Handout for "Y Chromosome and the SNPs STRs" by Marty Brady, May 2020

A chromosome is an organized package of DNA found in the nucleus of the cell. Humans have 23 pairs of chromosomes, 22 of which are autosomal and one pair of which is the sex chromosomes (X & Y). Men have an X and a Y chromosome. Women have two X chromosomes. So, only men have a Y chromosome.

The Y chromosome is one of the sex chromosomes. The SRY gene on the Y chromosome triggers the development of the male in the embryonic stage. The Y chromosome is 57,227,415 base pairs (bp) in length making it the second smallest chromosome. It contains only 63 genes. Compared to the X chromosome which is over 153 million bp in length and contains about 800 genes.

Autosomal DNA (atDNA) is scrambled through a process known as recombination. This scrambling prevents using atDNA for genealogical research back more than about 5 generations. For the most part, the Y chromosome does not go through the recombination process. The non-recombining region of the Y chromosome is referred to as the NRY or the MSY (male specific Y). This is the region that is tested in genealogical Y-DNA tests. It is the non-recombining characteristic of the Y chromosome that makes it particularly useful in genealogical research.

Without recombination the Y chromosome is passed down from father to son basically unchanged for thousands of years. It is therefore useful for research into deep ancestry. I said "basically unchanged" because if it were not for the occasional mutation, Y-DNA analysis would be of no genealogical value. Without mutation, all Y-DNA would be the same and we wouldn't be able to distinguish one man from another using Y-DNA. Y-DNA is also useful for Surname research and for adoptee research in searching for a biological father.

While there is tremendous potential for genealogical relevance, many Y-DNA test takers (myself included) do not receive much relevant information from Y-DNA tests. This is largely due to the fact that not many males have taken those tests, so the database for comparison is fairly small. If every male on the planet were to take a Big Y-700 test, not only would Bennett Greenspan be very happy, but the human haplotree (see definition below) would be relatively complete and all males would know their place in that tree.

Because a large portion of the Y chromosome consists of highly repetitive DNA, at this time, only about 40% of the Y chromosome is amenable to genealogical testing (about 23 million bp). Out of this 40% only about 15 million bp (or about 65% - 15/23) are analyzed in the Big Y-700 test (FTDNA). Someday maybe new technology (maybe Full Genome Corp.'s (FGC) long read Y test) will allow us to obtain more information from the Y, but for now, this is as good as it gets.

The First Type of Marker on the Y Chromosome are STRs (see definition of STRs below)

Genealogical testing of the Y began around the year 2000, which was long before autosomal testing was available for genealogy. The first tests only examined a few areas of Y DNA (i.e., 12 markers) so they had low resolution (ability to distinguish two test takers). The more markers that are tested, the greater the expectation that two test takers can be distinguished by the differences in their Y STR set. Each difference in repeat number contributes to the Genetic Distance (GD) assessment. GD is defined as the number of differences in repeat units between two test takers. However, not all STRs are alike. There are fast-mutating and slow-mutating STRs. Differences in the fast mutating STRs would be expected to show a closer match than differences in slow mutating STRs. So, a GD of 2 between a pair of test takers that differ on slow mutating STRs might not mean the same as a GD of 2 for another pair of test takers that differ on fast mutating STRs. However, using the Time Predictor (TiP) algorithm on the FTDNA match list, takes the rate of mutation of particular STRs into account.

FTDNA employs thresholds for matches of Y-DNA. However, you can't see the STR results of your matches. At 37 markers, only those with 4 or fewer differences will show as matches. At 111 markers, only those with 10 or fewer differences will show as matches. Therefore, it is possible that someone who does not show at the 37-marker level (i.e., GD = 5) will show as a match at the 111-marker level (i.e., GD = 9). The Big Y-500 claimed to provide at least 500 STRs. That means that the Big Y-500 provided an additional 389 STRs (although by a different technology). The Big Y-700 provides at least an additional 589 STRs. However, I chatted with someone at FTDNA recently and they said that the additional 589 STRs are currently extraneous information as they are not used for genealogical purpose. The system for analyzing these extra STRs has not been developed yet. The extra STRs contained in the Big Y-700 test may yield useful information in the future, but have not been used as yet.

There are several Surname Projects listed on the FTDNA website. Enrolling in a pertinent project can provide you with expert advice from the project administrator and also get you a discount in testing fees. Another reason to join a surname project is that the STR results of participants are reviewable which may allow you to see matches that did not meet the thresholds discussed in the previous paragraph. Also, the Project Administrator has the ability to group participants by relatedness which may allow you to make connections. Project Administrators can use the STR repeat numbers and frequencies to group participants. Look at Leo Little's spreadsheet for frequency data (also available through the ISOGG website). I recently noticed that I seem to have more similarities with the Carroll Surname Project members than the Brady Surname Project members. Interesting. If you share rare STR repeat values (alleles) with another man, it may mean that there is a more recent common ancestor with that man.

The Second Type of Marker on the Y Chromosome are SNPs (see definition of SNPs below)

SNPs determine the haplogroup. SNPs are very stable and do not readily back mutate. Even though there are four possible nucleotides, a SNP consists of a change from only one nucleotide to another. If you look at the SNP Index on the ISOGG website, you can see a listing of each SNP mutation and what nucleotide was changed. At one particular location, there is usually only one change. There is not a change from C to T in one man at the same site that there is a change from C to A in another man. If you want to test your SNPs, you can choose targeted SNP tests or go with the Big Y-700 test. Targeted SNP tests are less expensive, but you may waste time and money guessing what SNPs to test for. The Big Y-700 yields all known SNPs and can yield an average mutation rate of about 84 years, which is starting to get within a genealogical time frame. In the future, maybe 3rd generation (long read) sequencing technology will allow even more recent estimations.

