The Power od DNA in Unlocking Family Relationships

by Ugo A. Perego, PhD
Senior Researcher for Sorenson Molecular Genealogy Foundation (www.SMGF.org)
Scientific Consultant for GeneTree (www.GeneTree.com)

DNA and genealogy part I: The Y Chromosome

The Y chromosome (Ycs for short) is one of the 46 chromosomes that makes up a male's DNA. While females have two copies of the "X" chromosome (one inherited from the mother and one from the father), the male gender is determined by the presence of the Ycs received from their father alongside with one of the mother's X's. The Ycs is approximately 60 million bases long and, although this may seem a considerable amount of DNA, it is in reality one of the smallest chromosomes with a limitated number of known biological functions when compared to the 3.2 billion pieces of DNA found in the human genome.

Two peculiar characteristics of the Ycs make it a valuable tool in genealogical research:

a. It follows a straight paternal line, from father to sons (Figure 1), similar to the inheritance patter of surnames in many western cultures;
b. Sections of the Ycs do not mix with any of the other chromosomes and can be sequenced to create unique Ycs profiles called haplotypes.

In addition, specific locations on the Ycs have been linked with ancient migrations and expansion events and are geographically-limited. These mutations (called SNPs) have been organized in a structure similar to a paternal family tree called the World Y Chromosome Phylogeny (see www.isogg.org/tree and Figure 2). Scientists have been able to use this information to reconstruct the paternal history of humankind (Figure 3).

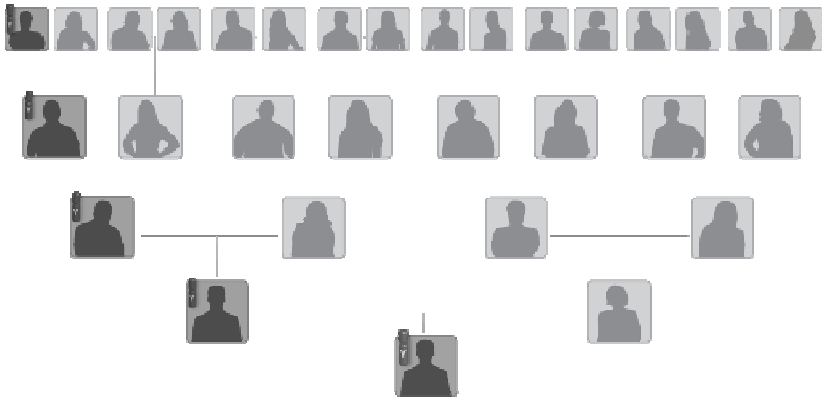

Figure 1 – The Ycs inheritance pattern

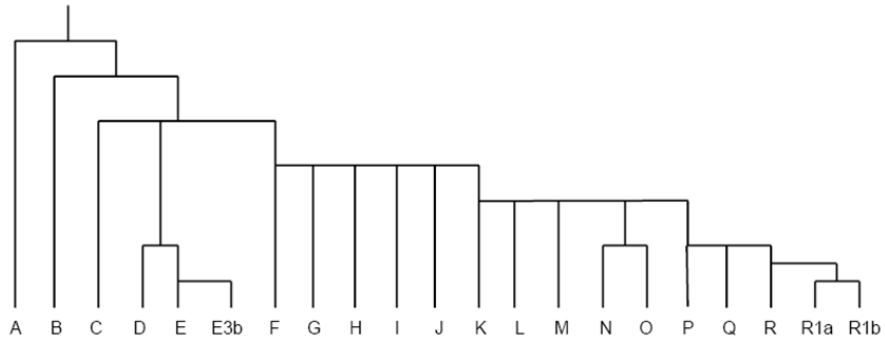

Figure 2 – The schematic Ycs world tree. Geographic distribution of each Ycs branch is shown in Figure 3.

