Genetic Genealogy Demystified: Understanding and Interpreting Your DNA Results

What is Genetic Genealogy?

DNA is the "gene" in genealogy. We all share some amount of DNA with our ancestors and relatives. Genetic genealogy is a tool to use in conjunction with traditional genealogy research that enables us to compare a small portion of our DNA to other people to determine if we're related.

Genetics Basics

23 pairs of chromosomes
- 22 pairs = autosomes (these determine who we are, what we look like, etc)
- 1 pair = gender chromosomes (X,Y). Women have XX, men have XY.

Mitochondrial DNA (mtDNA)

Y chromosome is inherited by men from their direct paternal line (their father's father's father etc).
mtDNA is inherited by both men and women from their direct maternal line (their mother's mother's mother etc).

The autosomes and X chromosome are a combination of DNA received from both parents, who also received a combination from their parents, etc. So the inheritance path can be a zigzag line through your ancestry rather than only a direct paternal or maternal line.

DNA is made up of four bases: Adenine (A), Cytosine (C), Guanine (G), Thymine (T)

Mutations (changes) in the DNA happen over time. Mutations happen randomly, but always happen between a parent/child pair. Mutations are what enable us to do genealogy and ancestry studies. They're needed to differentiate us from people who we're not closely related to, while matching us with people who we are closely related to. Genetic genealogy uses two different types of mutations: STR and SNP.

STR – Short Tandem Repeat
- DNA sequence of short repeating segments. Ex: **GATC**GATC**GATC**GATC**GATC**GATC**GATC**
- Great for genealogy. Family Tree DNA's Y chromosome "markers" are STRs.

Alleles – Number of repeats at an STR marker. Example above has 7 alleles (GATC repeated 7 times).
- Ex: DYS393=12 means you have 12 alleles (repeats) at marker DYS393 on the Y chromosome.

SNP – Single Nucleotide Polymorphism
- Base change at a single location in the DNA: Ex: C -> G, or A -> T
- Rare occurrences, great for deep ancestry studies.

Haplogroup – Broad division of the human population, defined by SNPs.
- Represents deep ancestry.
- Can be divided into subgroups, known as subclades.
- Separate Y chromosome and mtDNA haplogroups.
- Haplogroups are hierarchical, similar to an outline.
- Haplogroup names start with a capital letter, then alternate numbers and lower case letters.
- Ex: J1e. J is the main haplogroup name, J1 is a subclade of J, and J1e is a subclade of J1.

Y-DNA

Contains 60 million base pairs.
Genetic genealogy companies currently do not offer a product that tests the entire Y chromosome – expensive!!
Family Tree DNA offers test panels of 12, 25, 37 or 67 STR markers on the Y chromosome.

Y-DNA matches don't necessarily have to be exact to be significant. Since mutations happen randomly -- always between a parent/child – a father and son can have a mismatch on a marker. So relationship predictions are always provided in terms of probabilities.

Example: Two people who match on 36 out of 37 markers have a 90% probability of being related within 8 generations (meaning from 1 to 8 generations). On the flip side, in approximately 10% of the cases, a 35/37 match may be related more than 8 generations ago.

Locus	1	2	3	4	5	6	7	8	9	10	11	12
DYS#	393	390	19	391	385a	385b	426	388	439	389-1	392	389-2
Male 1	14	24	14	10	17	19	11	12	12	12	11	30
Male 2	14	24	14	10	17	18	11	12	12	12	11	30

In the above example, Male 1 and Male 2 match on 11 out of 12 markers. At DYS385b, they have a mismatch. This example came from two people who are known to be recently related, but an 11/12 match could also exist between two people who aren't recently related. Again, that's where the probabilities come in.

An 11/12 match has a 90% chance of being related within 39 (from 1 to 39) generations. That's quite broad.

Compare that to the 36/37 probability above, and you can see that the more STR markers you test and match someone on, the more confident you can be that the match is due to a recent common ancestor.

37 markers and higher is best for identifying matches within genealogical timeframe (hundreds of years).

FTDNA provides a haplogroup prediction for everyone who tests at least 12 Y-DNA markers. Add-on SNP tests can be ordered to determine the subclade. Each level of a haplogroup subclade is defined by a different SNP.

Example: E1b1b1 is defined by a SNP named M35. E1b1b1c is defined by a SNP named M123.

mtDNA

Contains 16,000+ bases.

Family Tree DNA offers tests on two sections of the mtDNA known as HVR1 & HVR2, which cover slightly more than 500 bases each, and Full Mitochondrial Sequence (FMS), which covers the entire 16,000+ bases.

mtDNA is reported as differences from the Cambridge Reference Sequence (CRS).

Mutations (differences) from the CRS in mtDNA are always SNPs, mtDNA mutations are rare.

HVR1 & HVR2 are best for deep ancestry studies, while FMS is the best level for genealogy.

As opposed to Y-DNA, your goal is to find exact matches with mtDNA. There are a handful of SNPs that tend to be more volatile (mutate more frequently), so people who match at everything except a volatile SNP may still be considered an exact match. However, in most cases, you want to find a true exact match.

