your mother's, which is exactly the same as her mother's, which is exactly the same as her mother's, and so on. This is not to say that changes do not occur. Mutation occurs in mtDNA and it occurs at a substantially higher rate than in

nuclear DNA. So, if you could look back in time you can find direct ancestors who may have had a difference or two in their mtDNA.

If you look at mtDNA sequences for two presumably unrelated people and if those sequences are the same, then those two people shared a common female ancestor at some unknown time in the past. However, if those same two people have identical mtDNA sequences in all but one or two positions, it is possible to use the mtDNA mutation rate to estimate how long ago they shared a common ancestor. This is what Anthropological Geneticists have done with mtDNA sequences from thousands of living people from all over the world. What they have found is astounding! Differences in world-wide mtDNA sequences form into distinct clusters, or lineages. Moreover, there are only nineteen of them. Antonio Torroni and colleagues have assigned letters to each of the groups. They are A,B,C,D,F,G,H,I,J,K,L,M,N,U,V,W, and X [13, 14, 15]. Even more important is that certain lineages are found only in certain parts of the world. Lineages H, I, J, K, T, U, V, and W are only seen in Europe, lineages L, N, and M in Africa and the Middle East, and lineages F and G in Asia. Lineages A, B, C, and D are found in Asia and in Native Americans while lineage X is found in Europe and in a small number of Native Americans.

In 1987 Rebecca Cann, Mark Stoneking, and Allan Wilson carried out a remarkable analysis. Mutations in mtDNA occur more or less at random and at a fairly constant rate. Cann, Stoneking, and Wilson began to compare mtDNA lineages and, using the known mutation rate, were able to estimate how long the lineages had been separated from each other. What they found was that, if you worked backwards, collapsing lineages mutation by mutation, you arrived at a common ancestral mtDNA lineage for every human being on Earth. This ancestor, the “mother of humanity” lived in Africa only 150,000 years ago [16]. This result has not only been confirmed by repeated analyses, the lineages that descended from Africa have been dated. The European lineages H, J, K, T, U, V, and X arrived in Europe 40,000 to 50,000 years ago, the Asian lineages A, B, C, D, F, and G arrived there between 60,000 and 70,000 years ago, and the Native American lineages A, B, C, and D arrived in the New World between 26,000 and 34,000 years ago (Figure 3).

How do we know that these statistically derived dates are correct? First of all, the estimated ages match the archeological record very well. Second, DNA sequence analyses from the human Y-chromosome, the one part of the nuclear genome that behaves like mtDNA except it is inherited exclusively from male to male, gives the same results. Even so, this is just more statistical inference. It is too bad that ancient peoples don’t leave mtDNA behind along with their artifacts. Or, do they? Bryan Sykes, a British molecular biologist, first showed in 1989 that ancient bones and teeth did contain mtDNA and that it could be sequenced. In the late 1990’s Sykes sequenced mtDNA from a well-dated human remain that had been found in a cave in England. This person had lived 9,000 years ago. A comparison of this ancient mtDNA to sequences on file turned up an identical match in an Englishman who lived not ten miles from the cave. Not only

had that exact mtDNA sequence been handed down for 45,000 generations, it had been handed down intact! The absence of mutation helped to verify the molecular clock for mtDNA which was set to one mutation every 10,000 years.