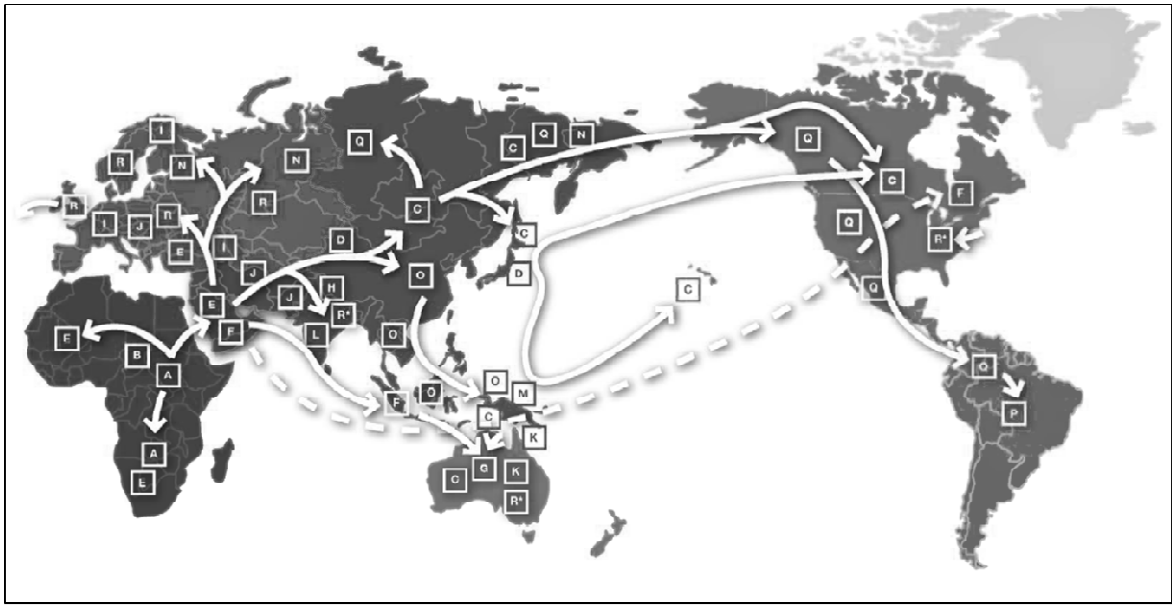

Figure 3 – MtDNA world migrations

Y chromosome in family history

This is what a typical Ycs profile would look like:

DYS385	DYS388	DYS389B	DYS389I	DYS390	DYS391	DYS392	DYS393	DYS394/19	DYS426	DYS437	DYS438	DYS439	DYS441	DYS442	DYS444	DYS445	DYS446	DYS447	DYS448	DYS449	DYS452	DYS454	DYS455	DYS456	DYS458	DYS459	DYS460	DYS461	DYS462	DYS463	DYS464	GGAAT1B07	YCAII	YGATAA10	YGATAC4	GATA H4.1
11																										9					15 16				
13	12	16	14	24	11	14	13	14	12	15	12	12	14	17	12	12	13	25	18	30	30	11	11	17	17	10	11	12	11	24	17	10	23	15	23	21

The set of locations tested and the corresponding numeric values are called your Ycs haplotype, or your Ycs profile. These values define you and your direct paternal line (past, present, and future). This type of testing is based on a set of Short Tandem Repeats (STRs), which varies from male to male. Ycs haplotypes can then be compared between individuals to prove, disprove, or identify a common paternal ancestor. Additionally, individuals can be tested for Single Nucleotide Polymorphisms (SNPs), which will assign a male to a specific group of Ycs with a known geographic origin and distribution (see Figures 3 and 4). The combination of STRs and SNPs testing may reveal valuable genealogical and historical information about a person's paternal line. Popular examples of such applications are recent studies involving the Ycs of descendants of Irish and Scottish royal lines, Genghis Khan, and others (see http://en.wikipedia.org/wiki/List_of_haplogroups_of_historical_and_famous_figures).

NOTE: some of the Ycs markers are reported differently among labs. A conversion table is available at http://www.smgf.org/ychromosome/marker_standards.jspx.

Some useful vocabulary

- DYS = DNA Y-chromosome Segment

- *Locus* (plural *Loci*) = Location on the Y chromosome (IE DYS388)
- STR = Short Tandem Repeat (IE ATT ATT ATT ATT ATT ATT ATT ATT)
- Allele value = number of repeats at a particular locus. In the example above, the repeat "ATT" is repeated 8 times. Therefore the allele value for that locus is 8.
- SNP = Single Nucleotide Polymorphism (IE Individual #1: ACTGAACCT; Individual #2: ACTTAACCT → The SNP, or mutation, took place at position 4: one individual has a G and the other one has a T at the same location).
- Haplogroup = a group (or family) of haplotypes sharing similar characteristics.