Haplo	HVR1	HVR2
K	224C,234T,311C,519C	
K	224C,234T,311C,519C	73G,114T,263G,315.1C,497T
K	224C,234T,311C,519C	73G,114T,263G,315.1C,497T
K1a1b1a	224C,234T,311C,519C	73G,114T,263G,315.1C,497T
K1a1b1a	224C,234T,311C,519C	73G,114T,195C,263G,309.1C,315.1C,497T
K2a2a	224C,311C,519C	73G,146C,152C,263G,315.1C,512C

In the above example, each row is a different person's HVR1 and HVR2 results. The first 5 people all have the same mutations in HVR1. However, you'll notice that the 5th row has a different set of mutations in HVR2 than rows 2, 3 and 4. So the person in row 5 probably isn't recently related to those in rows 2, 3 and 4 since they mismatch on HVR2, even though some of the HVR2 mutations are indeed the same.

For mtDNA tests, Family Tree DNA also does a test to determine the main haplogroup. So all of the above samples are confirmed to be in haplogroup. The subclade is only determined when the Full Mitochondrial Sequence (FMS) test is done. The rows with K1a1b1a and K2a2a listed have done the FMS, so they have a full haplogroup assignment. The members with K only may be in one of those subclades too, but it hasn't been determined yet since they didn't do the FMS test.

Autosomal (Family Finder)

Tests approximately 500,000 pairs of SNPs across the 22 pairs of autosomes. Each pair of SNPs is comprised of one SNP from your mother, and one SNP from your father. It's not possible to tell which SNP came from which parent just by looking at your own results.

IDB – Identical By Descent

Centimorgan (cM) – Unit of measurement along a chromosome, equal to a 1% chance that a SNP will have experienced recombination at a location in a single generation.

Family Finder clusters SNPs into blocks that are 50 to 100 SNPs long. Blocks are compared against other customers in the FTDNA database. Two blocks can be either "half-identical" or a non-match. Half-identical means that at least one SNP from every pair in the block matches between the two people.

Example 1: Person A has "CT" at a particular SNP, while Person B has "AC". They are half-identical because they have a C in common.

Example 2: Person A has "CC", while Person B has "TT". They are not a match at that SNP.

Family Finder uses linked segments (Haploblocks) to predict a cousin relationship.

Longer blocks typically identify a closer relationship. Family Finder requires the largest matching haploblock between two people to be at least 5 cM long to consider the two people a match. Once a 5 cM block is identified, the length of all matching blocks and total number of blocks are used to predict the degree of relatedness.

Challenges for Jewish Ancestry

Y-DNA: Ashkenazim tend to match people with a myriad of different surnames, even at the highest level tests. This is due to most Ashkenazi families having inherited surnames for only 200-300 years, while Y-DNA matches can go back 500 years or more. It's also due to having a relatively small founder population. So we're identifying cousins on our direct paternal line, but our paper trails typically don't go back far enough to document the relationship. Sephardim tend to have the opposite problem – they have fewer matches in the FTDNA database because there are fewer Sephardim who've tested and they're more diverse.

mtDNA: Matches at the FMS level may go back 500 years or more. So we are identifying cousins on the direct maternal line, but as with Y-DNA, our paper trails typically don't go back far enough to document the relationship.

Autosomal: Due to the small Ashkenazi founder population, we tend to be related to each other through many different ancestral paths, and we have many blocks that are common among the Ashkenazi population, even though we may not be recent cousins. This causes Family Finder to predict us as closer cousins than we really are. Family Tree DNA is working with the geneticists and mathematicians to refine the predictions for Ashkenazi and other groups that have small founder populations.

Make the Most of Your Genetic Genealogy Testing

1) Provide your most distant ancestor name and location, and surnames of your ancestors, where requested in your MyFTDNA account. This will help your matches know something about you before they even attempt to contact you, and vice versa.

2) Communicate! Don't be afraid to email your matches. Everyone who's listed on your match page signed the same release that you did, giving Family Tree DNA permission to provide their name and email address to their matches.

3) Join a surname project. If one doesn't exist for your surname yet, start one and invite others with your surname to take a Y-DNA test! This is the best way to be proactive about finding same-surname matches instead of waiting for them to show up in the database on their own.

4) Join a geographical project. Although Jews were certainly mobile, we also stuck close to our communities and had plenty of cousins who lived nearby, and certainly not all with our same surname.

5) Join a haplogroup project. This will help you learn more about the deeper ancestry of your direct paternal or direct maternal line.

6) Join the JewishGen DNA Discussion Group: http://lyris.jewishgen.org/ListManager

Existing Family Tree DNA customers can click the Join Projects link on their MyFTDNA page to find appropriate projects to join. Don't only go by the automated suggestions – those just look at your surname – be sure to scroll down and check out the geographical and haplogroup sections, as recommended above.

If you haven't ordered a DNA kit yet, you can visit JewishGen DNA Central to search for a Family Tree DNA project that interests you: http://www.jewishgen.org/dna/

All new DNA kit orders placed after clicking through JewishGen DNA Central will help support JewishGen ☺