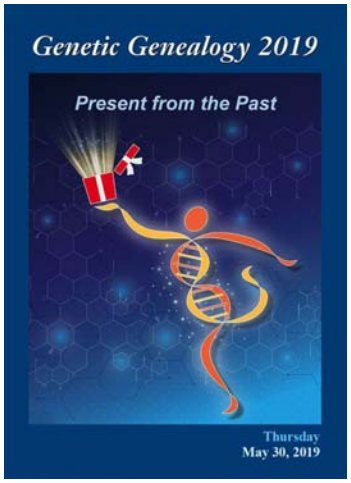
Why get the Big Y DNA Test?

FTDNA's Big Y-700 DNA test will be examined in terms of costs, markers, genealogy, and privacy concerns.



Brad Larkin

Prepared for the
Southern California Genealogical Society
Jamboree 2019

The Answer

- The Y-DNA Phylogenetic Tree
  - The most important reason to get the Big Y DNA test is that the results locate your family on the phylogenetic tree of mankind.
    - Distinctness
      - With millions of markers tested, every family will have unique marker(s)
    - Defined Kinship
      - Kinship between Y lineages on the tree is clear and getting more refined as more samples are tested.
      - Based on absolutes rather than probabilities.
        - » Either the two samples have the same mutation, or they do not.

Outline

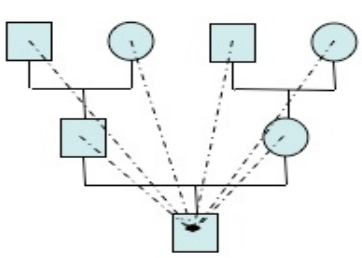
- Intro to Y-DNA Genetic Genealogy
- Phylogenetic Trees
- Comparing Big Y to other Y-DNA tests
- How to Use YOUR Big Y Results

Intro to Y-DNA Genetic Genealogy

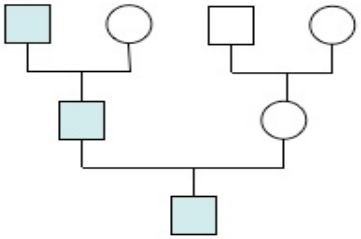
- Y-DNA vs MtDNA and Autosomal
- Y Chromosome Regions
- SNPs vs STRs
  - Co-evolution of SNPs and STRs
- Explanation of SNP naming and chromosome positions.

Chromosome Fit for Genealogy

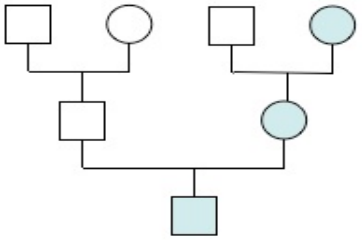
	Autosomal (Microarray)	Y-Chromosome (Y-37 STRs)	Mitochondrial (HVR1+HVR2)
Recombination - Mixing	Yes	No	No
# Coding Genes	~ 30,000	86	37
# Markers Initial Test	708,093	37	1,120
Mutation Rate	0.5 bp/gen = 354,047 per generation	μ = 0.0041 markers/generation 1 change per 165 years	0.48 bp/MY = 1 change per 1,860 years



Autosomal
(passed on in part, from all ancestors)



Y-Chromosome
(passed on complete, but only by sons)



Mitochondrial
(passed on complete, but only by daughters)

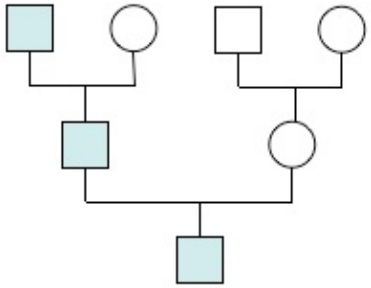
Recombination Inhibits Genealogy

- Recombination -> children might not be same genetically as a particular parent.
  - Makes it hard to track genealogical ancestry.
- Two areas that don't recombine:
  - MtDNA from mother
  - Y-DNA from father

Recap: Biology – Types of DNA Used for Genealogy

Chromosome Type	Strengths	Weaknesses
Y-DNA	Deep ancestry with measurable variation in surname era.	Only patrilineal ancestry. Only Male samples.
MtDNA	Sample any Gender Deep ancestry, strong signal	Only matrilineal ancestry. Slow mutation – no differentiation in surname era.
Autosomal	Sample any Gender Traces all lineages Adoptee research	Fuzzy signal lost after a few generations. Recombination makes interpretation challenging

Y-DNA



Y-Chromosome
(passed on complete, but only by sons)

- Biology
  - Non-recombining portion of the Y-chromosome
    - Holds signal over multiple generations
    - Mutates at a rate that we can see differences occur within families over the course of decades and centuries

Useful Y Segments

- Also called ‘Contigs’ for Contiguous Regions

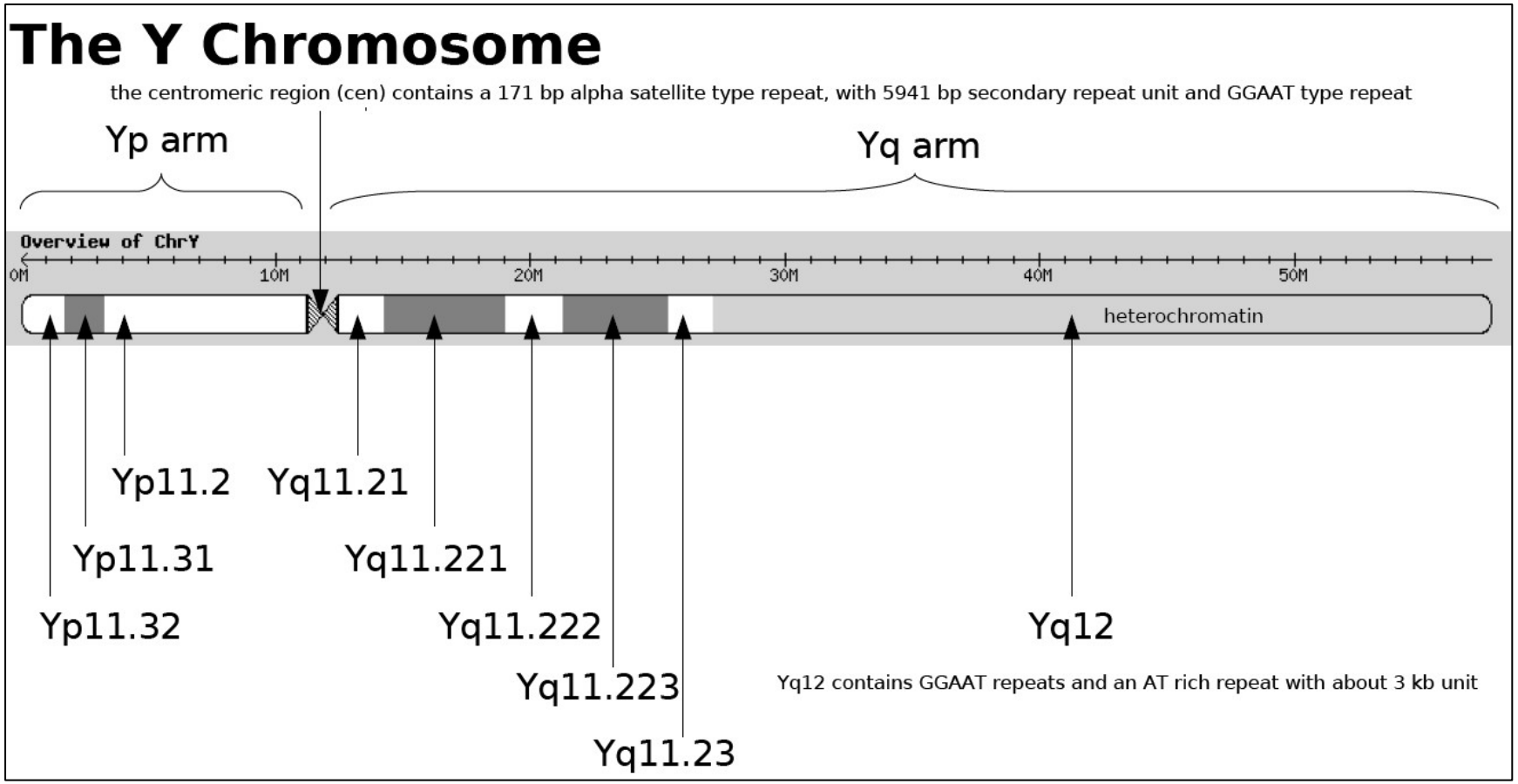


Image adapted from Thomas Krahn, *Walk On Y Project* presentation at Family Tree DNA 2007 Conference.