Public Ycs databases

- *Ancestry.com* – www.dna.ancestry.com
- *GeneBase* – www.genebase.com
- *GeneTree* – www.genetree.com
- Y-STR haplotype reference database – www.yhrd.org
- *YSearch* – www.ysearch.org
- *Sorenson Molecular Genealogy Foundation* – www.smgf.org

DNA and genealogy part 2: Mitochondrial DNA

Mitochondrial DNA (mtDNA for short) is a tiny circular genome of about 16,000 bases (Figure 4) found outside the nucleus in structures known as mitochondria (the cell's powerhouse for the production of energy). Among other applications, mtDNA is a powerful and accurate genealogical and ancestral tool as 1) it does not "mix" with the DNA in the nucleus, and 2) it follows a straight inheritance pattern along the unbroken maternal line. The first complete mtDNA sequence became available in 1981 and it was later re-sequenced in 1999. This sequence is known as rCRS and it is used as the standard against which every subsequent mtDNA sequence is compared against. Therefore, when a person receives his or her mtDNA profile (called a haplotype), the set of values reported are simply a list of all the differences from the rCRS. Nowadays, such tests are widely available commercially and are fairly inexpensive. A person may choose to have a segment of the mtDNA sequenced (usually an area called the Control Region of about 1,100 bases) or to sequence the whole mtDNA genome. Figure 5 is an example of a person's mtDNA Control Region profile, called a haplotype. NOTE: both males and females carry mtDNA, but only females will pass it on to the next generation! Therefore, mtDNA provides a recent and ancient history of mankind from a female standpoint (Figures 6, 7 and 8).

Some Useful Vocabulary

- Haplotype: your set of mutations as compared to the industry standard (rCRS).
- Control region: a segment of the mtDNA genome that contains the hypervariable segments known as HVR1, HVR2, and HVR3 (~1100 bases total).
- Full mtDNA sequence: the complete sequence of your mtDNA genome (~16,569 bases).

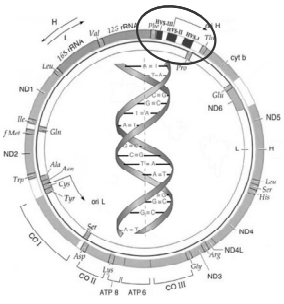

Figure 4 – MtDNA genome

Your test results for the range 16024 to 16569 and 1 to 576 are:

location	16224	16311	16519	73	189	195	263	315.1	497
your sequence	C	C	C	G	G	C	G	C	T
reference	T	T	T	A	A	T	A	--	C

