From Novelty to Necessity: Y-DNA Testing and Genealogy

Janine Cloud – Group Project Manager, FamilyTreeDNA
janinec@ftdna.com

The past 19 years have seen Y-DNA testing go from a novelty to an invaluable tool for tracing paternal ancestry, from 12 STRs to about 22 million base pairs from the non-recombining section of the Y chromosome, including thousands of Single Nucleotide Polymorphisms (SNPs), and at present, as many as 700 STRs. The exponential expansion of available Y-DNA data expanded the horizons for the use of Y-DNA testing in genealogical research and illustrated the limitations of previous tests in determining time to the most recent common ancestor (TMRCA).

This talk briefly explores the history of the use of Y-DNA for genealogy and the expansion of the Y haplotree before delving into the impact that Next Generation Sequencing (NGS) has had on the genealogical community and individual research.

Note: This syllabus material contains a lot of detailed information. Some of it may not make it into the final version of the presentation because of time constraints but you may find it useful.

Y-DNA Facts – the biological and technological basics

The first type of test available directly to consumers, Y-DNA results can confirm paternal relationships and point to a surname. Deeper analysis may reveal how closely (in generations) testers are related and from where the paternal line migrated. It cannot be used as a paternity test, however, because it can only confirm that two men share a paternal ancestor, not positively identify which ancestor is the father.

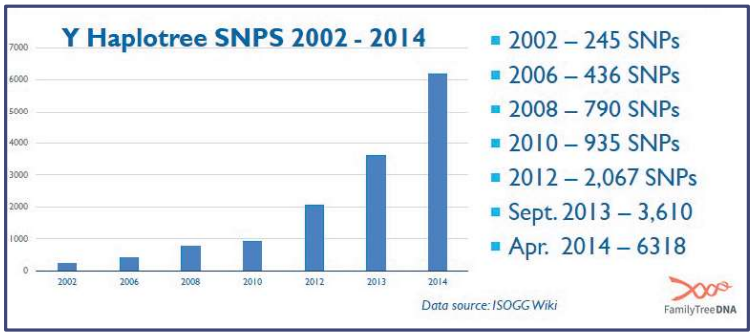
- Only males inherit from father.
    - Recombination limited to pseudoautosomal regions.
    - Genetic genealogy uses non-recombining portion of Y.
    - Used to match against testers to determine a common paternal ancestor.
    - Reveals migratory paths of paternal line.
- Types of mutations used:
    - STRs - Short Tandem Repeats.
        - Counts the number of times nucleotides repeat at specific locations.
        - Traditionally used for “close” matching to determine a common paternal ancestor.
        - Number of differences between two testers known as “genetic distance”, a measure of how closely testers may be related.

    - SNPs - Single Nucleotide Polymorphisms.
        - Change (mutation) from one base (i.e. G-C, A-T) at a single location.
        - Defines haplogroups and subclades showing positions on the Y haplotree.
        - Corresponds to migratory paths.
- Tests available:
    - 12-, 25-, 37-, 67-, and 111-marker STRs.
    - Individual SNPs.
    - SNP Packs - Curated groups of SNPs from a specific haplogroup or subclade.
    - Next Generation Sequencing (NGS) tests - reveal variations in Y chromosome unique to individual family lines, can differentiate among descendants of a specific ancestor.
        - Select portion of Y chromosome:
            - Big Y-500 (FamilyTreeDNA - no longer sold).
            - Big Y-700 (FamilyTreeDNA - from 1 Nov 2018).
            - Y-Elite (Full Genomes Corporation).
        - Exome (Full Genomes Corporation, Gene by Gene, Ltd.)
        - Whole Genome (Full Genomes Corporation, YSEQ, Gene by Gene, Ltd)
- Types of processing used for Y-DNA
    - Sanger sequencing
        - Developed by Dr. Frederick Sanger and associates at Cambridge, 1977.
        - Generally accurate and reliable, though slow and expensive.
        - Used in defunct products like Deep Clade and Walk the Y.
        - Now mostly used to either confirm Next Generation Sequencing results, or to fill in rare gaps in NGS coverage.
    - Mass Spectrometry
        - High throughput.
        - Confirmation (positive or negative) only, no discovery.
        - Inexpensive but time-consuming to design panels.
        - Not easy to change panels once created.
        - Can test dozens of SNPs.
    - Next Generation Sequencing (NGS)
        - High throughput.
        - Excellent for discovery since it sequences DNA fragments rather than testing preselected SNPs.
        - Tests millions of SNPs, hundreds of STRs.
        - Expensive, but best value for amount of data.
        - May miss some areas, not read enough of others, though updates have improved consistency.

Y haplotree – bringing order to discovery

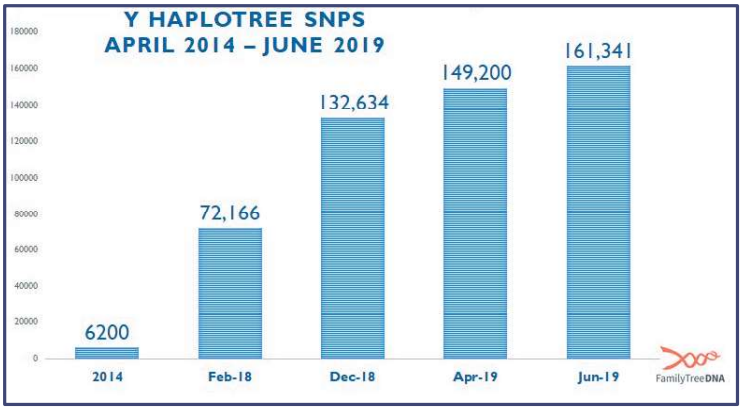
Since Y-DNA mutations form patterns reaching into the distant past, arranging SNPs phylogenetically, in a "tree" going from oldest to youngest, makes sense for displaying mutations and revealing migratory patterns. Combinations of letters for the main branches combined with the SNP name identify each tester's haplogroup, i.e., his position on the haplotree.

Since the release of Big Y in late 2013, FTDNA has analyzed over 32,000 Y chromosomes in ultra-high resolution, expanding the haplotree exponentially as seen in charts below.



1 Y haplotree growth 2002-2014

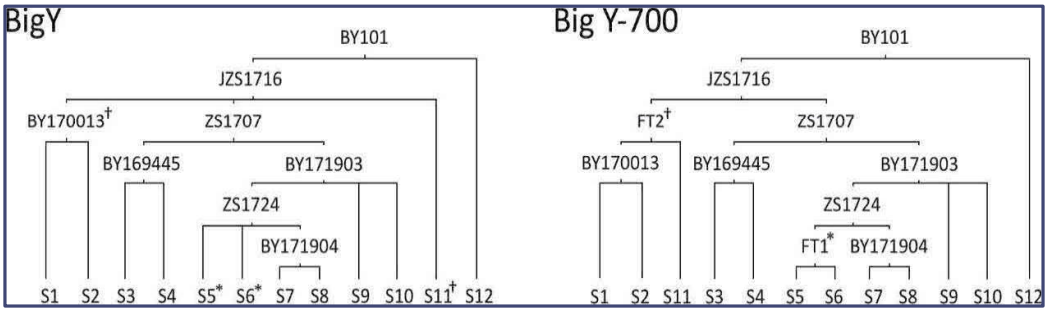
2 Y haplotree growth May 2015 -May 2018



3 Y haplotree comparison 2014, 2018, 2019

- SNP testing options
    - Individual SNPs.
    - SNP Packs – bundles of related SNPs.
        - Carefully designed panel of SNPs relevant to and encompassing a specific group (M343, L21, etc.)
        - Contain between ~30 -~175 SNPs.
        - Cost efficient method to move a tester down the tree.
        - Does NOT discover new SNPs.
    - NGS tests such as Big Y – advanced tests that sequences large section of the Y chromosome to reveal known SNPs as well as private variants.
- Big Y
    - Introduced November 2013.
    - Targeted non-recombining Y-DNA sequencing.
    - Illumina Novaseq platform.
    - Proprietary analysis technology.
- Big Y - 500
    - Introduced April 20, 2018.
    - No change in lab process from Big Y.
    - First 111 STRs from Y-111 product, added ~389 STRs called from Big Y data.
- Big Y - 700
    - Announced January 30, 2019, retroactive to November 1, 2018.
    - First 111 STRs still produced from Y-111 product.
    - Expanded NGS targeting of genealogically useful Y regions.
    - Revamped NGS chemistry for more uniform coverage within that region.
    - More consistent coverage improves comparison between any two results and allows more of the identified SNPs to be placed on the tree.
    - Additional 200 STRs chosen for ability to differentiate between lines.
    - Used 88 samples from diverse haplogroups, run on both Big Y and Big Y-700.
        - Compared the SNPs and STRs identified on each platform.
            - For STRs, confirmed 99.7% concordance between genotypes in Big Y and Big Y-700.
            - SNP coverage validation - of 87,816 unique SNPs in regions with between 200 -1000 variants called.
                - 51,417 (58.6%) identical between Big Y and Big Y-700.
                - 32,450 (37.0%) unique to Big Y-700.
                - 3,949 (4.5%) were unique to Big Y.
        - Findings predict Big Y-700 results will contain:
            - Fewer no-calls of branch-defining SNPs.
            - More unique SNPs per sample.
            - More new-branch-defining SNPs waiting to be discovered.

Big Y-700 Case Study



4 Comparison of results of Big Y vs. Big Y-700 under J-ZS1716

- From the 88-sample control group, 11 samples were chosen from haplogroup J-ZS1716 (TMRCA 1,000 - 1,500 years ago). Using Big Y-700 SNPs called in a minimum of 2 samples, plus J-BY101 sample as an outgroup to eliminate variants above that level.
    - Big Y showed 16 shared SNPs and 7 branches.
    - Big Y-700 retained all 16 shared SNPs from Big Y and produced 8 new variants for placement on the tree.
    - 6 of these 8 new variants proved to be equivalents to known branches, two were new branch SNPs.
        - The first SNP resolved relationships between J-ZS1724 samples S5 and S6, known by genealogy to be each other's closest match.
    - Big Y did not produce a unique SNP shared between them.
    - Big Y-700 did produce unique shared SNP, creating a new branch for these two samples (J-FT1).
    - The second SNP resolved relationships among samples positive for J-ZS1716.
    - Sample S11 appeared to be a lone member of a third branch of J-ZS1716 along with J-ZS1707 and J-BY170013.
    - Big Y-700 uncovered a new variant present in S11 and J-BY170013 but absent from members of J-ZS1707, forming a new branch upstream of J-BY170013 (J-FT2).

Group Projects

Surname projects, haplogroup projects and even geographical projects can provide Y-DNA testers context and additional clues about their genetic paternal lines. Volunteer administrators run group projects, usually to further their own research, and they analyze data to determine relationships among project members as well as origins of various Y lines.

Projects allow testers to pool resources and share information to which they might not otherwise have access.

- Types of Y-DNA Group Projects
    - Surname - studies the origins and variants of surnames
    - Haplogroup - studies an entire haplogroup or specific subclades
    - Geographical - studies the genetics of a region varying from a county or parish to a country or continent (also dual, with mtDNA)
- Y-DNA Projects and Your Research
    - Group Project Administrators provide:
        - Data comparison for grouping in project
        - Data analysis within matches and within project
        - Data integration with paper trail research
    - Administrators and project members help with:
        - Building the Y haplotree
        - Recruiting test candidates to further research
        - Financing tests - donations can be made to the Group General Fund.

Resources

FamilyTreeDNA Learning Center https://www.familytreedna.com/learn/dna- basics/ydna/
FamilyTreeDNA Search Projects https://www.familytreedna.com/projects.aspx
FamilyTreeDNA Public Y Haplotree https://www.familytreedna.com/public/y-dna- haplotree
National Institute of Standards and Technology (U.S. Dept. of Commerce) Summary List of Y Chromosome STR Loci and Available Fact Sheets https://strbase.nist.gov/ystr_fact.htm
National Institute of Health Genetics Home Reference
https://ghr.nlm.nih.gov/chromosome/Y
Alex Williamson's Big Tree (P-312 and subclades) https://www.ytree.net/ NCBI The Human Pseudoautosomal Region (PAR): Origin, Function and Future
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2435358/
Illumina Dye Sequencing https://wikivisually.com/wiki/Illumina_dye_sequencing ISOGG Wiki
https://isogg.org/wiki/Wiki_Welcome_Page