Big Y

FamilyTreeDNA's sequenced Y-DNA test, called the Big Y, revolutionized Y-DNA genetic genealogy. Other Next Generation Sequencing (NGS) tests yield excellent information, but DNA tests for genealogy are most valuable when compared against other testers, so FamilyTreeDNA's large global database of existing Y-DNA testers gave Big Y the competitive edge from the start. As the Big Y database has grown, the information gleaned has not only grown the Y-DNA haplotree but has created a paradigm shift in the way Y-DNA testing is used.

Conventional wisdom held that the Y-37 STR test held plenty of information to connect to direct paternal ancestors. Over the past five years, the Y-37 became insufficient in many cases, so upgrades to 67 and 111 markers increased. But as the flow of Y-SNP data from sequenced tests increased, the notion that Y-STRs are best for close matching and Y-SNPs are best for deep ancestry proved not to be true in many cases. The overhaul of the Big Y process in 2019, when the test went from Big Y-500 to Big Y-700, added granularity to the results and increased the probability of finding a more recent estimated time to most recent common ancestor (TMRCA).

Since Y-chromosomal DNA passes only from father to son, and since the part of the Y that recombines (mixes) with the X chromosome is not used, Y-DNA genealogical tests can be invaluable in tracing the direct line of paternal inheritance beyond the five to six generations autosomal DNA covers with no recombination to obscure the connection. The Big Y expands and refines the reach of Y-chromosomal testing.

Useful Terms

- **Sequenced DNA test:** Explores mutations on a segment of DNA as opposed to looking for specific mutations at specific locations.
- **Base pair:** Match-up of nucleotides adenine (A), cytosine (C), guanine (G), and tyrosine (T) that comprise DNA along the double-helix strand, such as A-T, C-G.
- **Mutation:** Change from one nucleotide base to another.

- STRs (Short Tandem Repeats):
  - Counts the number of times nucleotides repeat at specific locations on a chromosome. In genetic genealogy, usually refers to Y-DNA.
  - In Y-DNA, traditionally used for "close" matching of testers sharing a paternal ancestor.
  - **Genetic Distance:** Number of STR differences between two Y-DNA testers used to measure 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-DNA haplotree.
  - Corresponds to migratory paths in Y-DNA and mtDNA.
  - Analyzed in all four types of DNA (mtDNA, Y-DNA, X-DNA, and autosomal DNA).
- **Synonymous SNPs:** In Y-DNA, same mutation with two or more different names.
- **Equivalent SNPs:** In Y-DNA, different SNP name, different mutation, but on the same Y-DNA haplotree branch. Not enough information to place on a different branch yet.
- **Reoccurring SNPs:** Same Y-SNP mutation, same SNP name, different places on the Y-DNA haplotree.
- **Ancestral:** The original base state at a location on the chromosome. No mutation has occurred, and the tester is negative for the SNP.
- **Derived:** A change from one base to another at a location on the chromosome. A mutation has occurred, and the tester is positive for the SNP.

What is the Big Y?

An exploratory test that sequences between 16-22 million base pairs of the non-recombining portion of the Y chromosome.

- Includes SNPs and STRs.
- Provides data to build the Y-DNA haplotree.
- Maps the migration path of the tester's direct paternal line.
- Confirms and refines haplogroup.
- Reveals SNPs unique to the tester's family branch.
- Helps to estimate TMRCA to matches.
- Contributes to research and the advancement of science.

Big Y Components

- **Named Variants:** SNPs that are already known and have been given names.
  - Display shows:
    - Whether SNP is derived, meaning positive.
    - Whether the SNP is on the Y-DNA haplotree or not.
    - The Reference (ancestral allele) found for that SNP.
    - The Genotype (tester's allele) found for that SNP.

- **Private Variants:** SNPs that are newer mutations and have not been named yet.
  - Personal, family, or clan mutations, which can be used similarly to STRs.
  - When found in high confidence in at least 2 or more individuals, SNP is named and placed on the Y-DNA haplotree.
  - Display shows:
    - Position of the private variant.
    - The Reference (ancestral allele) found for that SNP.
    - The Genotype (tester's allele) found for that SNP.

- **Y-Chromosome Browser Tool:** Shows each SNP or Position and reads for both the Named and Unnamed Variants results.
  - Browser opens in a pop-up window when Position or SNP is clicked, displays number of positive or negative calls for that specific SNP or Position.
  - Pink indicates that a mutation (derived value) was found for the SNP or Position versus the ancestral value found in the reference model.
  - Blue rows indicate that the read was on a forward strand of the helix and green rows indicate it was on a reverse strand of the helix. The more intense the shade, the higher the confidence of the read.

- **Big Y Matches:**
  - Provided only for SNPs at this time. STR matching stops at 111 markers.
  - Available matches, if any, displayed at the tester's terminal SNP and the four SNP/branches above it.

○ Must share 30 or fewer SNP differences (Named and Unnamed Variants combined).

○ Matches' names and contact information displayed and searchable.

