Tracing History Through Single DNA Molecules

Two DNA molecules in the cells of humans and other mammals have very distinct patterns of transmission. One of these is the DNA molecule present in the mitochondrion, an intra-cellular organelle responsible for transformation of energy to fuel cellular processes. Because mitochondria are provisioned by a female to the egg during its production, and not normally contributed by a fertilizing sperm, the mitochondrial DNA (or mtDNA) of each individual is inherited maternally. Your mtDNA in each of your cells was inherited from your mother, which she inherited from her mother, and so on. Inheritance of mtDNA traces the series of mothers in your single, continuous maternal line, or matriline. Modern assisted reproductive technologies are resulting in novel transmission patterns for mtDNA. Each difference in the sequence of bases between the mtDNA molecules of any two individuals is a consequence of a mutation in one descendant lineage since the two individuals last shared a common maternal ancestor. Mutations in the mtDNA base sequence shared among individuals provides evidence of common ancestry.

In contrast to the maternal transmission of the mtDNA, Y chromosomes are paternally transmitted - but only from fathers to their sons. The Y chromosome contains a gene that controls the initial genetic switch directing development toward the male pathway. Thus, a male inherits his Y chromosome from his father, who inherited it from his father, and so on. The Y chromosome is only present in males and traces the single, continuous paternal lineage, or patriline. Just like mtDNA, changes in the Y chromosome shared between males indicates common ancestry.

Both the mtDNA and the Y chromosome were utilized early on in the DNA era of molecular evolutionary studies as excellent recorders of human history. Allan Wilson's lab at UC Berkeley initiated during the late 1980's analyses of mtDNA data to examine the origin and diversification of modern humans. Findings from these studies are nicely summarized in a Scientific American article originally published in 1992 and published in revised form in 2003. Studies of both mtDNA and Y chromosomes indicate that modern human diversity descends from common maternal and paternal ancestors (the metaphorical "Eve" and "Adam", respectively) that lived in Africa about 150,000 years ago, because the oldest existing lineages persist today among individuals of African ancestry. Furthermore, individuals with pre-colonial ancestry outside of Africa descend from a more recent common ancestor reflecting a migration about 72,000 years ago of humans out of Africa.
Comprehensive studies of mtDNA sequences obtained from individuals distributed across the globe reveal an essentially long period of evolutionary history on modern humans in Africa prior to migration of the M and N lineages out of Africa.

Personal genetic analysis of your mtDNA is most informative about your deep evolutionary ancestry. Analysis of your mtDNA can determine where your mtDNA fits within the larger phylogenetic context of human mtDNA variation. Family Tree DNA offers services to determine the complete sequence of the mtDNA, or only the sequence of the hypervariable regions (HVR1 & HVR2). Matches are reported listing other customers in their database with identical mtDNA sequences or customers successively different by 1, 2, 3, etc. mutations. Mutations are most often present in the relatively short HVR1 & HVR2 within the “d-loop” control region involved in replicating or copying the mtDNA. As the name implies the hypervariable regions accumulate mutations at a higher rate than most other regions of the mtDNA, which largely encode products essential to cellular functions. Even within this most labile region of the mtDNA, a new mutation occurs approximately once every 87 transmission events (Sigurðardóttir et al. 2000). Thus, most individuals matched with identical, or nearly so, mtDNA sequences are unlikely to have family records that extend back 500-plus years to include their common maternal ancestor. In contrast, if historical records indicate individuals descend from a common matriline, no matter how distant the connection, the mtDNA of two collateral descendants should have identical (or nearly so) sequences.
Y Chromosome Ancestry Analysis

Comprehensive surveys of Y chromosome diversity reveal the phylogenetic structure present among Y chromosomes and provide a reference for a nomenclature identifying haplogroups. In contrast to the small mtDNA molecule, the Y chromosome is much larger and amenable to a variety of tools that assess genetic variability among individuals. Variants in DNA sequence (Single Nucleotide Polymorphisms or SNPs) of the Y chromosome represent one source of comparison; however, a number of locations along the DNA sequence of the Y chromosome have been characterized that contain Short Tandem Repeat (STR) markers. STR variants exhibit different numbers of sequential repeats of a short sequence, such as CA, so differ in length of the DNA sequence at these locations. These differences can be used to identify common ancestry at a bit finer resolution than mtDNA. The Y-DNA 37 marker test from Family Tree DNA provides good resolution of matches within their database. Still, most individuals matched with identical, or nearly so, profiles at 37 markers are likely connected by a common ancestor that pre-dates the individuals’ genealogical records.