MSY Complexity

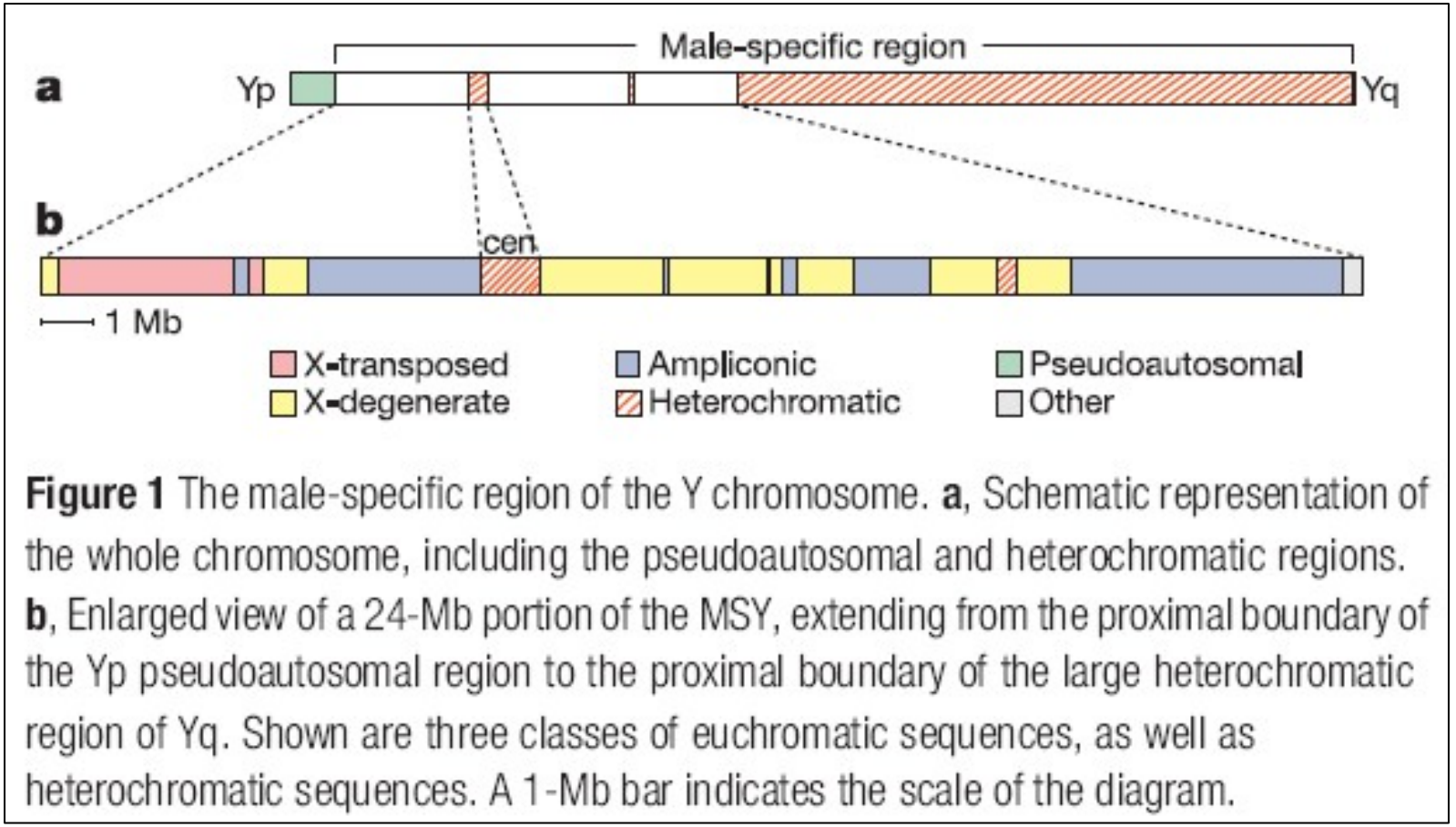


Figure 1 The male-specific region of the Y chromosome. **a**, Schematic representation of the whole chromosome, including the pseudoautosomal and heterochromatic regions. **b**, Enlarged view of a 24-Mb portion of the MSY, extending from the proximal boundary of the Yp pseudoautosomal region to the proximal boundary of the large heterochromatic region of Yq. Shown are three classes of euchromatic sequences, as well as heterochromatic sequences. A 1-Mb bar indicates the scale of the diagram.

Image adapted from Figure 1 from Skaletsky et al (2003), The male-specific region of the Y chromosome is a mosaic of discreet sequence classes, Nature, Vol 423, https://doi.org/10.1038/nature01722

Y Chromosome Structure

- Imagine the curled Y-chromosome stretched out like one spaghetti noodle.
- In DNA sequencing paradigm, scientists assign a '**position**' to each nucleotide chemical base.
  - From 1 to about 58 million base pairs
  - Large portions are repeating, un-useful segments (grayed out in diagram)

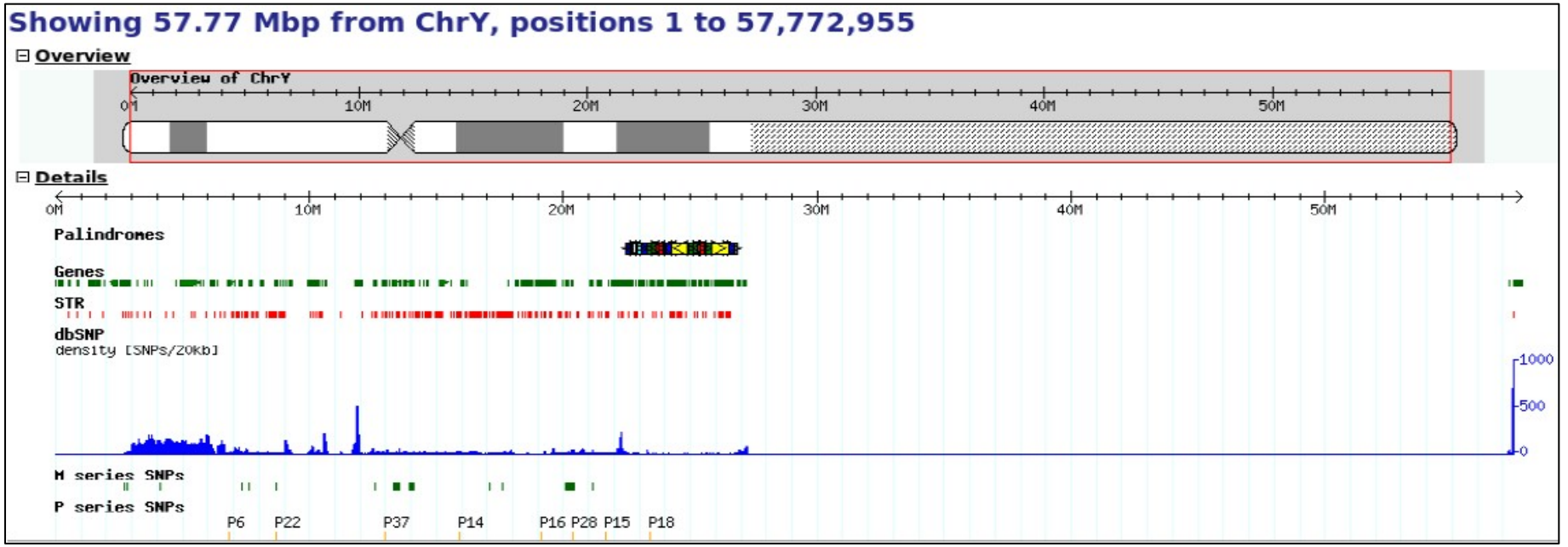


Image adapted from Thomas Krahn, *Walk On Y Project* presentation at FTDNA 2007 Conference.

SNP vs. STR Measurement

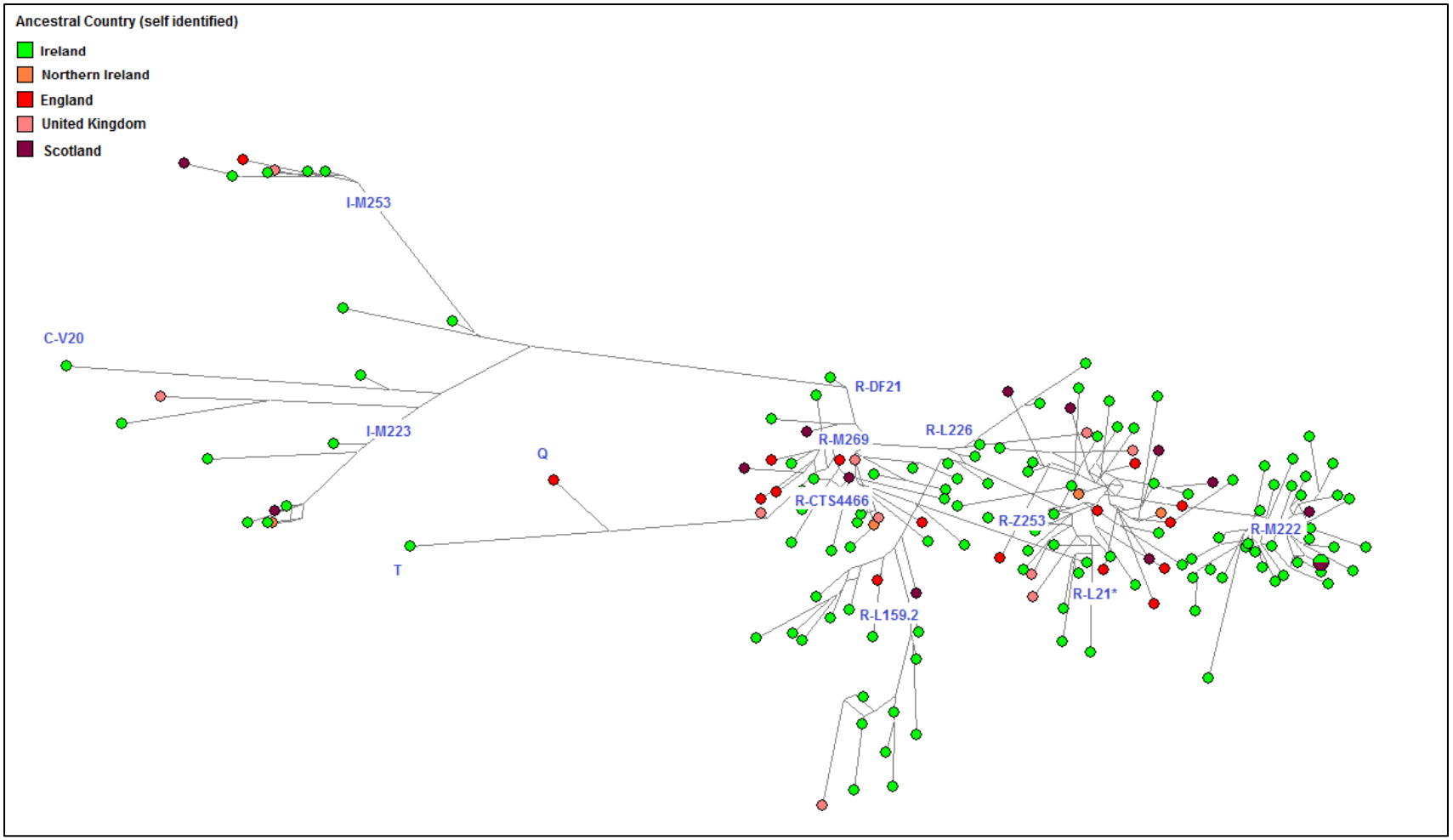
- SNP = Single Nucleotide Polymorphism
- Mutation in a single base pair at a specific position
- Expressed a 'positive' when different from ancestral allele value.
  - e.g. mutation *rs1019875*
  - Person1 T ATCC T = -
  - Person2 T ACCC T = +
- Analogous to '*Trunk and Branches of the Tree*'

- STR = Single Tandem Repeat
- Repeating patterns of multiple base pairs
- Allele Count = number of repetitions of particular pattern
  - e.g. *DYS389*

  Person1 T AACC T = 1

  Person2 T AACC AACC T = 2
- Analogous to '*Leaves on the Tree*'

Y-STR Clusters with Haplogroups



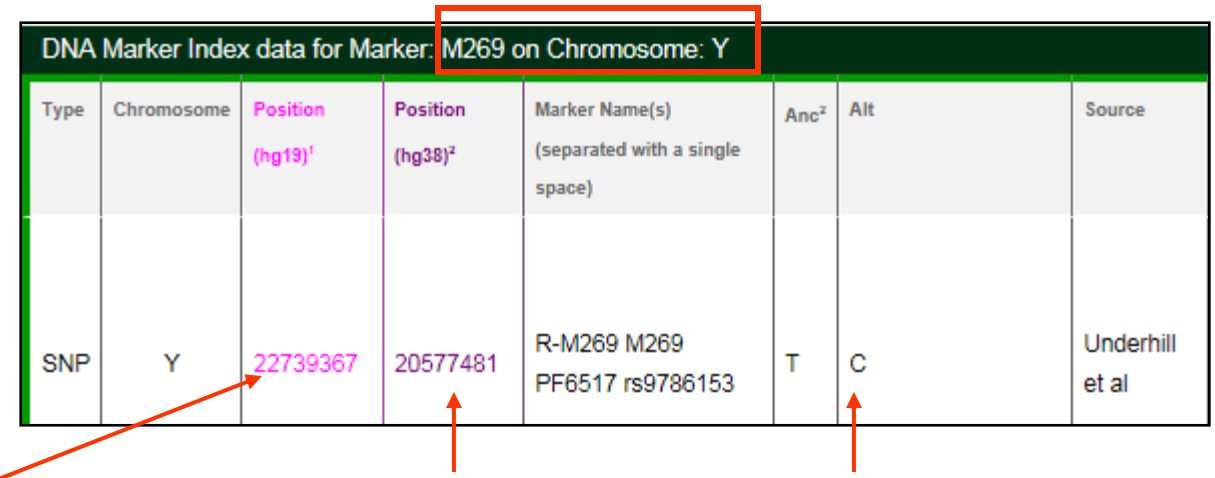
Brad Larkin, Irish Mapping DNA Project , 2014, samples with uniform 37 markers and ancestral county identified, n=165

Types of DNA Markers for Genealogy

- STRs
  - Used for clustering Y-DNA results and surname studies.
  - Change fast, thus providing good dividing points in the past 1000 years when surnames have been in use.
- SNPs
  - Used for breaking mankind into groups of people, called *Haplogroups*. Change infrequently and thus serve as major branches in the tree of man.
  - With availability of Big Y, we can do lots of group classifications at an economically feasible rate.

SNPs: Naming Biological Phenomena

- SNP Name: An SNP is a name given to a particular biological mutation at a particular position on a chromosome.
  - Giving the mutation a name is helpful as chromosome positions often shift in each version of the human genome assembly.
    - e.g. with M269, the position was 22739367 in the previous (hg19) assembly.



DNA Marker Index data for Marker: M269 on Chromosome: Y

Type	Chromosome	Position (hg19)[1]	Position (hg38)[2]	Marker Name(s) (separated with a single space)	Anc[2]	Alt	Source
SNP	Y	22739367	20577481	R-M269 M269 PF6517 rs9786153	T	C	Underhill et al

Here the SNP M269 is enumerated as a T->C nucleotide mutation at position 20577481 on chromosome Y in assembly hg38.

Also note that the current *Genome Reference Consortium Human Reference 38* (GRCh38) is a composite based on three (3) individuals in Y haplogroup R – so the Reference allele is not always the same as what we now believe the ancestral allele was.

e.g. M269's Ref nucleotide allele is listed as *C* in GRCh38 but we believe the actual ancestor had allele T.

SNPs: What's In a Name

- SNP names typically reflect the discovering entity and a serial number for that discovery.

– The letter prefix(es) typically represent the laboratory which identified the marker.
  - e.g. "M269"
  - See list of more SNP prefixes on ISOGG Tree
– The numbers following the prefix are just a discovery sequence number and have no biological meaning.
  - e.g. "M269"

DNA Marker Index data for Marker: M269 on Chromosome: Y

Type	Chromosome	Position (hg19)[1]	Position (hg38)[2]	Marker Name(s) (separated with a single space)	Anc[2]	Alt	Source
SNP	Y	22739367	20577481	R-M269 M269 PF6517 rs9786153	T	C	Underhill et al

"M" of "M269" => Peter Underhill, Ph.D. of Stanford University.
"PF" => Paolo Francalacci, Ph.D., Università di Sassari, Italy

SNPs: Multiple Names

- Because multiple laboratories and researches may discover and name mutations independently, the same position and mutation may be called by multiple names.
  - Despite its original abbreviation, in practice the term *SNP* is used to refer to single and multi-nucleotide mutations as well as insertions and deletions of DNA.
    - **Insertions** may be represented with "ins" in the ancestral or reference allele or by including the preceding nucleotide.
      - e.g. T->TC
    - **Deletions** may be represented with "del" in the derived or Alt allele values.
      - e.g. AT->A

DNA Marker Index data for Marker: M269 on Chromosome: Y

Type	Chromosome	Position (hg19)[1]	Position (hg38)[2]	Marker Name(s) (separated with a single space)	Anc[2]	Alt	Source
SNP	Y	22739367	20577481	R-M269 M269 PF6517 rs9786153	T	C	Underhill et al

In this example, SNP names *R-M269, M269, PF6517*, and *rs9786153* are all synonymous. They are just different expressions or labels for the same biological phenomenon: a T->C mutation at position 20577481 in the hg38 assembly.

SNPs: Multiple Historical Instances

- Because the same mutation can occur at the same position multiple times over human evolution, nomenclature distinctions for each mutation instance have arisen.
  - Preferred Method: **Append a decimal suffix** to the SNP names
    - Such as .1, .2, etc
    - Typically .1 would be first discovered instance; .2 would be second instance
  - Alternative Method: **haplogroup-SNP nomenclature**
    - Prepend the root haplogroup to the SNP name.
      - Such as *I-S7642* versus *R-S7642*
- Weakness of this method is that same mutation can occur multiple times within the SAME major haplogroup.
  - e.g. S6932.1 and S6932.2 both found under haplogroup R1b would be ambigious two instances of **R-S6932**.

DNA Marker Index data for Marker S7642 on Chromosome: Y

Type	Chromosome	Position (hg19)[1]	Position (hg38)[2]	Marker Name(s) (separated with a single space)	Anc[2]	Alt	Source	Notes
SNP	Y	16651591	14539711	I-S7642 S7642.1 S7642 rs758721439	T	C	Jim Wilson 2013	Found in haplogroup I. See also S7642.2 found in haplogroup R1b. Phylogenetic Parent Marker: S4767 Phylogenetic Child Markers: FGC23267 FGC29185 S11236 Y17933 Y17938 Y21671 Y24819 Y6339 Y6340 View Pedigree
SNP	Y	16651591	14539711	R-S7642 S7642.2 rs758721439	T	C	FTDNA 2019	Found in haplogroup R1b. See also S7642.1 found in haplogroup I. Phylogenetic Parent Marker: S10621 View Pedigree

Here the T->C mutation called S7642 has been found in two separate sets of people (haplogroups). The mutation may be labeled *S7642.1* or *I-S7642* for the first set and *S7642.2* or *R-S7642* or for the second set.

SNPs: Multiple Positions

- While ostensibly an SNP should occur at only one chromosome position in a given assembly, there are special-case exceptions.
  - This multi-position phenomenon occurs most often when the mutation is located on an unstable region of DNA called a *pallindrome* where the entire region may have shifted.
  - SNP name nomenclature often uses **Underscore Suffixes** such as *_1*, *_2* indicate instances where the same mutation has been found at different chromosome positions.

DNA Marker Index data for Marker Z39589% on Chromosome: Y

Type	Chromosome	Position (hg19)[1]	Position (hg38)[2]	Marker Name(s) (separated with a single space)	Anc[3]	Alt	Source
SNP	Y	4439912	4571907	R-Z39589 Z39589 Z39589_2	GCAGCTTCACTCCTGAGG	del	Alex Williamson 2016
SNP	Y	4439912	4571871	R-Z39589_1 Z39589_1	GCAGCTTCACTCCTGAGG	del	Alex Williamson 2016

In this example, mutation Z39589 (an 18 base pair deletion) has been observed both at positions 4571907 AND 4571871 of the hg38 assembly. The suffixes clarify which instance is being referred to.

Y Chromosome Structure

- Recap with note of concentration of Pallindromes, Genes, STRs, and SNPs along the length of the Y-chromosome.

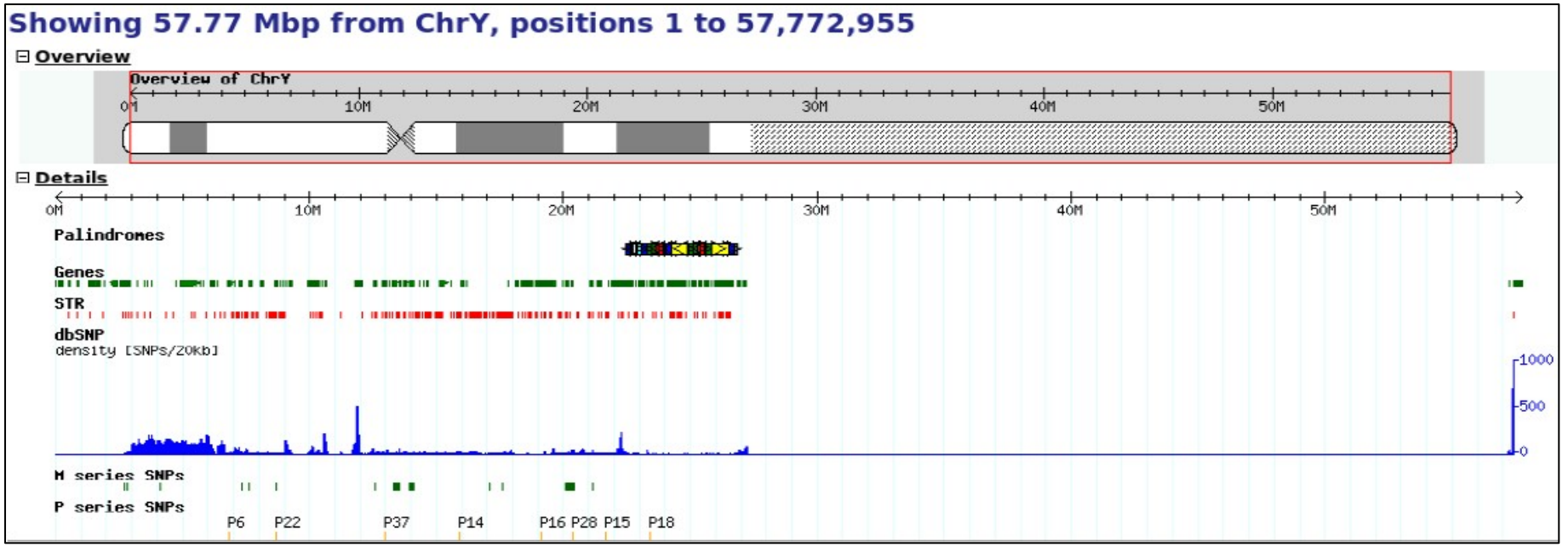


Image adapted from Thomas Krahn, *Walk On Y Project* presentation at FTDNA 2007 Conference.