The Big Y-700 test shows you which of the more than 290,000 known Y-SNPs you have. On FTDNA's website when it says "you are derived for a particular SNP," it means you have that SNP. The Big Y-700 also shows you which variants (SNPs) are private. A private variant means that you are the only test taker that has that variant. In order to be given a SNP name and placed on the Y haplotree, at least one other test taker has to be derived for that variant. One person having a variant does not define a "group." You are not a group of one. It is considered a "singleton" until at least one other person is shown to have that private variant.

The SNP naming system allows test takers to see where they sit in the Y haplotree (see slides for a description of the systems). The terminal SNP is the SNP that is the furthest out (furthest downstream) on the Y haplotree. As more men take the test, the tree will be further resolved and a man's terminal SNP will almost certainly change because a man's "Private" variants will almost certainly cease to be private. Entering your terminal SNP on the GeneticHomeland.com DNA Pedigree feature will allow you to see your history of SNP mutations back 200 to 300 thousand years ago to Y Adam. You can enter two terminal SNPs and see just where those two lineages diverge. As more SNPs are discovered, the Y haplotree is revised as seen in FTDNA's Big Y-700 white paper (see slides).

There are several other companies that can analyze your Y-DNA raw data for a small fee (YFull and Full Genome Corp. to name two). You can also post your raw Big Y-700 data on Alex Williamson's "Big Tree" using the Y-DNA Data Warehouse.

Definitions

Allele - any one of two or more genes that may occur alternatively at a given site (locus) on a chromosome.
Complementarity refers to the fact that each nucleotide has a base (A, C, T or G), and each base on a DNA strand can pair up with the appropriate (complementary) base (A with T and G with C) from the opposing nucleotide on the second DNA strand. Complementarity is maintained during DNA replication.
DNA Polymerase is the enzyme responsible for synthesizing new strands of DNA (i.e., making copies, replication).
DNA Polymerase works in one direction adding new nucleotides (dNTP) to the 3 position of deoxyribose backbone of the DNA strand.
Enzymes are protein molecules in cells which work as biological catalysts. Enzymes speed up chemical reactions in the body, but do not get used up in the process, therefore can be used over and over again. Almost all biochemical reactions in living things need enzymes.
Haplotype – A 111 marker STR test reveals the number of repeats at 111 individual locations on the Y chromosome. When these are all listed together, it is called a Y-STR haplotype. Haplotypes can be used to predict the haplogroup fairly accurately, but they are conservative predictions (i.e. the predictions don't go very far out on a limb…ha, ha)
Haplogroup – Related haplotypes belong to the same genetic family which we call a haplogroup. A haplogroup is a major branch on either the maternal or paternal tree of humankind. Haplogroups are further divided into subclades defined as originating with a SNP mutation.
Haplotree – The tree of humankind (either maternal or paternal) displaying the relationship of all known haplogroups and subclades.
ISOGG – International Society of Genetic Genealogists.
A nucleotide is the building block unit of nucleic acids such as DNA. When we talk about sequencing, we are talking about the sequential order of the different nucleotides as we progress down the DNA strand.
SNP: A SNP is a single nucleotide polymorphism. That means that it is a single small change in your DNA code. These changes are rare. Once they happen, they seldom change back (back mutate).
STR: An STR is a short tandem repeat. This is a place in your DNA code where a letter sequence is repeated. For example, AGTAAGTAAGTA is three repeats of the sequence AGTA. STRs have a fast mutation rate. Some STRs mutate faster than others. When they change, it is an increase or decrease in the number of repeats. STR values change back (back mutate) more commonly.
Terminal SNP - A terminal SNP determines the terminal (final) subbranch on the Y-DNA Tree to which someone belongs. It is the SNP that is the farthest out on a limb of the Y-DNA Tree (i.e., farthest downstream).

Links

www.nature.com/ncomms/2014/141202/ncomms6631/full/ncomms6631.html
https://yhrd.org/pages/resources/locus_information or the ISOGG website
FTDNA Learning Center
http://freepages.genealogy.rootsweb.ancestry.com/~geneticgenealogy/yfreq.htm
https://sites.google.com/site/wheatonsurname/beginners-guide-to-genetic-genealogy/lesson-14-more-with-the-y

Resources

Publications

"The Family Tree Guide to DNA Testing and Genetic Genealogy" by Blaine Bettinger

"Tracing Your Ancestors Using DNA, A guide for Family Historians" edited by Graham S. Holton
"FTDNA Big Y-700 white paper" on FTDNA blog site.
"The Y-chromosome Tree Bursts into Leaf: 13,000 High Confidence SNPs Covering the Majority of Known Clades" (numerous authors); Mol.Biol.Evol. 32(3):661-673
"A Brief Review of Short Tandem Repeat Mutation", Fan Hao, Chu JY., Genomics Proteomics Bioinformatics. 2007 Feb;5(1):7-14. Review.
"Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups" by the Y Chromosome Consortium, Genome Research, www.genome.org

YouTube videos

"How Y DNA Can Help Your One Name Study" by Dr. Maurice Gleeson
https://www.youtube.com/watch?v=vOx971zy6LI
"Using Y-DNA to Research Your Surname" by Dr. Maurice Gleeson
https://www.youtube.com/watch?v=u5xHtppMNJY
"Exploring New Y-DNA Horizons with Big Y-700" by Dr. Iain McDonald
https://familytreewebinars.com/download.php?webinar_id=1443