Phylogenetic structure among Y chromosomes is also recognized by shared changes in bases at SNPs within DNA sequence. Many SNPs have been characterized and used to classify Y-chromosome haplogroups. Determining the bases present at these SNPs on an individual Y chromosome reveals its placement among the reference haplogroups. Services are also available for an individual to obtain the DNA sequence of a region of the Y chromosome. New sequence variants are discovered with these data, and these results are used to resolve relationships within terminal haplogroups. These fine-scale studies are primarily conducted by DNA project administrators working with individuals that volunteer samples for analysis, such as in a surname study.

Haplogroup Assignment from SNP Data

The genome-wide assessment of DNA sequence variation through SNPs by 23andMe and The Genographic Project includes sites within the mtDNA and Y chromosome that define the different haplogroups. Bases present at these sites are used to determine the mtDNA haplogroup of all customers, and for males the Y chromosome haplogroup is also determined. 23andMe also reports famous people included in each haplogroup - so customers can associate with someone with which they may share a 100x great grandmother.

Specialized Services

Other specialized services (e.g., African Ancestry & AfricanDNA) have appeared recently that advertise to determine relationships of customers to African populations based on the analysis of mtDNA and Y chromosome DNA. This is the same analysis as placement of an individual within the tree of human variation, except that presumably the services compare the individual to an extensive database of reference individuals from different regions of Africa. It is important to recognize that this specialized test of mtDNA would be of little use if a person's matriline includes an individual of European ancestry, and the same for assessment of the Y chromosome if the patriline includes a European male.
### Links to Additional Information

<table>
<thead>
<tr>
<th>Topic</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial DNA</td>
<td></td>
</tr>
<tr>
<td>mtDNA and Mitochondrial Diseases</td>
<td></td>
</tr>
<tr>
<td>PhyloTree mtDNA Tree</td>
<td></td>
</tr>
<tr>
<td>Maps of mtDNA Haplogroup Distributions</td>
<td></td>
</tr>
<tr>
<td>Y Chromosome</td>
<td></td>
</tr>
<tr>
<td>ISOGG Y Chromosome Tree</td>
<td></td>
</tr>
<tr>
<td>Maps of Y Chromosome Haplogroup Distributions</td>
<td></td>
</tr>
<tr>
<td>Comparative Timeline of mtDNA and Y Chromosome Diversification</td>
<td></td>
</tr>
</tbody>
</table>

### Reading List


Surname Studies
Surname studies represent a straightforward phylogenetic application of single molecule inheritance to investigate a familial characteristic. Because in many cultures the family name of the father is commonly taken by children, the inheritance of this surname should follow the inheritance of the Y chromosome. In the simplest scenario, if a given surname originated only once and all biological relationships strictly followed the paternal transmission of the surname, the Y chromosome of all men with that surname would trace their ancestry to a single common man who identified with that surname. However, things are a bit more messy in the real world. One simple complexity is that men in different families independently identified with the same name at the inception of the surname era. Thus, common surnames usually have many familial origins. The other complexities are that naming conventions do not capture the underlying biology of adoption and infidelity. In spite of these complexities, some surnames do exhibit a strong affinity to Y-chromosome transmission (Jobling & King 2009).

Given a long-term culture of paternal transmission of the surname, a living man that tests his Y chromosome would expect to find other closely related males with the same, or variant, surname. For example, the illustration below shows male descendants of two Irish immigrant brothers, Patrick and Thomas McAllister, that arrived in the US about 1848 during Irish Potato Famine. The Y chromosome in males descended from these brothers would be expected to be identical, or nearly so, and more similar to the Y chromosome in males that also have variants of the MacAlistair surname than males with other surnames.
Things did not quite work out as expected in the case. The Y chromosome in male descendents of the two brothers Patrick and Thomas McAllister are indeed similar, exhibiting only a single difference among 37 STR markers examined. However, the profile of their Y chromosome is most similar to profiles present in men with various forms of the Irwin surname. These results indicate that while Patrick and Thomas may share a common McAllister paternity, at some more distant point in the genealogy there was a transition in the surname from Irwin to McAllister through a Non-Paternity Event (NPE).