Y-DNA; How it Works, Why we need it and How Surname Groups Can Help

Y-DNA is passed down from father to son relatively unchanged. Meaning it does not undergo any significant recombination (i.e., shuffling or chopping up). Now, if Y-DNA didn't change at all, it would be of no use to the genealogist because every male would have the same Y-DNA and you would not be able to distinguish patrilineal lines. But every few generations on average, Y-DNA does add mutations that are specific to a patrilineal line. Because it doesn't change much, we can follow the Y-DNA mutations of a specific patrilineal line for thousands of years.

Mutations are created when mistakes are made during DNA replication. There are two primary types of mutations present in Y-DNA: STRs (**S**hort **T**andem **R**epeats) and SNPs (**S**ingle **N**ucleotide **P**olymorphisms). STRs can back mutate more readily than SNPs and mutate faster than SNPs. STRs are therefore not used to establish a haplogroup but can be used to predict a haplogroup at relatively low resolution (i.e., not very far out on a limb of the tree of mankind). STRs are also used for connections in more recent times. SNPs establish haplogroups. As more people test and more of the Y chromosome is analyzed, SNPs will also be used for more recent connections. But currently they are more often used for older associations.

Because STRs can back mutate, different patrilineal lines can converge. This can make the Time to Most Recent Common Ancestor (TMRCA) for two patrilineal lines appear more recent than it really is. This effect is most prominent when a smaller amount of the Y chromosome is tested. For example, convergence has a more profound affect with a 12-marker test than the 111-marker test. The 111-marker test is said to have greater coverage than the 12-marker test because it covers more of the Y chromosome. David Vance estimates that up to 70% of matches at the Y12 level may be due to convergence.

Family Tree DNA (FTDNA) provides a TiP (Time Predictor) feature to help you estimate the TMRCA for you and your match. Just remember that this is an estimate and therefore is not exact. Another thing to remember is that it is highly unlikely that all the mutational differences between you and a match will be on one patrilineal line (either yours or your matches). We can assume that on average half of those mutations are on your line and half are on your matches line and at a coverage of 111 markers (from Dr. Iain McDonald's chart) the average time between mutations is 120 years/mutation. So, if you and your match have tested at a coverage of 111 markers and you have a genetic distance of 6 (meaning 6 mutational differences between you), the projected TMRCA would then be 3 X 120 = 360 years ago.

A SNP is a single nucleotide polymorphism. That means it is a single small change in your DNA code. These changes are rare. Once they happen, they seldom change back. So, convergence is not an issue with SNPs. A SNP originates in one man and one man only (with extremely rare exceptions). For a mutation (aka variant) to be named and placed on the tree of mankind, at least two men in the database must have tested positive for the

variant. If only one man has the mutation, it is called a private variant and is not placed on the tree of mankind. You can find a list of named variants on the International Society of Genetic Genealogy (ISOGG) website. An explanation of the SNP naming prefixes is also on the ISOGG site. 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. As more and more men test, your private variants will be paired with other men and your terminal SNP will change and move closer to the current time. As of 13 October 2024, there were about 684,000 named variants in the FTDNA database, and it is growing rapidly. The Discover page on FTDNA provides a lot of good information about the Y haplogroup of you and your matches.

There are generally three periods of ancestry: deep ancestry, lineages, and genealogies. Surname projects usually go back to about 1000 CE (AD) when surnames started to appear. Haplogroup projects go back much further. SNPs are better for deep ancestry but can be used up to present times. The limiting factor with SNPs is the long distance between mutations and the relatively small number of males who have tested for SNPs (i.e., Big Y-700 test). This will change as more people test and more of the Y chromosome is covered. STRs are usually unreliable over more than 1500-2000 years. STRs are usually needed to fill in gaps between SNPs especially in genealogical timeframes.

A while back a genealogist named Leo Little compiled a chart of STR value frequencies for 7 haplogroups (E3a, E3b, G, I, J2, R1a and R1b) for the first 67 STRs. Leo has since passed away and I don't think anyone has continued his efforts. It would be nice if FTDNA would tabulate these frequencies for all Y haplogroups using their current database. These frequencies can sometimes give you a clearer picture of which surname group you belong in as they did for me.

The Big Y-700 test is so named because it tests for 700 STRs. FTDNA actually tests for 838 STRs, but some of those extra STRs don't yield conclusive results. FTDNA does, however, promise to give you at least 700 STR results. The first 111 STRs are determined by one methodology and the rest of the STRs are determined by Next Generation Sequencing (NGS) which yields more inconclusive results than the first methodology. I don't know how polymorphic the additional STRs are or if anyone is currently using the additional STRs for genealogical purposes. Maybe some clever Project Administrator could answer that question.