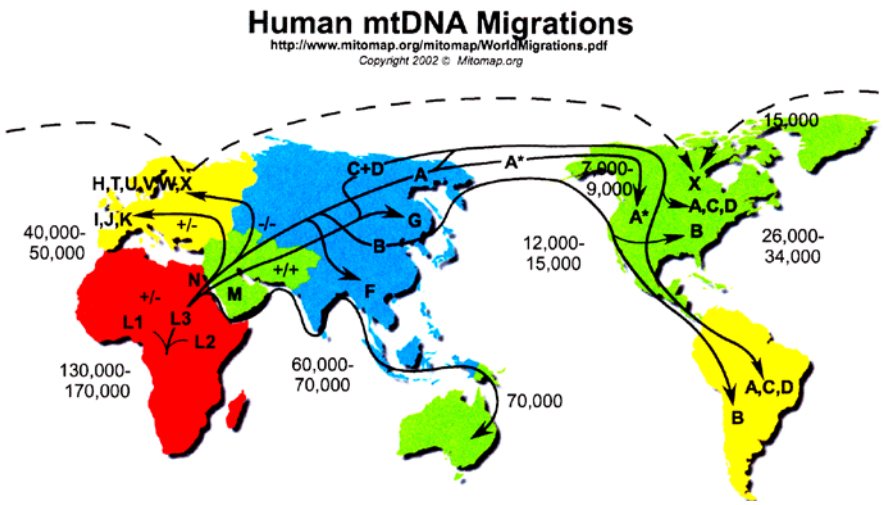

Fig. 5. Estimated migration routes and ages of the human mitochondrial DNA lineages. Source: www.mitomap.org.

Subsequent refinements of mtDNA techniques and the addition of thousands more samples culminated in the recent elucidation of the ages and geographic origins of the seven European core lineages. These are shown in Table 2.

Table 2.
"The Seven Daughters of Eve"

Lineage	Est. Age	% of Europeans	Geographic Origin
U	45,000yrs	11	Southern Greece
X	25,000yrs	6*	Caucasus
H	20,000yrs	47	Southern France
V	17,000yrs	5	Pyrenees (Spain)
T	17,000yrs	9	Appenines (Italy)
K	15,000yrs	6#	Dolomites (Italy)
J	10,000yrs	17	Mesopotamia

*Lineage-X is also found in Circumpolar peoples and Native Americans.
#Includes Oetzi, the Ice Man.
Source: [17].

While it is Oetzi, the IceMan, and the so-called Cheddar Man from England that get the headlines when it comes to ancient mtDNA analysis, there have been many studies of

ancient human remains from elsewhere in the world that continue to verify and solidify the conclusions reached from earlier research. One important example is the work reported by Anne Stone and Mark Stoneking in 1998. Stone and Stoneking extracted and sequenced mtDNA from 108 prehistoric Oneota Indians associated with the Norris Farms archeological site in Illinois. This site dates to ~A.D.1200. Results from their study showed that these people had representatives of all four founding mtDNA lineages; A, B, C, and D. They also had a few examples of rare lineages suggesting that there was more variation in the New World than is seen today. That is to say, some lineages have become extinct. The most important result from the Norris Farms population is that the temporal relationships among the lineages confirm that migration of people to the New World happened only once, that "wave" lasted from about 37,000 to 23,000 years ago as had been suggested by contemporary mtDNA analyses, and that there was more lineage diversity among the migrants than survives today among Native Americans [18].

Beyond Anthropology

We have focused on the relationship between archeology and molecular biology but there is mitochondrial DNA in all animals and plants and this DNA can be extracted and used to answer many historical questions. Sica et al. extracted mtDNA from five equine skeletons found in the excavation of Pompeii (A.D. 79) [19]. They were able to demonstrate that only two of those skeletons were actually horses. The other three were from donkeys. Bar-Gal et al. have used mtDNA analyses from 5,000 to 10,000 year old animal bones to show that the domestication of goats in the Middle East occurred about 8,000 years ago and spread from there to other parts of the world in a very short time [20]. Finally, Deakin et al. are using mtDNA from plant seeds to determine the age and geographic origin of sorghum domestication in the Old World [21].

It is fair to say that there are few people who are unaware of the significance of molecular biology and, in particular, the polymerase chain reaction (PCR) and DNA sequencing, on medical research. The influence of these techniques has reached far beyond biomedical research. Here, you have seen how historical geology, archeology, and molecular biology have combined to produce a picture of the genetic relatedness of the entire world. It is also seen how molecular biology can be used to study the past in very great detail. It is not just biology that is becoming a molecular science but anthropology and history as well and there is much more work to be done!