Figure 5 – A typical mtDNA Control Region haplotype

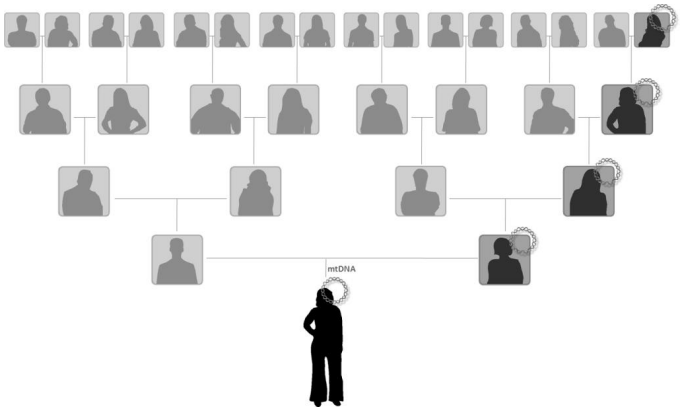

Figure 6 – MtDNA inheritance pattern

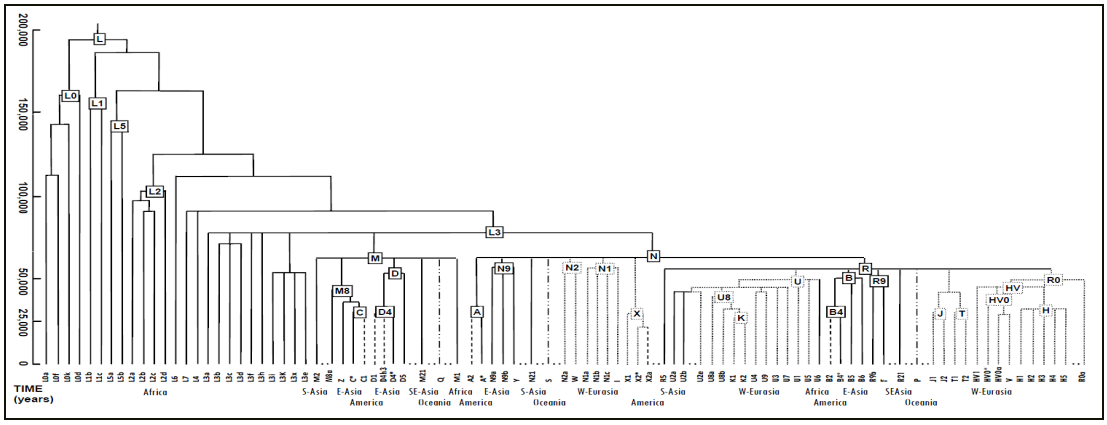

Figure 7 – The mtDNA world tree

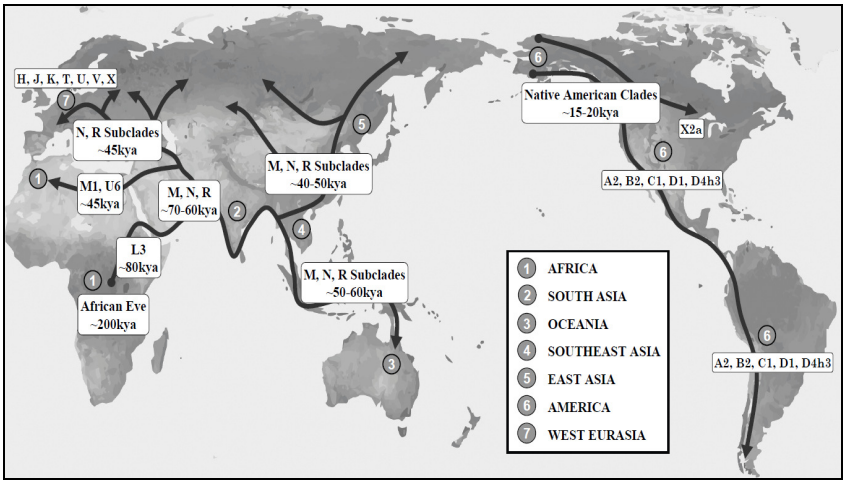

Figure 8 – MtDNA world migrations

Mitochondrial DNA and family history

The benefits from employing mtDNA in genealogical research are quite obvious. Since the inheritance pattern is straightforward (along the maternal line) and since the genetic information from this small genome does not undergo mixing with other genetic components found in the human cell, a person's mtDNA profile is a true genetic legacy of his or her maternal history. In addition, the ability to link an individual's mtDNA fingerprint to a specific population or geographic region, two individuals believing to have a common female ancestor may compare their mtDNA information as additional proof of such a relationship. This process can be done in two ways:

1. Both individuals purchase their mtDNA profile from an accreditated commercial laboratory such as GeneTree.com, FamilyTreeDNA.com, Ancestry.com, etc. and then they would manually compare the data with each other. If the profiles are identical, then they are most likely related along the maternal line. If the profiles are different, then there is no shared biological relationship on that part of the family tree. NOTE: not all laboratories sequence the same length of the mitochondrial genome. So when choosing a test, in addition to the price, it is important to check how much data is provided. Complete mitochondrial sequences will offer more data and would be more accurate for comparison purposes. However, fewer individuals have such data available and therefore most comparisons are done on the Control Region segment only.
2. An individual purchases his or her mtDNA profile and then uses it to query a public database for possible matches. A list of such databases is given below. NOTE: Currently, there are considerable limitations for individuals with a complete mtDNA sequence. FamilyTreeDNA.com does internal queries and contacts their customers when meaningful matches are found. Another database of complete sequences is the scientific repository GenBank (http://www.ncbi.nlm.nih.gov/genbank), but it is not set up for queries and only experts in the field are usually able to retrieve the necessary information.

Public mtDNA databases

- *Ancestry.com* – www.dna.ancestry.com
- *GeneBase* – www.genebase.com
- *GeneTree* – www.genetree.com
- *MitoSearch* – www.mitosearch.org
- *Sorenson Molecular Genealogy Foundation* – www.smgf.org

Additional services offered at www.GeneTree.com:

- Pre- and post-test consultations (Ycs, mtDNA, and autosomal from any testing company);
- Specialty testing (extended Ycs haplotype, Ycs SNPs, autosomal STRs, and X chromosome STRs);
- FREE genealogical and genetic utilities and tutorials;
- FREE social networking for genetic genealogists.

Follow us: