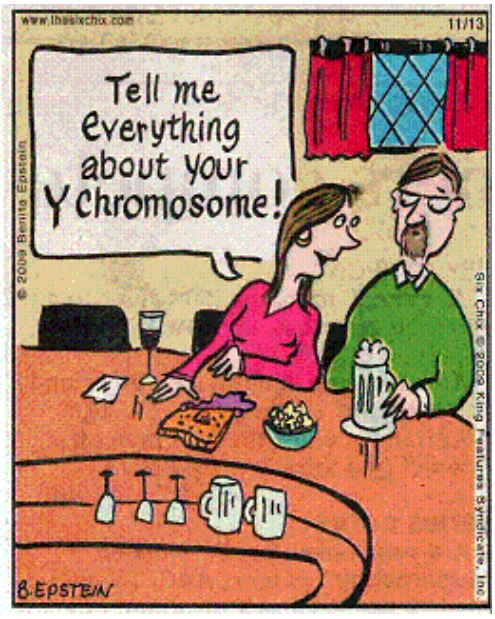
Advanced SNP Tests

Lesson 2. Advanced SNP Tests - Big Y and others

Objective: Understand the Big Y test results and what they mean, plus other SNP tests

Tools: This lesson will walk you through the result of a BigY test and how to interpret the results.

- Words and phrases in **bold** indicate important terminology. Please consult the glossary included with the course material.
- Click any entry in the Table of Contents to jump to that part of the lesson.
- Class notes are at the end of the lesson.



© 2009 - King Features Syndicate, Inc

Lesson written by Gale French

Version 1.0 (Apr. 2021)

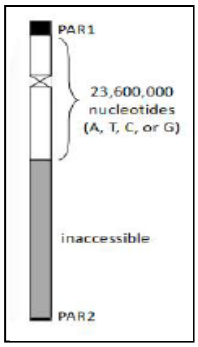
Layout and Design Gale French and Karin Corbeil

Contents

Section 1. BigY

This section covers the BigY test at FTDNA.

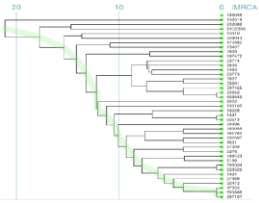
1.1 Overview



Big Y tests over 11M of known 23.6M branch markers (**SNP**s). It is intended for expert users with an interest in advancing science. It tests known markers as well as **loci** (places) where there may be new branch markers. Big Y is able to detect new branch markers that are unique to your paternal lineage, your surname, or even you. Just like **STR**s, SNPs are passed intact from father to son with high fidelity.

BigY is actually <u>two</u> Y-DNA tests in one!

1. STRs - over **700** Markers (ideal for determining closer relationships)
2. SNPs - SNPs allow scientists and genealogists to learn about human migration and the relationships of their **<u>own</u>** distant ancestors. SNPs also mark 'Branching Points' within a family during the most recent 10-20 Generations.



The BigY test has evolved over the years. The first version in 2013 used the Human Genome (**HG**) version 19 and requirements for previous STR testing and gave 111 STRs. In 2015 FTDNA released the BigY-500 that added 389 more STRs, introduced the Step Chart and used the HG38 version. BigY-700 was released in 2019 and reported **838** STR markers, replaced the Step Tree with the Block Tree and removed any previous test requirements.

The results of all 3 tests can be compared to each other. If you took a BigY-500 in the past, there is an upgrade to Big-Y700.

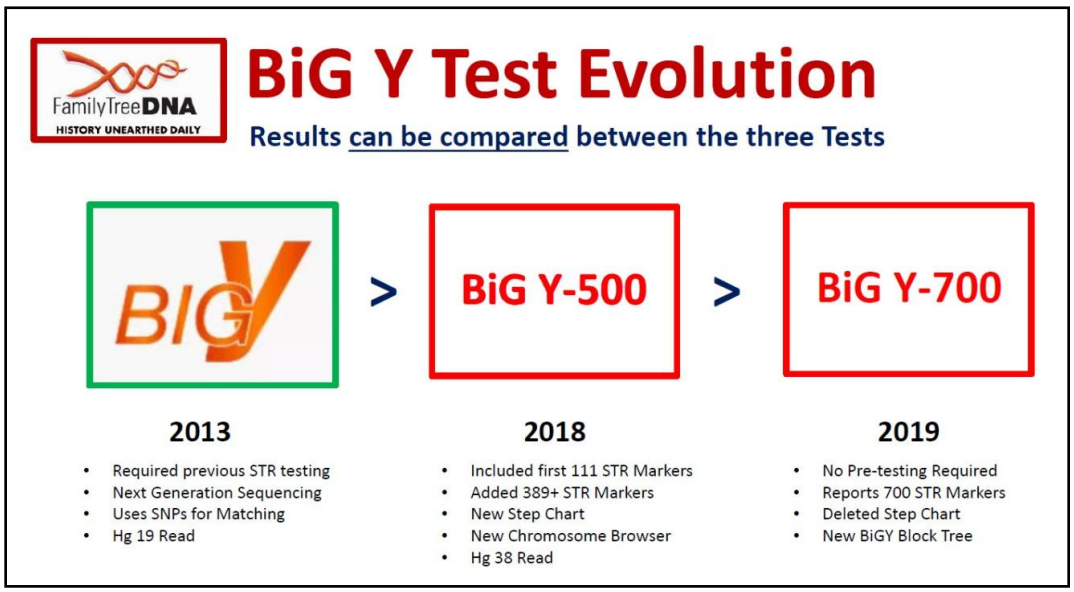


Figure 1 - History of Big Y
Chart courtesy of Bill Wood - BigY Facebook Group

Pop Quiz #1

How many **STR** marker values result from the Big Y-700 test?

1.2 BigY Buttons

When you get your results for the BigY, your FTDNA dashboard will show 4 buttons with information about your BigY data (see Figure 2). The buttons are:

1. Block Tree - current terminal Haplogroup data
2. Matches - matches with named & unnamed variants
3. Results - named & unnamed variants with Chromosome Browsing Tool
4. Y-STR Results - over 700 STRs

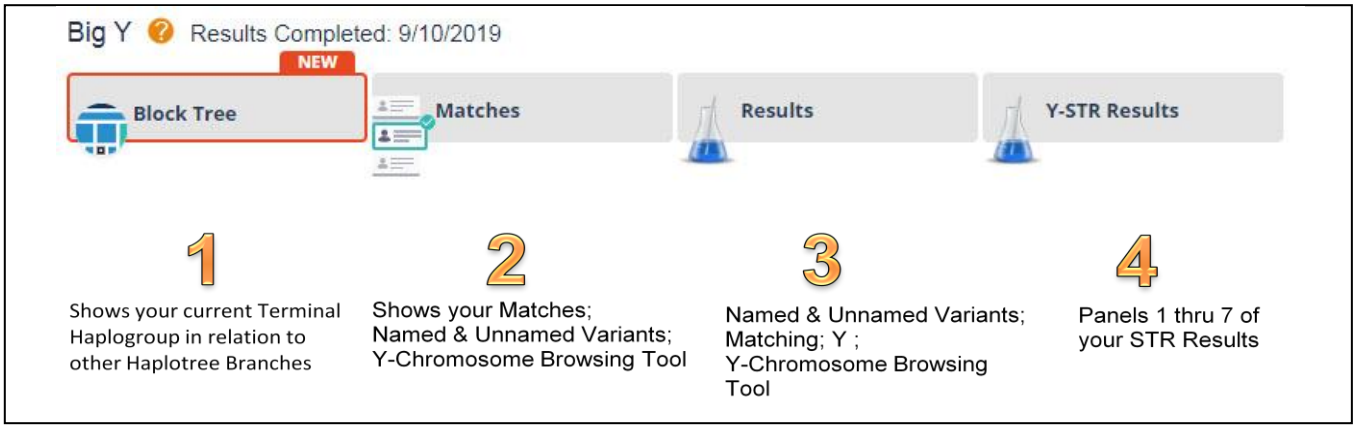
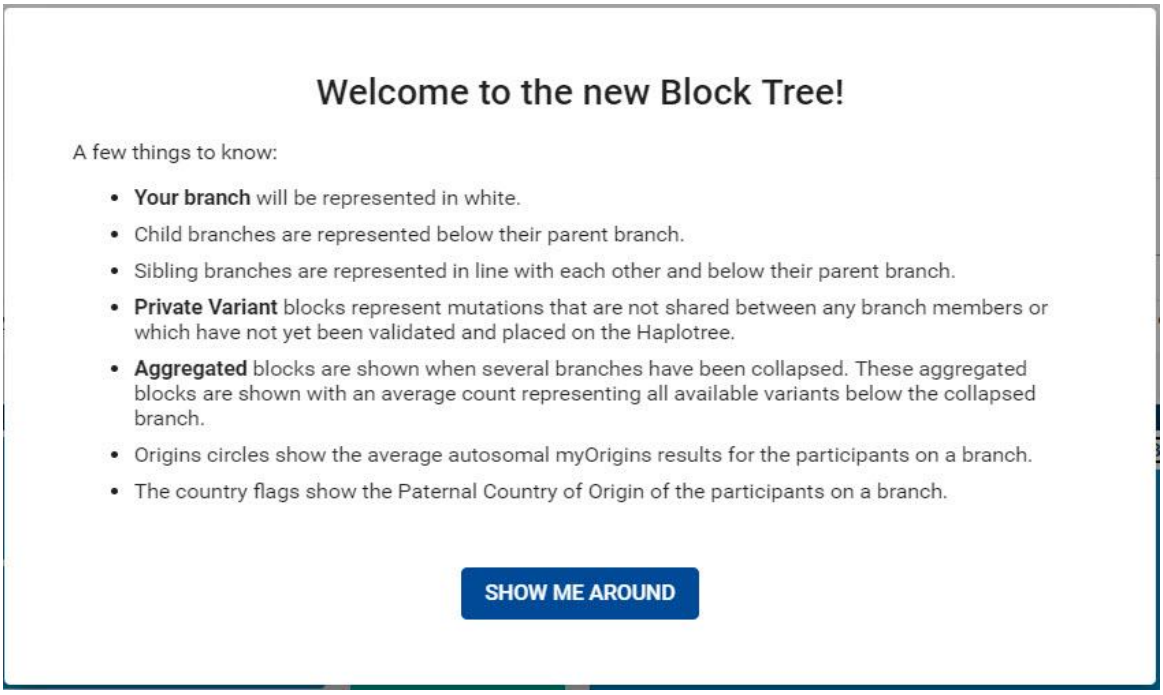


Figure 2 - Big Y Results buttons

1.2.1 Block Tree

Once you click on the 'Block Tree' button, you get an info banner that welcomes you and gives information about the Block Tree. Read and understand the notes in the banner. The ' Show Me Around ' button gives you a tutorial on the Block Tree.



When you get to the Block tree, there are several sections that you need to learn (by color/icon code):

1. **Equivalents** – require more tests to create a unique branch
2. Subclades of Block Equivalent - child branches are shown under their parent, sibling branches are in horizontal line with each other below their parent
3. **Private Variants** – Mutations not shared between any branch members; useful to find **TMRCA**. Aggregated Blocks are shown where several branches have been 'collapsed'. These blocks show an average count of the variants below the collapsed branch
4. Your Branch – white with a black border
5. Origins – Family Finder results; shows the average autosomal 'myOrigins' for the participants on a branch
6. Countries – country origin associated with **EKA/MDKA**
7. DNA Matches – BigY matches
8. 5 Number of SNPs (scale on far left) between this/your Kit and the SNP identified to the right on the chart

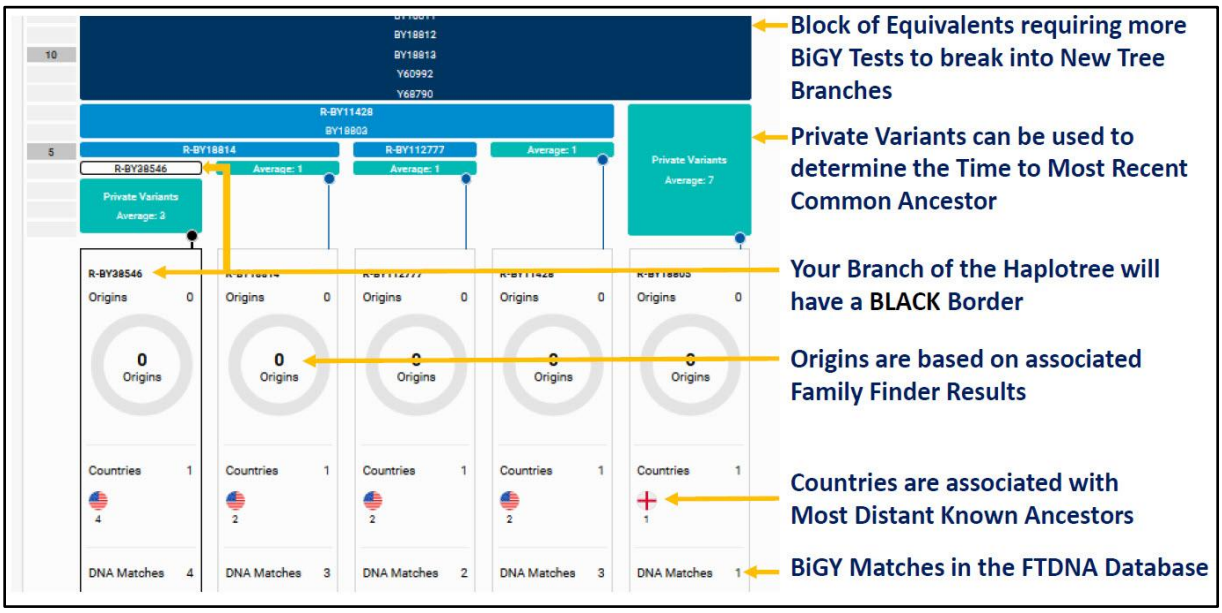
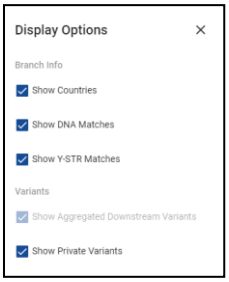


Figure 3 - Big Y Block Tree

Chart courtesy of Bill Wood - BigY Facebook Group

Your results may look <u>quite</u> different based on the Haplogroup, the number of other testers, matches, etc. You can change the data displayed by clicking on "Display Options" at the top of the page to turn elements on/off.



Pop Quiz #2

When looking at the Block Tree for your kit, which 'block' is yours?

The SNP Blocks are shown from left to right as Youngest (L) SNP to Oldest (R) SNP. The Private Variants may be used to determine TMRCA. TMRCA is calculated with BigY in two different ways:

1. Average #SNPs to the bottom box **x** 144
2. Private Variants **x** 144

For example, in the chart below, **R-PH367** (using the 'bottom box' method) is 1 x 144 which formed about 144 years ago. It's the same with the Private Variant method (144). **R-BY23335** is 3 x 144 which formed about 432 years ago. **R-A5919** is 7 x 144 which formed about 1008 years ago (by either method - see yellow arrows ← and braces }7 for sum of Private Variants). YFull has a tool (see Lesson #3) that calculates SNP age (1100-600 YBP) which you see correlates with the BigY age computed using these variables.

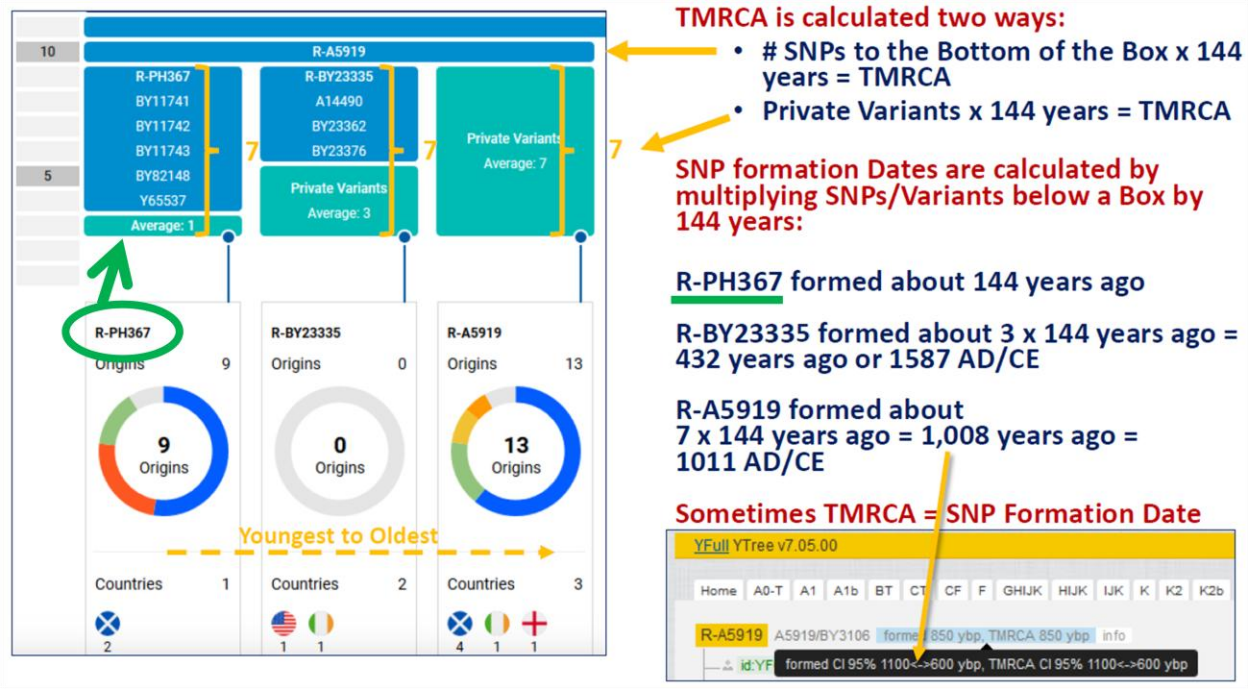
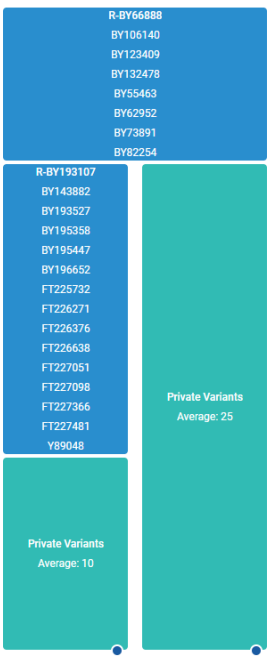


Figure 4 - TMRCA from Private Variants
Chart courtesy of Bill Wood - BigY Facebook Group



Pop Quiz #3

In the Block Tree snippet to the left, what is the estimated TMRCA for **R-BY66888**?

1.2.2 Matches

Button #2 displays your **terminal SNP** matches. A person is considered a match if they have 30 or fewer differences in SNPs with you. The columns on the page are:

1. ***Match Name*** - This column displays the names of your matches. Clicking on the name will show you match's Profile (e-Mail, Haplogroup(s), EKA, Surnames). The icons allow you to e-Mail your match or add notes about your match. Once a note is added, the callout icon turns green .
2. ***Non-Matching Variants*** - This column displays the known variants (SNPs) within your subclade that you and the specified match *do not* share. When you and a match do not have any non-matching variants, then no SNP names will be listed for that match.
3. ***Shared Variants*** - This column displays the number known variants (SNPs) within your subclade that you and the specified match *do* share.
4. ***Match Date*** - This column displays the date you and the specific match were matched.
5. - The sort arrows located at the upper-right corner of each column to sort the data in ascending or descending order.
6. The Name, Non-Matching and Date fields have search boxes to filter/narrow your results. You may enter full or partial entries (e.g. 'Smit' for Smith or Smithers)

The following chart (Figure 5) is my Matches result with my father; it shows that I have one mutation at Y Chromosome location 151707. The box on the right gives information about the match; total number of variants considered is **791,185**.

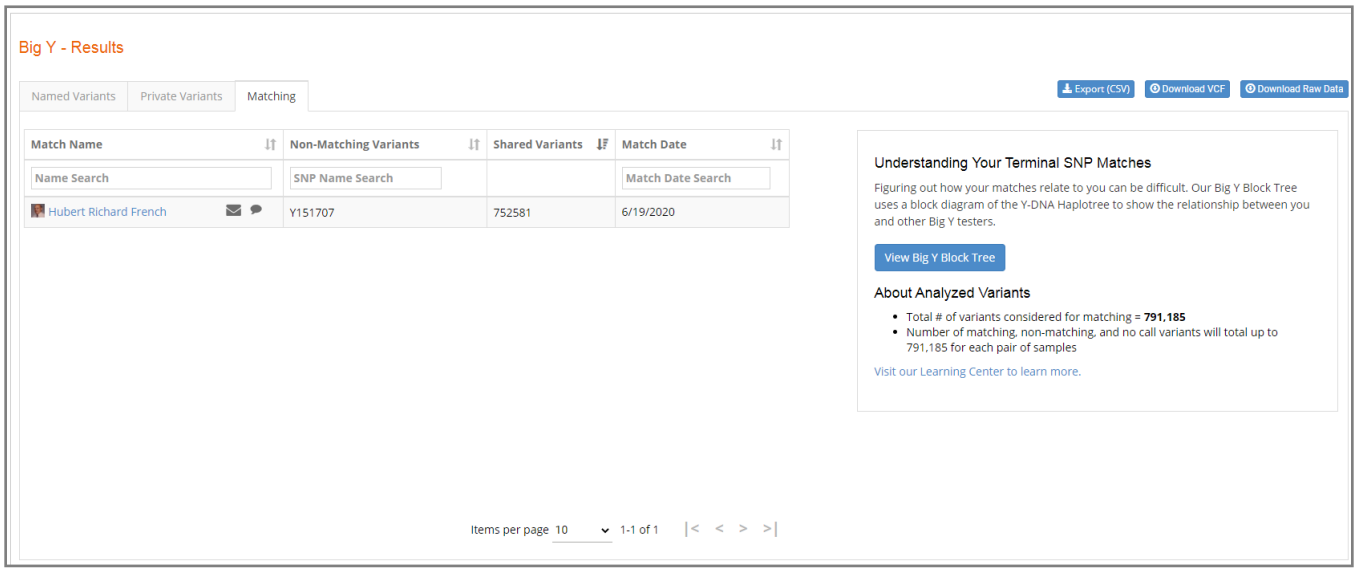


Figure 5 - Matches page

Caution!!

To be able to use the mail icon, you must have installed an e-Mail client (e.g., Outlook).

1.2.3 Results

This page shows each **SNP Name**, whether it was **Derived** (+ - ?)[1], **On the Y-Tree** (whether or not the SNP exists on the FTDNA Y-DNA Haplotree) and the allele **Reference** (HG38) and the **Genotype** (<u>your</u> base at the given position). If you click on one of the SNP names, it takes you to that spot in the Y-Chromosome with the Browser. The Y-Chromosome browsing tool allows you to view forward and reverse strands for any SNPs in your profile.

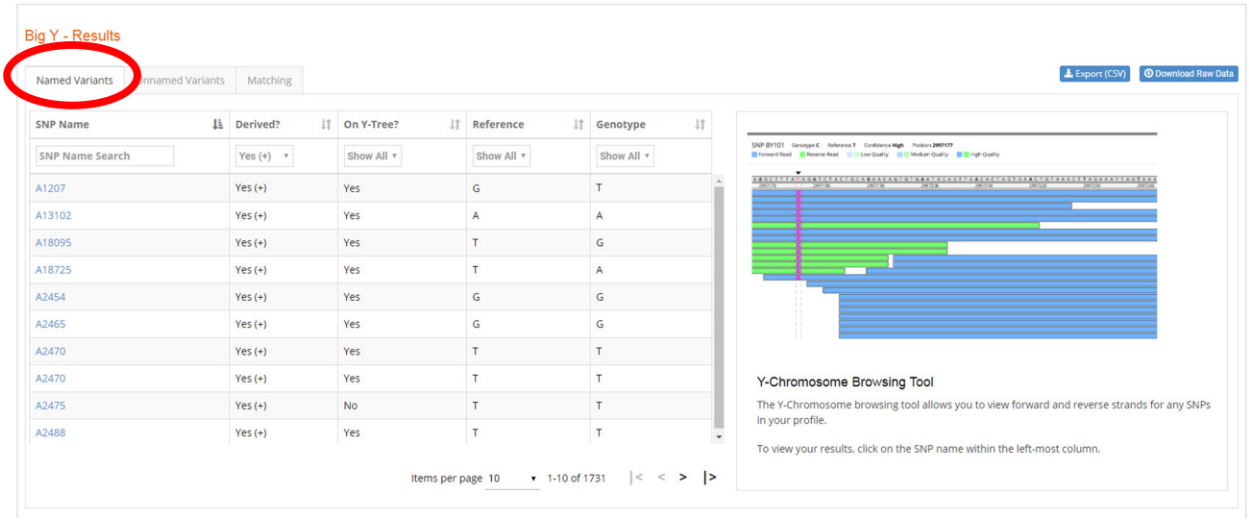
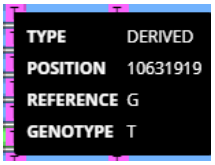


Figure 6- BigY Results page

The Y-Chromosome Browsing Tool opens in a new window, displaying the location of the SNP. The SNP you selected (A1207) is displayed in pink with a downward arrow pointing to the position of the SNP (see **red** box). The other pink locations display your other nearby SNP positions. You can click on any of your SNP positions to see more information for that SNP.



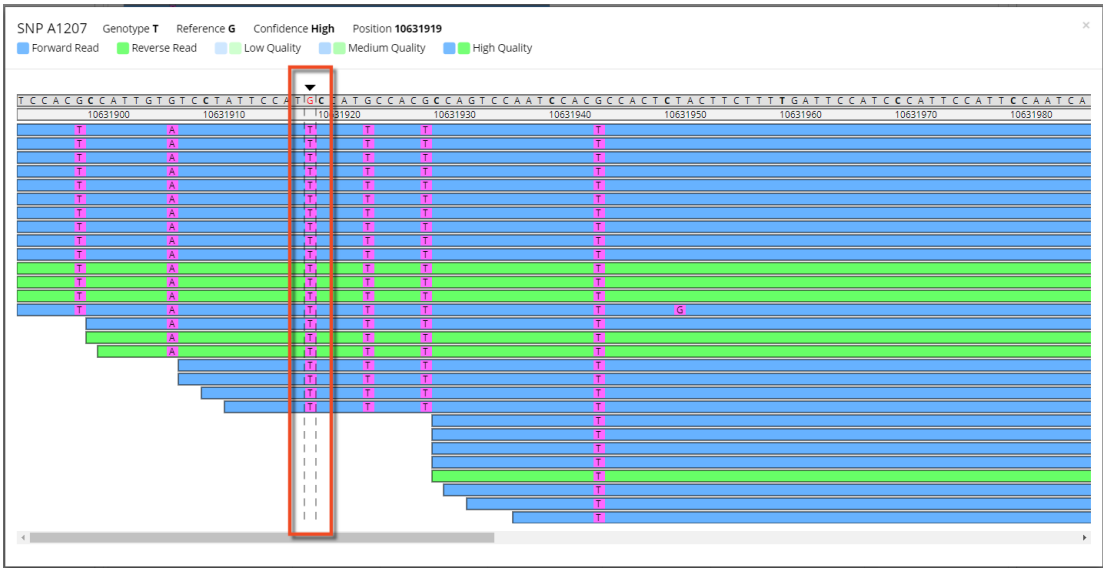
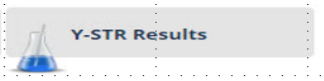


Figure 7 - BigY Results page

[1] ***?*** – Indicates that FTDNA did not get good coverage on the SNP, therefore, were unable to confirm whether or not the called genotype is derived.

1.2.4 Y-STR Results



This final page lists the allele results for all your STR markers. Panel 7 indicates that you have at least 700 markers from the BigY 700 test (actually 838!).

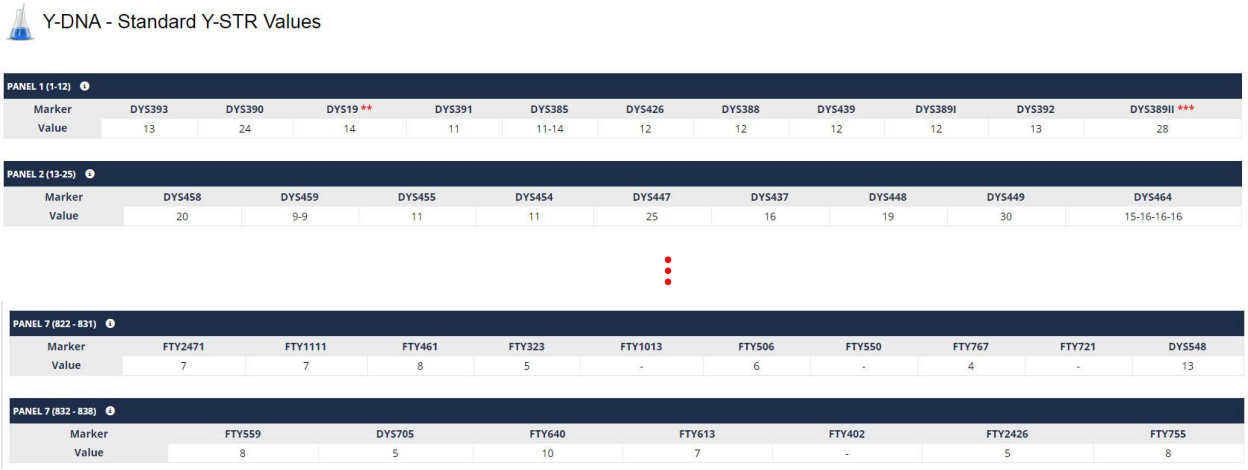


Figure 8 - Y-STR Results Values page

1.3 Post Processing

What happens to your BigY test during and after the processing? According to Michael Sager at FTDNA:

- After sequencing, samples undergo variant calling and matching
- New SNPs are identified at this stage and named
- Within 24 hours of result posting, all new SNP names are published on the ISOGG wiki https://isogg.org/wiki/FT_SNP_index
- Quick naming was done to cut down on double/triple naming
- Samples are then placed onto the FTDNA Haplotree
- Matching variants and equivalent breakers are identified and subsequently reviewed for inclusion into the tree

Project Administrators

FT-DNA Project Administrators are unpaid volunteers that donate their time and experience to help users understand their results and where they "fit" in the paternal DNA picture puzzle.

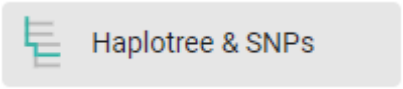
Section 2. Other SNP Tests

Why would I ever want to do a SNP test other than BigY? Let's say that you know your terminal SNP, and you have a candidate that matches you (on STR test) with a low GD. To determine if you are a closer match you could do a "targeted" SNP test to see if the two of you are close without the expense of a full BigY test.

2.1 FTDNA

In addition to the Big Y, FTDNA offers other "lesser" SNP tests.

1.1.1 SNP Packs



If you look at the Haplotree (click on the '**Haplotree & SNPs**' button on the dashboard), you will see "little **red** baskets" next to a SNP name. This means that this SNP is part of a SNP Pack which bundles many known SNPs into a test that will show if you are positive (or -) for that particular SNP. Click on the 'Add' button to see the details. In this example (Figure 9 - Haplotree & SNPs page), **R-DF84** is part of the '**R1b - DF27**' SNP Pack that tests 153 SNPs for $119.[00].

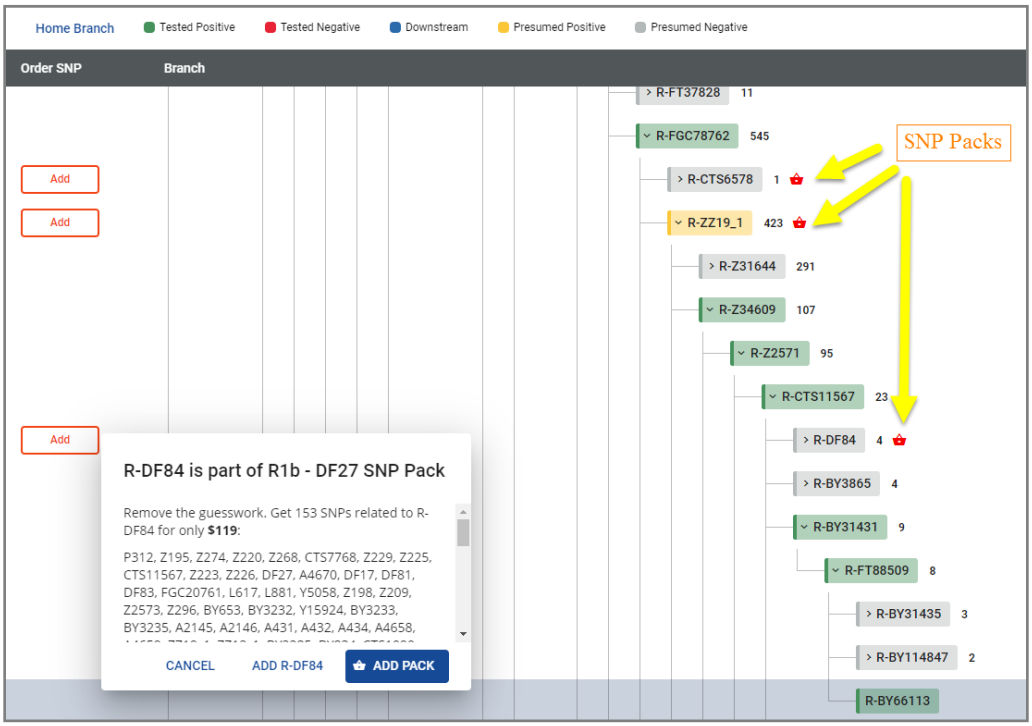


Figure 9 - Haplotree & SNPs page

2.1.1 Single SNPs

If you see the 'Add' button on a SNP without the **red** basket, that means that this is an individual SNP (not part of a SNP Pack) that is available for $39.[00]. You can see how economical a SNP Pack is!

Pop Quiz #4

I have a match with Genetic Distance = 3. How can I tell which markers are different (and if they are "fast changing" markers)?

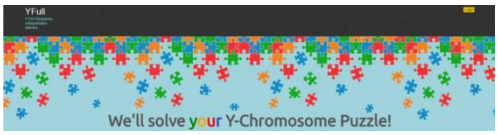
2.2 YSEQ

YSEQ provide STR and SNP testing. They also have a set of utilities to analyze your Y-DNA but you must purchase the analysis kit. You can chose either single SNPs or "Panels" for your Haplofroup.

They also do Whole Genome testing. YSEQ is owned and operated by Thomas and Astrid Krahn (former employees of FT-DNA). https://www.yseq.net/

2.3 YFull

YFull's NextGen Sequence Interpretation reports about your BAM file (from FT-DNA BigY test) are provided online ($49). They also provide other downloads and reports. https://yfull.com



2.4 Full Genomes

The Full Genomes Corporation (https://www.fullgenomes.com) have announced the launch of a new Y-chromosome next generation sequencing product known as the **Y Elite 2.1**. The technical details are as follows:

- SNP Length coverage: 13.2+ megabases (a conservative estimate)
- Read length: 250 base pairs
- Coverage: 50x

They say that it's "*better SNP calling and better STR calling quality*". Cost is $645.^{oo} (U.S.). They also provide whole genome sequencing ($1295.^{oo}).

Pop Quiz #5

If I have a highlighted allele value on one or more of my markers, can my Genetic Distance be zero?

Section 4. Glossary [2]

- **Allele** - One of the possible values for a marker. For example, with Y chromosome DNA, the values of "13" or "14" would be different alleles for DYS393.
- **Base** - A base is a unit or building block of DNA. Adenine (A), cytosine (C), guanine (G), and thymine (T) are the four primary bases in DNA. The order of bases is the sequence of DNA.
- **EKA/MDKA** - **E**arliest **K**nown **A**ncestor -or- **M**ost **D**istant **K**nown **A**ncestor
- **GEDCOM -** An abbreviation for "Genealogical Data Communications". A standard file format for genealogy records.
- **Genetic Distance** - Genetic Distance is the number of differences, or mutations, between two sets of results. A genetic distance of zero means there are no differences in the results being compared against one another. See following pages for 11 markers.

# of Markers	Genetic (Allelic) Distance											
		0	1	2	3	4	5	6	7	8	9	10
	12	Related	Possibly Related	Probably Not Related	Not Related							
	25	Related	Related	Probably Related	Probably Not Related	Not Related						
	37	Very Tightly Related	Tightly Related	Related			Probably Related	Possibly Related	Not Related			
	67	Very Tightly Related	Tightly Related		Related			Related		Probably Related	Possibly Related	Not Related

- **Haplogroup** - A Haplogroup is a major branch on either the maternal or paternal tree of humankind. Haplogroups are associated with early human migrations. Today these can associated with a geographic region or regions.
- **Haplotree -** A Haplotree is the phylogenetic tree that defines your Haplogroup and all its sub-clades
- **Haplotype** - A haplotype is the set of DNA values. For example, the results of the Y-DNA12 test for one person are their haplotype.
- **Locus -** A locus is a specific location in your genetic code. In a genetic map of our DNA, the locus tells us where to find any base. The plural of locus is loci.
- **Marker** - A marker is a physical location (locus) on the chromosome. The term is often used colloquially in genetic genealogy to refer to a short tandem repeat (STR).
- **Most Distant Ancestor** - the person along your paternal line that your paper research shows as the "oldest" male ancestor you have found
- **Most Recent Common Ancestor (MRCA)** - the most recent ancestor from which the two individuals are descended. Also **E**arliest **K**nown **A**ncestor.
- **Named Variant -** a named SNP identified by Haplogroup (A-T) and followed by a SNP name (e.g., M269).

[2] **NOTE:** the Glossary "grows" with each lesson (and includes definitions from previous lessons) so that you don't have to go back to other Lessons to get definitions.

- **Non-Matching Variants -** SNPs that are different between kits, usually denoting different branches.
- **Novel SNPs** - newly found mutations that have not been registered in a Haplotree data base. They will get a new SNP name.
- **Non-Paternal Event (NPE)** - a break in the Y chromosome line due to adoption, infidelity or numerous other causes. Also called **M**is-**A**ttributed **P**arentage.
- **Phyloequivalents** - similar SNP(s) discovered independently, but it is unclear of the order in which they occurred. More testing may show the true order of mutations.
- **Private Variant/SNP -** SNP(s) unique to one person's kit or an immediate family (father, brother, son).
- **Shared Variants -** SNPs that exist in two or more kits.
- **Short Tandem Repeat (STR**) - A short DNA motif (pattern) repeated in tandem. ATGC repeated eleven times would give the marker a value or allele of 11.
- **Singleton** - one person with a unique SNP mutation shared with nobody else.
- **SNP -** a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population.
- **Subclade** - term used to describe a subgroup of a subgenus of a Haplogroup.
- **Terminal SNP -** the last **named** SNP tested for a kit. This may or may not be the last SNP in the ancestral line; only the last tested for this person.
- **Unnamed Variant -** an unnamed SNP, identified by coordinates along the Y chromosome.
- **Variant -** a mutation on the Y chromosome.
- **Y-Chromosome** - One of the two sex chromosomes, X and Y. The Y-Chromosome passes down from father to son. Females do not receive it. As the Y-Chromosome is passed on through the paternal line, it is valuable for surname based genealogy studies.
- **Y-DNA Backbone test** - If a person's Y-DNA Haplogroup cannot be predicted with 100% confidence, the SNP Assurance Program at FTDNA will test your sample with our Backbone SNP test for FREE. This test is a "deep" multiple SNP test. Specifically, if we cannot predict a person's Y-DNA Haplogroup with sufficient confidence that they can join the National Geographic's Genographic Project, we will automatically perform a Backbone SNP test in order to identify the Haplogroup assignment. Backbone tests take about 6-8 weeks from the time they are ordered.

- https://dnaadoption.org/acronyms/

  https://dnaadoption.org/glossary/

Section 5. Resources

Once you have

5.1 Internet

- The resources we've used in class, of course: FamilyTreeDNA.com.
  - https://www.familytreedna.com/learn/using-the-kit/family-tree-dna-test-kit/
  - https://www.familytreedna.com/public/y-dna-haplotree/A
  - https://learn.familytreedna.com/y-dna-testing/big-y/big-y/
  - https://ghr.nlm.nih.gov/primer/genomicresearch/snp
- ISOGG: http://www.isogg.org

  https://isogg.org/wiki/Y-SNP_testing
- BLOGS: http://dna-explained.com/2014/09/04/what-does-and-doesnt-a-y-dna-match-mean/

  https://dna-explained.com/2019/01/24/family-tree-dnas-new-big-y-block-tree/

  https://dna-explained.com/category/big-y-700/

  https://blog.familytreedna.com/2019-review-of-big-y/

  https://blog.familytreedna.com/human-y-chromosome-testing-milestones/
- There are many surname project websites, often associated with Y-DNA tests.
  https://www.familytreedna.com/my/group-join
- If you haven't found what you want yet, visit Cindy's List http://www.cyndislist.com/ for an incredible list of genealogy resources organized by category.

5.2 Videos

- https://www.youtube.com/watch?v=ekB9LY_aL04 Family Tree DNA Results Explained: Y-DNA Markers, Matching & Genealogy (FT-DNA Webinar 1:55:40)
- https://www.youtube.com/watch?v=zy0QnSLnFtw How to enhance your Y-DNA results through Surname and Haplogroup projects (1:11:42)
- https://www.youtube.com/watch?v=qiv0Ny6nvAY Help, My Y-DNA Matches Have a Different Surname! (FT-DNA 1:29:56)
- https://www.youtube.com/watch?v=Wicb2_bEIYo Adoptions & illegitimacies - using DNA to solve adoption mysteries (Dr. Maurice Gleeson 1:03:40)
- https://www.youtube.com/watch?v=f_IPIPVALnE Michael Sager Feb 2020 "The Tree of Mankind" at Genetic Genealogy Ireland conference

5.3 Books

- Bettinger, Blaine T. - "***Guide to DNA Testing and Genetic Genealogy"*** ©2016, Family Tree Books, Cincinnati, OH [ISBN-13: 9781440343326 9781440345395] *<< available as book or e-Book>>*
- Aulicino, Emily - "***Genetic Genealogy: The Basics and Beyond***", ©2014, AuthorHouse LLC, Bloomington, IN [ISBN13: 9781491840900]
- Smolenyak, Megan & Turner, Ann - "***Trace Your Roots With DNA: Use Your DNA to Complete Your Family Tree***", ©2004, Rodel Books, New York [ISBN 1594860068 (ISBN13: 9781594860065)]
- Pomery, Chris - "***DNA and family history : how genetic testing can advance your genealogical research***", ©2004, Dundurn Group, Toronto, Ontario
- Dowell, David R. (Ph.D.) - " ***NextGen Genealogy: The DNA Connection***", ©2015, Santa Barbara, California : LIBRARIES UNLIMITED, an imprint of ABC-CLIO, LLC
- Hill, Richard - "***Finding family : my search for roots and the secrets in my DNA***", ©2012, Richard Hill (self published), Grand Rapids, Michigan. *<< available as book or e-Book>>*
- Griffeth, Bill- "***The Stranger in My Genes***", ©2016, UPNE Book Partners, Lebanon, NH [ISBN-13: 978-0-88082-344-9]

Section 6. Pop Quiz Answers

1. **838** - Although the test is called Big Y-700, it actually reports allele values for 838 markers (shown in panel 7 of the Y-STR Results page.

2. The one with the black border.

3. **3600** - 25 x 144 = 3600. Using the first method, there are 25 Private Variants below **R-BY66888** in the right block. Using the second method, there are 15 SNPs } plus 10 SNPs [(15+10)*144] below SNP **R-BY66888** to the bottom of the left two blocks.

   In the next lesson you will see the YFull tool that will correlate the age to 600 - 4500 YBP.

4. The only way to see the STR markers of your matches is in the Surname Project (if they have joined). Use the browser 'Search' to find if you know his kit number. There, he has been grouped with "like" markers/Haplogroups and you can see the values for each marker. If you look at the 'colorized' version it will highlight the differences in markers in the group. Then you can see if the differences were on fast changing markers.

   The Y-Chromosome DNA (Y-DNA) results chart headings are color coded in two ways. First, each testing level (Y-DNA1-12, Y-DNA13-25, Y-DNA26-37, Y-DNA38-67, and Y-DNA68-111) is coded with a different shade of blue. Second, the STR (short tandem repeat) markers that have faster mutation rates and are more likely to change within the genealogical time frame are coded with a red background.

Y-DNA1-12	Y-DNA13-25	Y-DNA26-37	Y-DNA38-67	Y-DNA68-111	Faster Changing STR Marker

5. No. Highlights mark differences. With each highlighted difference, that increases your Genetic Distance (by one or more!). See definition of Genetic Distance in the Glossary.

Section 7. Epilogue

Class Notes

Summary

Y-DNA can trace the origins of a surname and can be used to find other people who shared an ancestor on this paternal line. Y-DNA Testing is the process of getting your Y-Chromosome results for comparing with other testers.

Access to Materials

Your Moodle login will continue to be available to you. If you'd like to retake the class, you are welcome to sign up as space is available. No additional donation will be requested (although they are certainly welcome!). Email DNAAdoptionHelp@gmail.com if you wish to re-take this class.

Class Forum

The class forum will be available for follow-up questions for an additional 4 weeks subsequent to the posting of Lesson 3. Participants will also have the option of re-taking this class for free within the next year. Contact DNAAdoptionHelp@gmail.com if you wish to re-take this class.

Stay in Touch! Let us know how your journey is going.

DNAAdoptionHelp@gmail.com