Maurice Gleeson has some excellent videos on using Y-DNA for investigating surnames and lineages. One that is particularly appropriate is his video on creating a Mutation History Tree (MHT) for Lineages. The link is in this handout and available on his blogsite. You can also download a PDF of his Gleeson Lineage MHT from his blogsite. His video provides many reasons for creating an MHT such as giving a clear picture of the evolution of the surname and defining the branching points over time. He also recommends developing an MHT manually and then checking your work with the SAPP program.

The SAPP (Still Another Phylogeny Program) program was developed by David Vance and Maurice Gleeson has also collaborated with Mr. Vance to improve the program. SAPP is available at https://www.jdvsite.com and is free to the public. Under FAQ the site provides much information about the program. There are links to at least 4 videos and I highly recommend watching all of them, but especially the Getting Started video. I decided to try out the program.

I recruited a cousin to investigate my 3X great grandfather Laurence O'Loughlin. My cousin had a match at 111 markers with a genetic distance of 6. This could place the MRCA of my cousin and his match at about 360 years ago. Which is worth investigating. I contacted the match and discovered that his ancestor lived in a part of Ireland about 13 miles from my ancestor. When I checked Griffith's Valuation for the area in between, I found several Loughlan's. I bought a Big Y-700 upgrade for my cousin's match but have not received the results yet. I decided to try SAPP anyway. I went to a surname project page on FTDNA and copied the STR data for a few members into an Excel spreadsheet.

Keep in mind that this is sensitive data and identifying information should not be distributed. After cleaning the data as suggested in Mr. Vance's video, I saved the file as a text file, added the function /STRDATA to the top and executed the SAPP program. SAPP can correct for potential convergence anomalies and generate a Y-DNA tree for the selected members. Mr. Vance also provides information about estimating the TMRCA and gives examples of the kind of information you get with a typical Y-DNA Surname Project group and a typical Y-DNA Haplogroup Project group.

Resources:

The Genealogist's Guide to Y-DNA Testing for Genetic Genealogy 2nd Edition by David Vance
The Family Tree Guide to DNA Testing and Genetic Genealogy by Blaine Bettinger
DNA and Family Tree Research: https://dnaandfamilytreeresearch.blogspot.com
Videos by Maurice Gleeson (esp. on mutation history trees):
https://www.youtube.com/c/DNAandFamilyTreeResearch/videos and
https://youtu.be/ZyGAid-EgsY
DNA Lecture series, especially by John Cleary:
https://www.youtube.com/channel/UC7HQSiSkiy7ujlkgQER1FYw
The Tree of Mankind Youtube video by Dr. Michael Sager
https://www.youtube.com/watch?v=f_IPIPVALnE
Exploring New Y-DNA Horizons with Big Y-700 a video by Dr. Iain McDonald
On Legacy Family Tree Webinars
Genetic Genealogy Tips and Techniques
https://www.facebook.com/groups/geneticgenealogytipsandtechniques
Other resources in the "Some Useful Links" section of the Vance Surname Project Results page:
https://www.familytreedna.com/groups/vance/about/results

David Vance Exploring Group Project Administration Videos
https://www.youtube.com/watch?v=OrXI5HvhYgM Part 1 (Subgroups: Organizing and Grouping Members (Y-DNA)
https://www.youtube.com/watch?v=qZQGy4AHdqo Part 2 (Analyzing a Project Subgroup (Y-DNA))
https://www.youtube.com/watch?v=iBHPeiEDfKs Part 3 (Recruiting New Participants (Y-DNA))
David Vance Youtube channel:
https://www.youtube.com/channel/UCae_YVCJhhLunuefx1mrw2A
How to Use the SAPP Tool: https://www.jdvsite.com
Ybrowse site: https://ybrowse.org/gb2/gbrowse/chrY/?
This database contains position, allele, and other information for over 2.5 million SNPs and is updated regularly.

Other resources:

SNP Tracker site: http://Scaledinnovation.com/gg/snptracker.html
Other tools at Scaled Innovation: http://scaledinnovation.com/gg/adminUtils.html
FTDNA Y Haplotree: https://familytreedna.com/public/y-dna-haplotree/A
Big Y, Yseq, NGS Facebook Discussion group:
https://www.facebook.com/groups/257810104756408
FTDNA Big Y Facebook Group: https://ww.facebook.com/groups/familytreednabigygroup
ISOGG Wiki: https://isogg.org/wiki/
SNP tests and information: https://YSeq.net
Big Y results are in Now What presentation:
https://hatcherfamilyassn.com/documents/Big-Y-results-are-in-Now-what.pdf