○ Non-Matching Variants displayed; currently includes variants in noisy regions.

○ Number of Shared Variants with each match.

■ Click the number to view all of them. Total based on the number of variants used for matching on the match date.

■ Number of variants used for matching changes over time, so the number of Shared Variants the tester has with a match will also change over time, but the number in this column does not change.
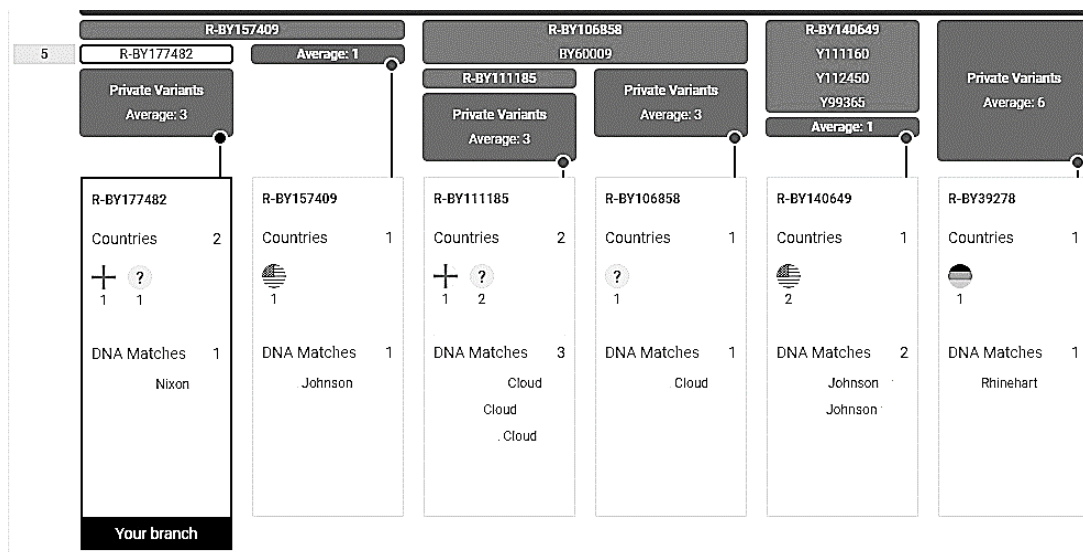
● **Block Tree:** Layout based on the Big Tree by Alex Williamson, volunteer administrator of the R-P312 Project.

○ Provides alternate view of data and matches, incorporating information from the Public Y-DNA Haplotree.

○ Tester's branch shown in white.

○ Child branches shown below parent branches.

○ Sibling branches represented beside each other, below their parent branch.

○ Blocks labeled "Private Variant" represent mutations that are either:

- - Not shared between any branch members.
    - Not yet been validated and placed on the Y-DNA haplotree.
  - Certain "noisy" regions not counted here but are still reported on the Unnamed Variants list; may be relevant upon further manual analysis of specific kit.
  - Aggregated blocks, representing several branches where display is collapsed, shown with average number of available variants below the collapsed branch.
  - Country flags show self-reported Paternal Country of Origin of opted-in testers.
  - Tester's Y-37, 67, and 111 STR matches who also tested Big Y show on the Block Tree and SNP cards whether they qualify as a Big Y match or not.

Group Projects

Surname projects, Y-DNA 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-DNA lines. Group 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 may provide:
    - Data comparison for grouping in project.
    - Data analysis between matches and within project.
    - Data integration with paper trail research.

- **Administrators and project members often help with:**
  - Recruiting test candidates to further research.
  - Financing tests - Donations can be made to a Group Project's General Fund.

Resources

Davis, C., Sager, M., Runfeldt, G., Greenspan, E., Bormans, A., Greenspan, B., & Bormans, C., Big Y-700 [White paper] 2019: https://blog.familytreedna.com/big-y-700-white-paper/
Biology Dictionary: https://biologydictionary.net/dna-sequencing/
Estes, Roberta. DNAeXplained – Genetic Genealogy blog http://dna-explained.com FamilyTreeDNA Learning Center: https://learn.familytreedna.com/
FamilyTreeDNA Public Y-DNA Haplotree: https://www.familytreedna.com/public/y-dna-haplotree McDonald, Dr. Iain. Recent human genetic anthropology: http://www.jb.man.ac.uk/~mcdonald/genetics.html
National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/
Vance, David – DNA Concepts for Genealogy videos:
 -Y-DNA part 1 https://www.youtube.com/watch?v=RqSN1A44lYU&t=16s
 - Y-DNA part 2 https://www.youtube.com/watch?v=mhBYXD7XufI&t=355s
 - Y-DNA part 3 https://www.youtube.com/watch?v=03hRXVg9i1k&t=4s
Vance, David The Genealogist's Guide to Y-DNA Testing for Genetic Genealogy: Self-published, 2020
Y-DNA Warehouse: https://ydna-warehouse.org/ and https://ydna-warehouse.org/statistics.html