Summary

When the sequence of the human mitochondrial genome was published in 1981, a companion piece in Nature by Borst and Grivell carried the title "Small is beautiful…" Given that there are hundreds, maybe thousands, of nuclear genes that are larger than human mtDNA and the that mtDNA from mammals in general has proved to be of great, and increasing, importance, that excusable hyperbole in 1981 is as appropriate today as it was then [22]. The mitochondrial genome is, and should be, a focus of research all on its own. It has relevance for medical and veterinary genetics, evolutionary genetics, and

population genetics of all species, especially humans. The mitochondrial genome is a pretty remarkable, albeit tiny, piece of DNA.

References

1. Boore JL. (1999) Animal mitochondrial genomes. Nucleic Acids Res, 27: 1767–1780.

2. Saccone C, Gissi C, et al. (2000) Evolution of the mitochondrial genetic system: an overview. Gene, 261: 153–159.

3. Margulis L. (1971) The origin of plant and animal cells. Am Sci, 59: 230–235.

4. Martin W and Muller M. (1998) The hydrogen hypothesis for the first eukaryote. Nature, 392: 37–41.

5. Muller M and Martin W. (1999) The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. Bioessays, 21: 377–381.

6. Conway DJ, Fanello C, et al. (2000) Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. Mol Biochem Parasitol, 111: 163–171.

7. Unseld M, Marienfeld JR, et al. (2000) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nature Genet, 15: 57–61.

8. Anderson S, Bankier AT, et al. (1981) Sequence and organization of the human mitochondrial genome. Nature, 290: 457–465.

9. Richter C. (1998) Oxidative stress, mitochondria, and apoptosis. Restor Neurol Neurosci, 12: 59–62.

10. Wallace DC, Brown MD, and Lott MT. (1997) Mitochondrial genetics. In: Rimoin DI, Connor JM, et al. (eds.). Emory and Rimoin's Principles and Practice of Medical Genetics. London: Churchill Livingstone, 277–332.

11. Jazin E, Soodyall H, et al. (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. Nature Genet, 18: 109–110.

12. Relethford JH. (2001) Genetics and the Search for Modern Human Origins. New York: Wiley-Liss.

13. Torroni A, Huoponen K, et al. (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics, 144: 1835–1850.

14. Torroni A, Bandelt J-H, et al. (1998) mtDNA analsysi reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet, 62: 1137–1152.

15. Torroni A, Bendelt H-J, et al. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet, 69: 844–852.

16. Cann RL, Stoneking M, and Wilson AC. (1987) Mitochondrial DNA and human evolution. Nature, 325: 31–36.

17. Sykes B. (2001) The Seven Daughters of Eve. New York: WW Norton.

18. Stone AC and Stoneking M. (1998) mtDNA analysis of a prehistoric Oneota population: Implications for the peopling of the New World. Am J Hum Genet, 62: 1153–1170.

19. Sica M, Aceto S, et al. (2002) Analysis of five ancient equine skeletons by mitochondrial DNA sequencing. Ancient Biomolecules, 4: 179–184.

20. Bar-Gal GK, Khalaily H, et al. (2002) Ancient DNA evidence for the transition from wild to domestic status in Neolithic goats: A case study from the site of Abu Gosh, Israel. Ancient Biomolecules, 4: 9–17.

21. Deakin WJ, Rowley-Conway P, and Shaw CH. (1998) A sorghum of Qasr Ibrim: Reconstructing DNA templates from ancient seeds. Ancient Biomolecules, 2: 117–125.

22. Borst P and Grivell LA. (1981) Small is beautiful—portrait of a mitrochondrial genome. Nature, 290(5806): 443–444.

Resources

http://www.mitomap.org/

http://www.mywiseowl.com/articles/Mitochondrial_DNA

http://www.jpac.pacom.mil/mtDNA.htm

http://www.artsci.wustl.edu/~landc/html/cann/

http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/CellularRespiration.html