Outline

- Intro to Y-DNA Genetic Genealogy
- Phylogenetic Trees
  - Trees from SNPs
  - Definition of Terminal SNP
  - Growth of Y-DNA Tree in Haplogroups R, I, and J
  - Aggregate Statistics for SNPs and Trees
  - Sampling Summary by Haplogroup and Country
- Comparing Big Y to other Y-DNA tests
- How to Use YOUR Big Y Results

Phylogenetic Trees

- Trees from SNPs
- Definition of *Terminal SNP*
- Growth of Y-DNA Tree in Haplogroups R, I, and J
- Aggregate Statistics for SNPs and Trees
- Sampling Summary by Haplogroup and Country

Phylogenetic Tree from SNPs Truth, not Probability

- Abraham was the father of Isaac[1]
- Isaac the father of Jacob
- Jacob the father of Judah and his brothers
- Judah the father of Perez

- A man with SNP B253 had a male descendant with a mutation we call M207
- A man with M207 had a male descendant with a mutation we call M173
- A man with M173 had a male descendant with a mutation we call M343
  - M343 is the marker shared by all R1b descendants

[1]Biblical genealogy from the Book of Mathew, Chapter 1, published by The International Bible Society

Tree Branches are not usually Consecutive Generations

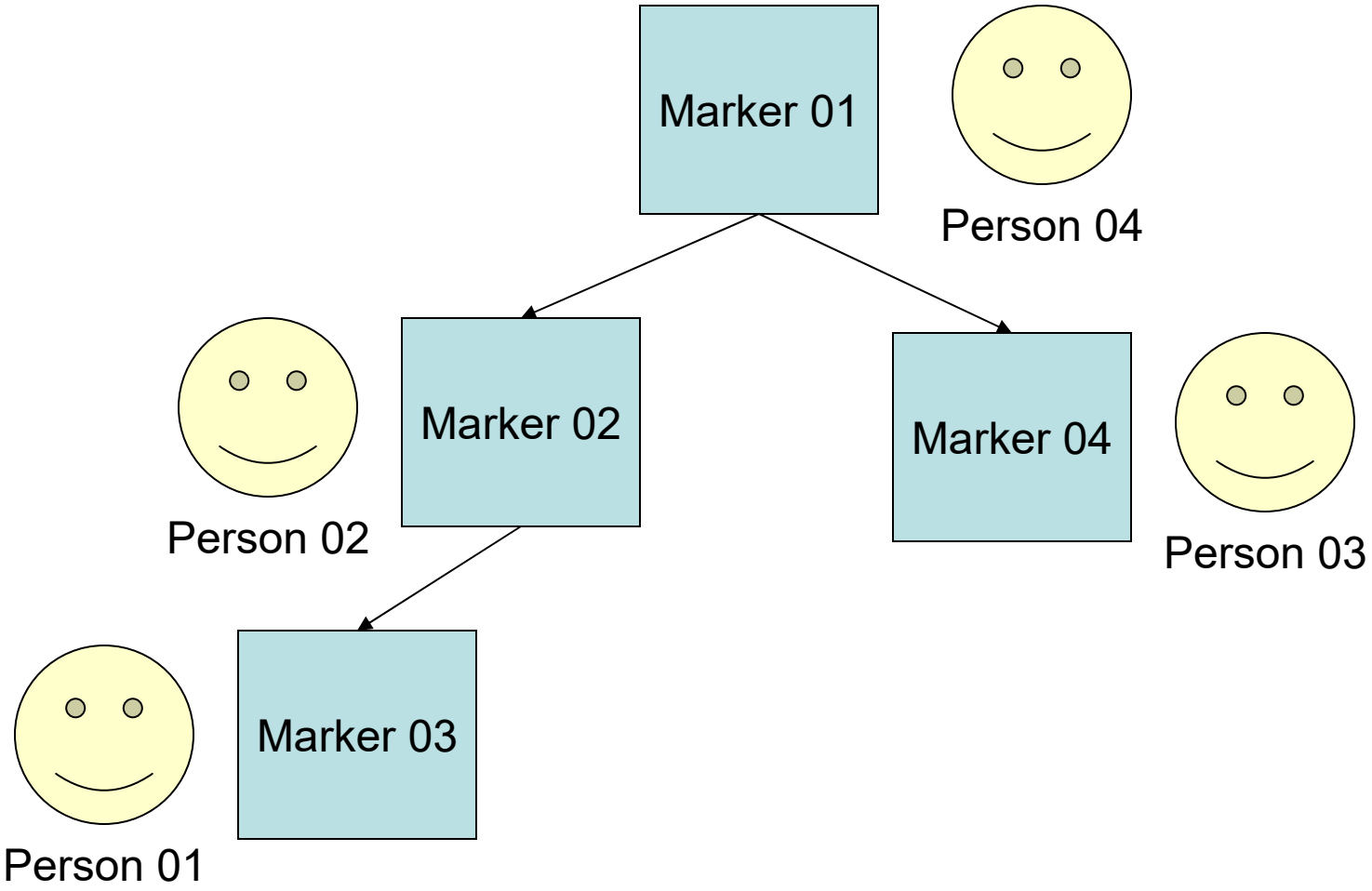
- Bear in mind the tree is not every mutation observed.
  - Just an assembly of those mutations which we have found which clearly distinguish groups of human beings.
- There are hundreds of thousands of mutations observed, but only a fraction of those have been found useful for tree branching.

Phylogenetic Classification Exercise with SNP-like binary markers

Participant	Marker 01	Marker 02	Marker 03	Marker 04
Person 01	+	+	+	0
Person 02	+	+	0	0
Person 03	+	0	0	+
Person 04	+	0	0	0

Here “+” means the participant is Positive for the SNP marker in question. “0” means negative result.

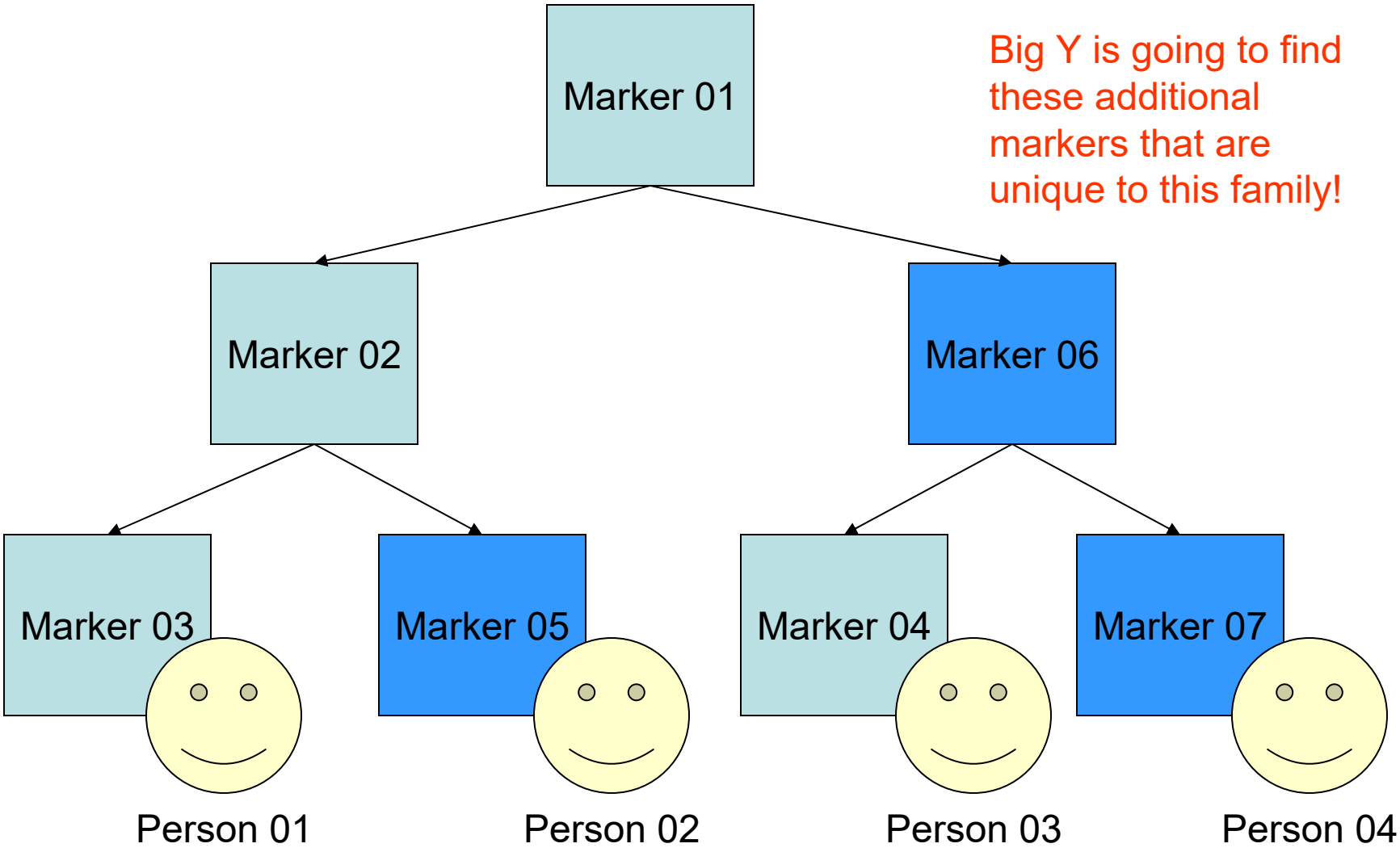
Marker Phylogeny Solution



Terminal SNP

- The term *Terminal SNP* refers to the deepest SNP marker on the Phylogenetic Tree that an individual has tested positively for.

Extending Exercise with More Markers so that Everyone has Distinct Markers



Y Phylogenetic Tree - 2002

- Diagram fit on one page!
- 18 Haplogroup letter designations
- My tree level from Homo Erectus would have been 10 steps.
  - *R1b1c7* labeling

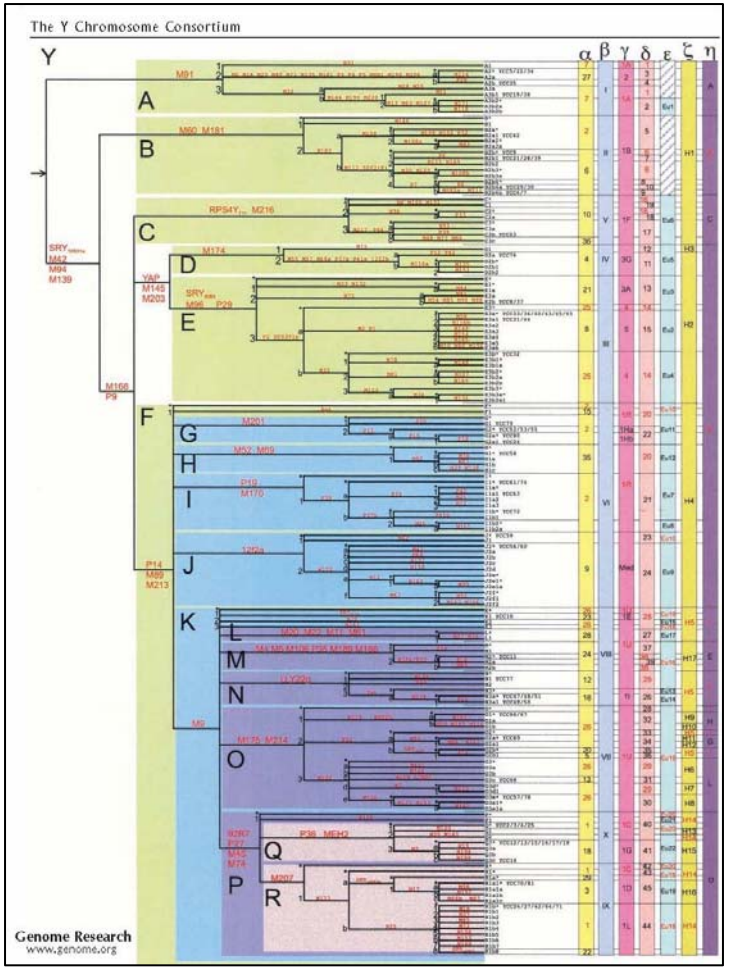


Image taken from Figure 1 from Michael F Hammer et al (2002) as The Y Chromosome Consortium in "A Nomenclauture System for the Tree of Human Y-Chromosomal Binary Haplogroups" in Genome Research 12:339-348 www.genome.org

Y-DNA Phylogenetic Inheritance Tree

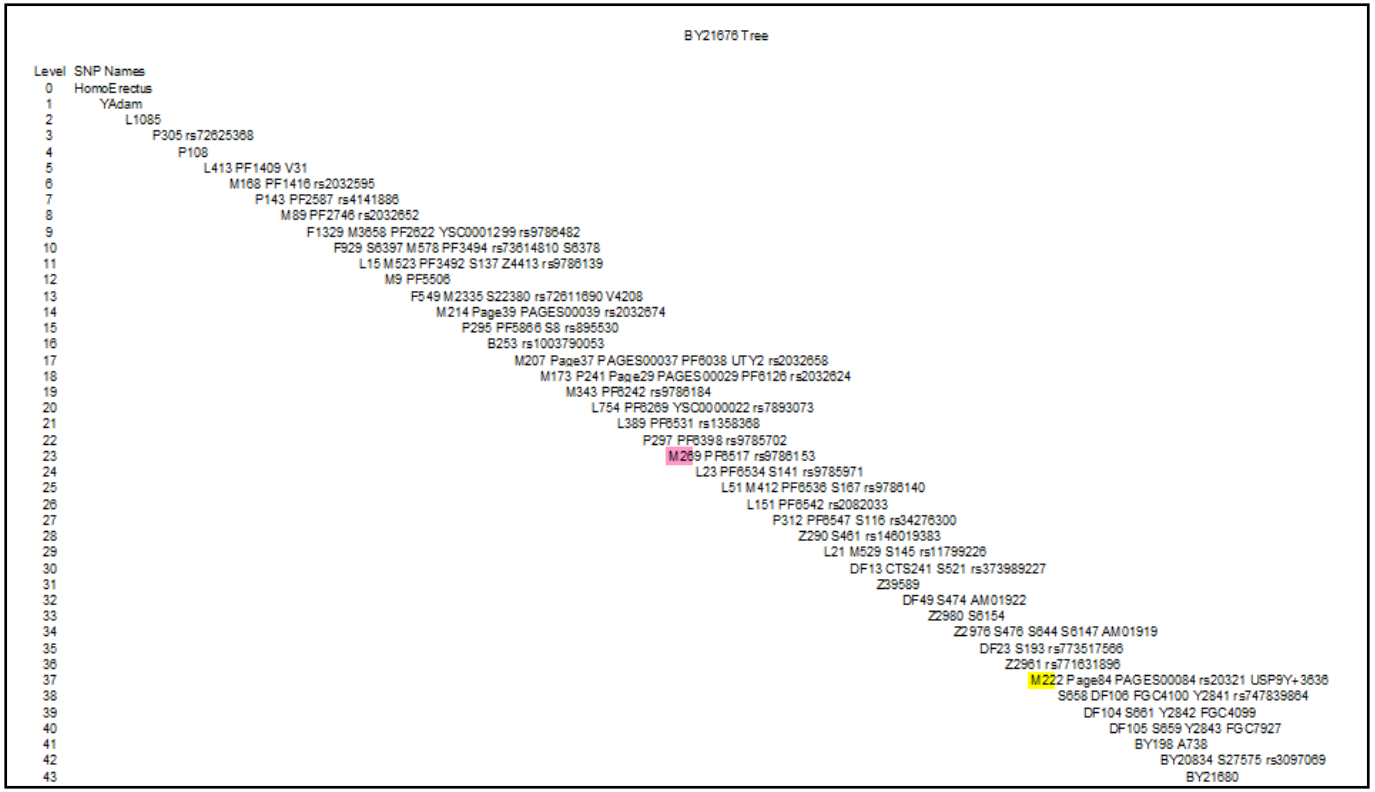
- Major haplogroups almost unchanged from 2003
  - 20 lettered haplogroups + early African branch A0
- Illustration below shows tree branches for SNP marker BY21680 as of 2018

•43 branches from Homo Erectus

•20 branches below M269

•6 branches below *M222*

•Overwhelmed original labeling convention.



Full Phylogenetic Tree for BY21680 in 2019

- Illustration of phylogenetic tree from HomoErectus to BY21680
  - 46 branch levels using Genetic Homeland Ancestral DNA Marker Pedigree Tool

1	HomoErectus	hg38:21292569-T-G	Human and Denisovan diverge from ancestral allele T found in chimpanzees at this position. hg38 Ref is G.
2	YAdam	hg38:2844421-A-G	Ancestral allele in chimpanzee is A implying this is probably a human-defining SNP for homo sapiens. hg38 Ref is G.
3	L1065	2922685-T-C	Defining mutation for near-root haplogroup A0-T (aka A0'1'2'3'4). That is to say father of ALL modern Y-DNA lineages except A00. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
4	P305	rs72625368 2842113-A-G	Defining mutation for haplogroup branch A1. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
5	P108	13314368-C-T	Defining mutation for haplogroup branch A1b. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
6	L413	PF1409 V31	Defining mutation for haplogroup branch BT. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
7	M168	PF1416 rs2032595	Defining mutation for haplogroup branch CT. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
8	P143	PF2587 rs4141886	Defining mutation for haplogroup branch CF. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
9	M89	PF2746 rs2032652	Defining mutation for haplogroup F. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
10	F1329	M3658 PF2622 YSC0001299 rs9786482	Defining mutation for haplogroup branch above GHIJK (sic). hg38 has Incorrect Ref versus SNP tree (2/24/2018).
11	F929	86397 M578 PF3464 rs73614810 86378	Mutation at haplogroup branch HIJK. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
12	L15	M523 PF3482 8137 Z4413 rs9786139	Defining mutation for haplogroup branch IJK. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
13	M9	PF5606	Defining mutation for haplogroup K. Predecessor of haplogroups T, P, and NO. Arose about 46,000 bce. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
14	F549	M2335 822380 rs72611690 V4208	Defining mutation for haplogroup branch NO (aka K2a). hg38 has Incorrect Ref.
15	M214	Page39 PAGES00039 rs2032674	Defining mutation for haplogroup branch NO1 (aka K2a1). hg38 has Incorrect Ref.
16	P295	PF5866 88 rs896530	Defining mutation for haplogroup P (K2b2). Predecessor of haplogroups R and Q. Arose about 43,000 bce. hg38 has Incorrect Ref.
17	B253	rs1003790053	Unconfirmed defining mutation for haplogroup branch P2 above QR. Arose about 38,000 bce. hg38 has Incorrect Ref.
18	M207	Page37 PAGES00037 PF5038 UTY2 rs2032658	Designates major haplogroup R* (ancestor to R1b, R1a, and R2) aka K2b2a2 in early years. Arose about 30,000 bce. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
19	M173	P241 Page29 PAGES00029 PF6126 rs2032624	Designates major haplogroup branch R1*. Arose about 26,000 bce. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
20	M343	PF6242 rs9786184	Designates major haplogroup R1b. Arose about 22,000 bce. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
21	L754	PF6269 YSC0000022 rs7893073	Defining mutation for ISOGG haplogroup branch R1b1. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
22	L389	PF6531 rs1358368	hg38 has Incorrect Ref versus SNP tree (2/24/2018).
23	P297	PF6398 rs9785702	hg38 has Incorrect Ref versus SNP tree (2/24/2018).
24	M269	PF6517 rs9786153	Designates WAMH major European branch of R1b. Arose about 11,000 bce. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
25	L23	PF6534 B141 rs9786971	Found in haplogroup R1b. Arose about 6,000 bce. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
26	L51	M412 PF6536 B167 rs9786140	Under R1b M269 > L23. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
27	P310	PF6546 B129 rs9786283	Under R1b M269. Believed coincident with P311, CTS 7650, L52, YSC0000082. Changed Ref value per YBeq 11/22/2017
28	L151	PF6542 rs2082033	Under R1b M269 > L23 > L51 according to FTDNA. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
29	P312	PF6547 B116 rs34276300	Major block under R1b. Arose about 5,000 bce. hg38 has Incorrect Ref versus SNP tree (2/24/2018).
30	Z290	8461 rs146019383	Largest branch under under M269
31	L21	M529 B145 rs11799226	Largest European group under R1b P312 > Z290. Highly correlated with geography of ancient Celts.
32	DF13	CTS241 8521 rs373989227	Directly under R1b L21.
33	Z39589		Under R1b > L21 > DF13. 18 bp deletion aka 4439911-TGCAGCTTCACTCCTGAGG-T
34	DF49	8474 AMO1922 rs769171919	Large branch under L21 Z39589
35	Z2980	86154 rs1006335787	Under L21 DF49
36	Z2976	8476 8644 86147 AMO1919 rs770179167	Under L21 DF49
37	DF23	8193 rs773917986	
38	Z2961	rs771631896	Most concentrated in Ireland.
39	Z2966	AMO1923 rs1006912931	
40	M222	Page84 PAGES00084 rs20321 USP9Y+3636	Often called Northwest Irish, concentrated in Ireland and western Scotland. Associated with Niall of the Nine Hostages and Ui Neill clans.
41	8658	DF106 FGC4100 Y2841 rs747839864	Under R1b L21 > M222
42	DF104	8661 Y2842 FGC4099 rs752415261	Under R1b L21 > M222 > DF106
43	DF105	8659 Y2843 FGC7927 rs755714899	Under R1b L21 > M222 > DF106 > DF104
44	BY198	A738	Under M222. Discovered by Thomas Krahn & Iain Kennedy. Includes Larkin Type O1 and others.
45	BY20834	827575 rs3097069	Found in R1b M222 individuals.
46	BY21680		Found in Larkin Type O1 of Ireland.

Terminal SNP Recap

- Means the deepest SNP marker on the Phylogenetic Tree that an individual has been tested for.
- Phrase gets used a lot and can be confusing as it can change:
  - When we take a new Y-DNA test
  - When results from existing tests get re-evaluated thanks to expansion of the Phylogenetic Tree.

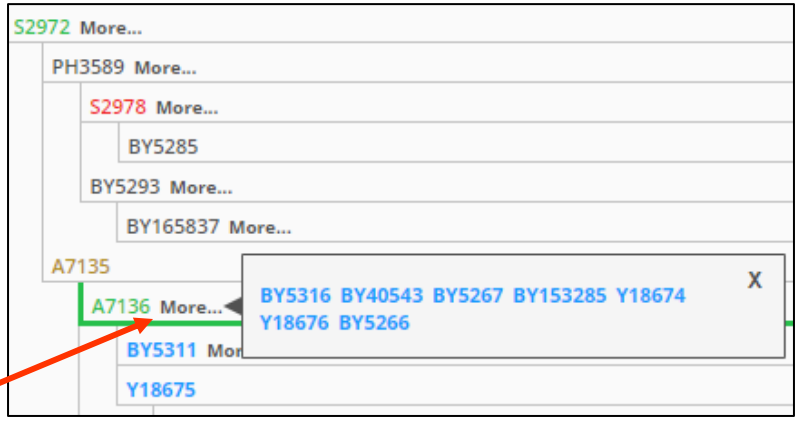
Discovery Time vs Ancestral Time

- Branch Level 40 ~ about 2200 years of ancestral time.
- M222 marker was only discovered in 2006
  - (13 years ago)
- A738 / BY198 discovered 2014
  - then only 2 examples
  - today there are 43
- BY21680 correctly located in tree in 2018

Phylogenetic Ancestral Tree for [BY21680] on Chromosome Y (Tree 1226825)

Tree Level	Marker / Branch Name	Alternative Names	Notes
40	M222	Page84 PAGES00084 rs20321 USP9Y+3636	
41	S658	DF106 FGC4100 Y2841 rs747839864	
42	DF104	S661 Y2842 FGC4099 rs752415261	
43	DF105	S659 Y2843 FGC7927 rs755714899	
44	BY198	A738	
45	BY20834	S27575 rs3097069	
46	BY21680		

Coincident Mutations



- On any single branch, coincident mutations are often observed.
  - they will be shown in a pop-up when the *More…* link is clicked on FTDNA's tree.
    - e.g. BY5316 and BY40543 are expected to be positive whenever A7136 is positive.
  - These mutations may become separate branches in the future as additional samples are collected.
  - Or they may merely represent mutations that occurred prior to the next branch in the tree.
- Coincident mutations can be very helpful in cases where a 'back-mutation' of one base pair to the ancestral condition occurs.

Status of Human Y-DNA Sampling and Database Size

- In the long run, we will assemble a definitive phylogenetic tree for all humanity
  - for all geographies
  - for all haplogroups
- In terms of the immediate value of your Big Y results, some perspective on the database size and sample attribution may be helpful.
- Remember that your father's Big Y result is going to have distinct SNPs that will extend the Phylogenetic Tree of mankind at some point
  - Regardless of whether your paternal ancestry is from a well-sampled population

How many SNPs

- Currently 645,721 SNPs cataloged on the Y chromosome (excluding RSIDs).[1]
- Number of branches on the phylogenetic trees[2]:
  - Y chromosome: 25,728
  - M chromosome: 6,590
- Not every SNP is on the tree.
- Haplogroups are not uniformly sampled.

Figures from Genetic Homeland [1]DNA Marker Index and [2]DNA Ancestral Pedigree Tree as of 30 May 2019.

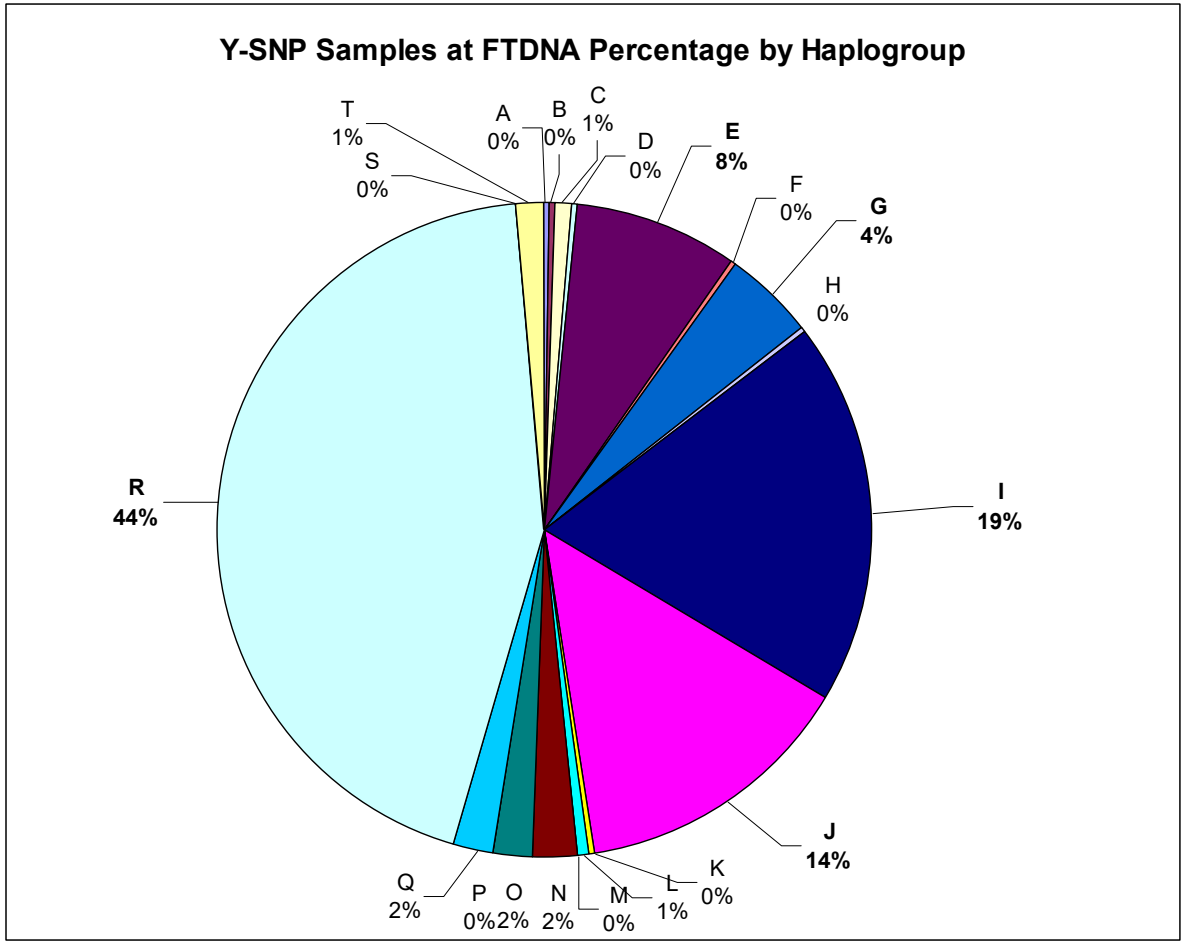
Big Y Database Size

- FTDNA's Y-Database is believed to be the world's largest.
- Official statistics as of 2 Feb, 2019[1]:
  - 713,091 Y-DNA records in database
  - 369,088 Y-37 STR results
- *Unofficial Estimate* based on Big Y vs. Y-37 adoption rate of 47%[2]:
  - ➢173,000 Big Y records in database

[1]Family Tree DNA, Why Family Tree DNA? About The Family Tree DNA Database https://www.familytreedna.com/why-ftdna.aspx

[2]Irish Mapping DNA Project at Family Tree DNA, Project Statistics, https://www.familytreedna.com/groups/irishmapping/about/project-statistics

Y-SNP Samples by Haplogroup

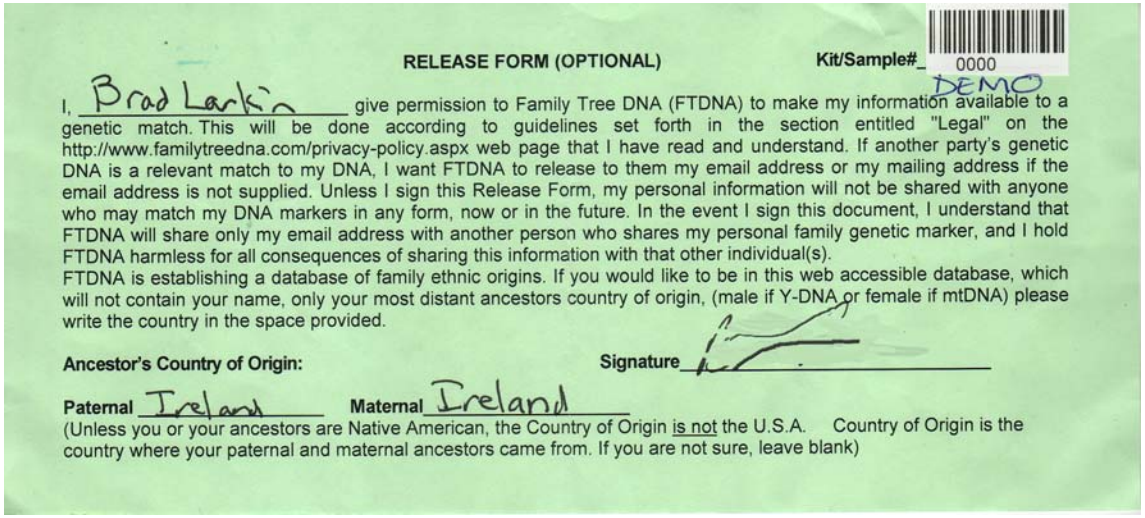


Some High Level Y-DNA Haplogroups of Western Europe

- R-L21 ~ Celtic
- R-U106 ~ Germanic
- R1a & I1 ~ Scandanavian
- J ~ Sephardic, Middle Eastern
- G ~ Ancient Hunter Gatherers
- E ~ African

Sampling Penetration by Country

- Number of Matches can only be function of how many other persons from your group have already been tested with results in the same database.
  - Count samples by the country which is attributed by participants to be the origin of their earliest known paternal ancestor.



RELEASE FORM (OPTIONAL) **Kit/Sample#** 0000 DEMO

I, Brad Larkin give permission to Family Tree DNA (FTDNA) to make my information available to a genetic match. This will be done according to guidelines set forth in the section entitled "Legal" on the http://www.familytreedna.com/privacy-policy.aspx web page that I have read and understand. If another party's genetic DNA is a relevant match to my DNA, I want FTDNA to release to them my email address or my mailing address if the email address is not supplied. Unless I sign this Release Form, my personal information will not be shared with anyone who may match my DNA markers in any form, now or in the future. In the event I sign this document, I understand that FTDNA will share only my email address with another person who shares my personal family genetic marker, and I hold FTDNA harmless for all consequences of sharing this information with that other individual(s).

FTDNA is establishing a database of family ethnic origins. If you would like to be in this web accessible database, which will not contain your name, only your most distant ancestors country of origin, (male if Y-DNA or female if mtDNA) please write the country in the space provided.

Ancestor's Country of Origin: **Signature**

Paternal Ireland **Maternal** Ireland

(Unless you or your ancestors are Native American, the Country of Origin is not the U.S.A. Country of Origin is the country where your paternal and maternal ancestors came from. If you are not sure, leave blank)

Y-DNA World Map - Pre-Colonial

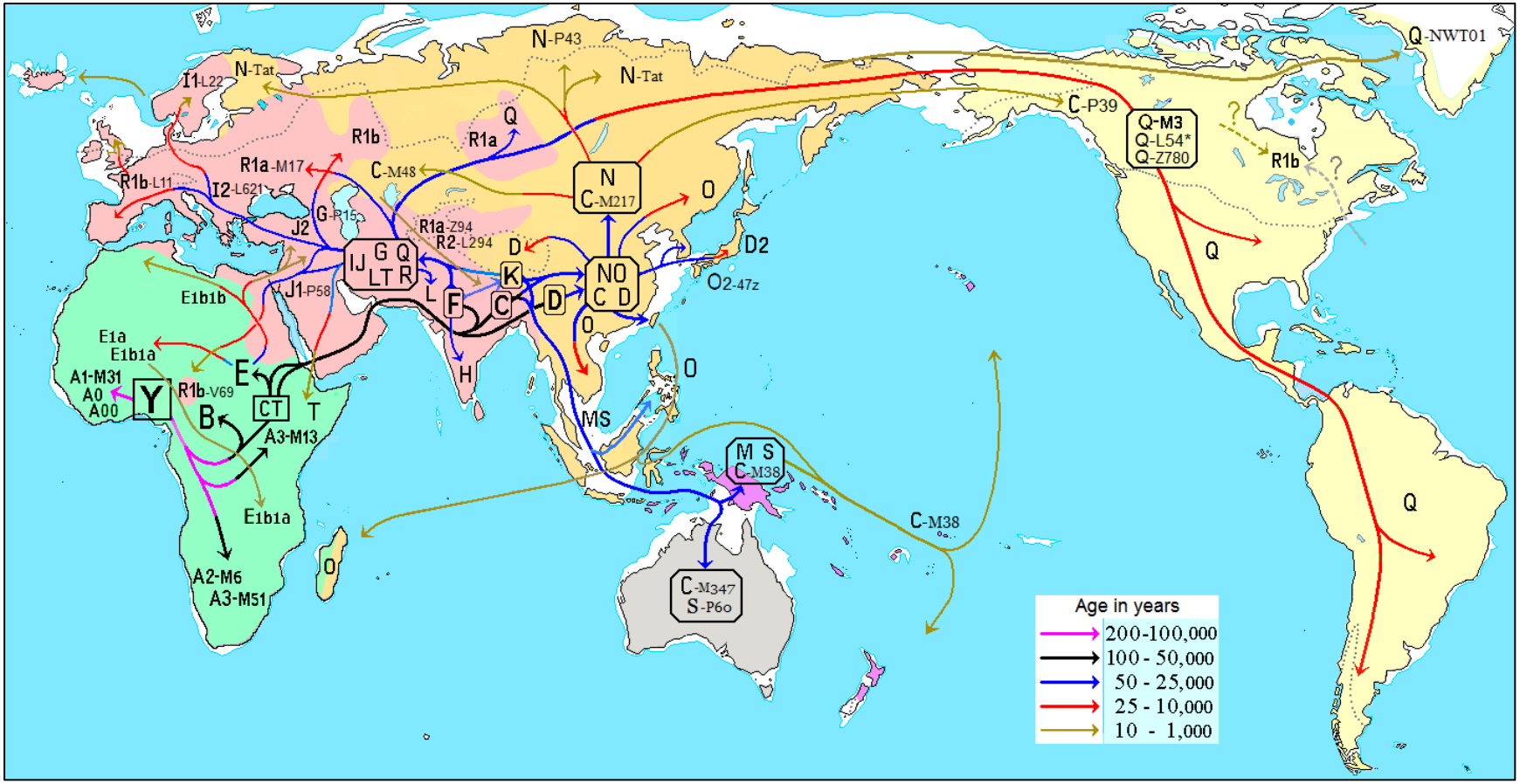


Image by Chakazul (2013) on Wikipedia, [World Map of Y-DNA Haplogroups]

Y-DNA World Map - Today

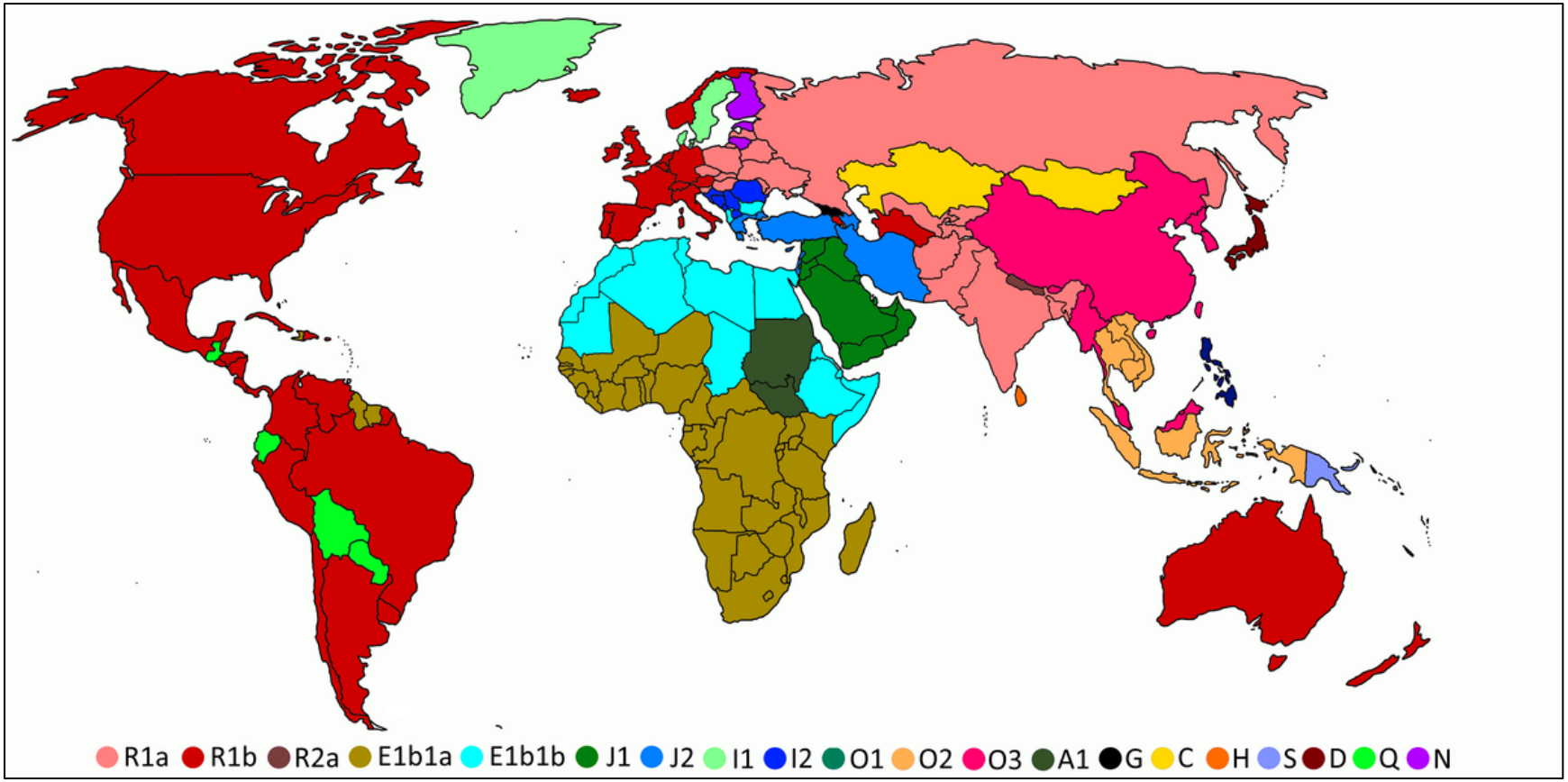
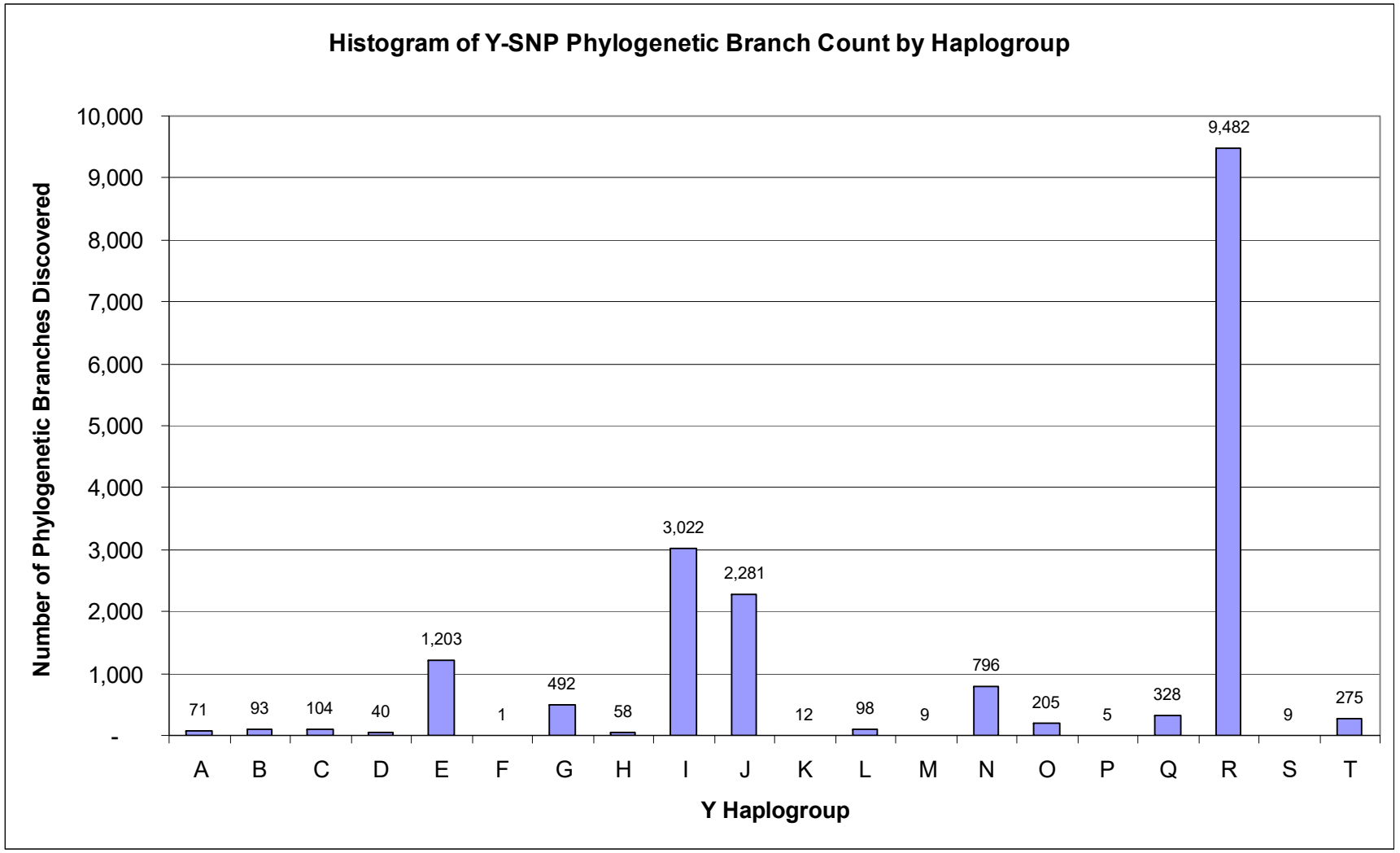
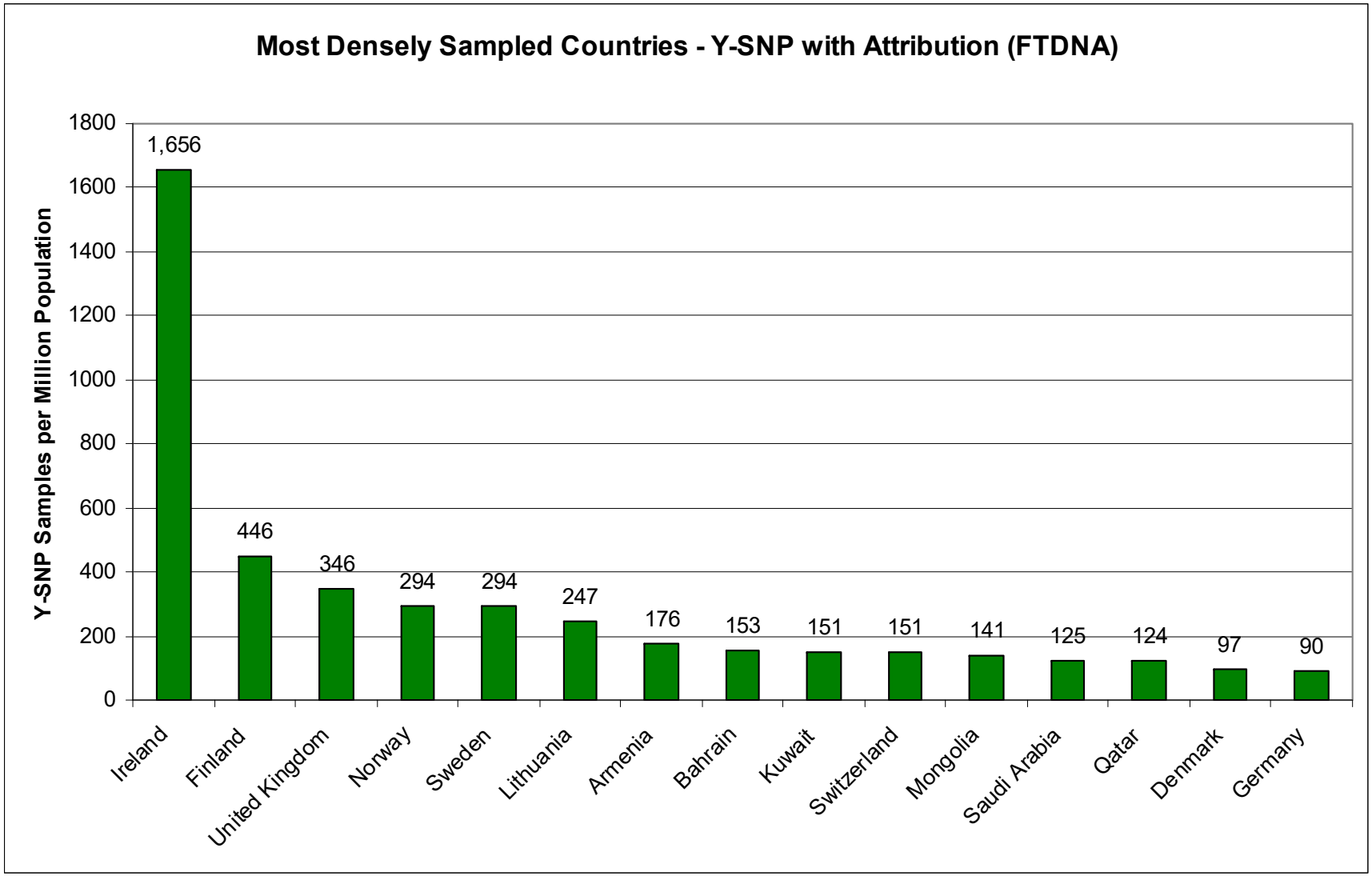


Image by Costa (2017) on genetics – Haplogroups message board on TheAPricity.com



Figures from Genetic Homeland DNA Ancestral Pedigree Tree as of 23 Feb 2019.



Population data from United Nations 2017 World Population Prospects on Wikipedia
https://en.wikipedia.org/w/index.php?oldid=879939654

Review

- The Phylogenetic Tree
  - Tells us who our paternal ancestors were.
  - Let's us compare our result to another person based on terminal SNPs.
- Big Y or comparable test gives you thousands of SNP results
  - Many SNPs will already be on the phylogenetic tree
  - Some will be added in future as your paternal kin become better sampled

How Complete is the Y Phylogenetic Tree?

- Getting very complete for Haplogroup R1b.
- Y-SNP sample coverage over 400 samples per million from British Isle lineages
- Still very low coverage in Asia and Africa

Outline

- Intro to Y-DNA Genetic Genealogy
- Phylogenetic Trees
- Comparing Big Y to other Y-DNA tests
  - History of Big Y
  - Competitors and Costs with Equivalent Tests
  - SNP Packs
  - STR packages
  - Privacy and Law Enforcement
- How to Use YOUR Big Y Results

Comparing Big Y to Other Tests

- History of Big Y
- Competitors and Costs with Equivalent Tests
- SNP Packs
- STR packages
- Privacy and Law Enforcement

Big Y Product Timeline

- 2013 Product Announced by FTDNA
  - Only SNPs, no STRs, long results time.
  - Could not be used on initial order, required an STR order then upgrade.
- 2017 Big Y-500
  - 500 STRs added, SNP matching enhanced.
  - Could be used as initial order.
- 2019 Big Y-700
  - 200 more STRs added
  - Extraction & chemical prep improvement for more consistent coverage
  - Expanded range of chromosome sampling

Markers Tested

- Original Big Y yielded an average of about 10.3 million base pairs per sample.
- New Big Y-700 test expected to expand the coverage regions AND provide more consistent reads of those regions.
  - FTDNA using new series of SNP-designators based on Big Y-700 results: *FTx*
  - Leads to more SNP tree branch identification.
    - Example from haplogroup J-ZS1716 went from 7 to 9 SNP-based tree branches among a set of 11 samples.[1]
    - Roughly 30,000 more measurable SNPs with Big Y-700 versus prior version.[1]

    => SNP branches among known cousins

[1] Big Y-700 White Paper by Davis, Sager, et al. as of 22 Mar 2019.

Original Big Y Tested Regions

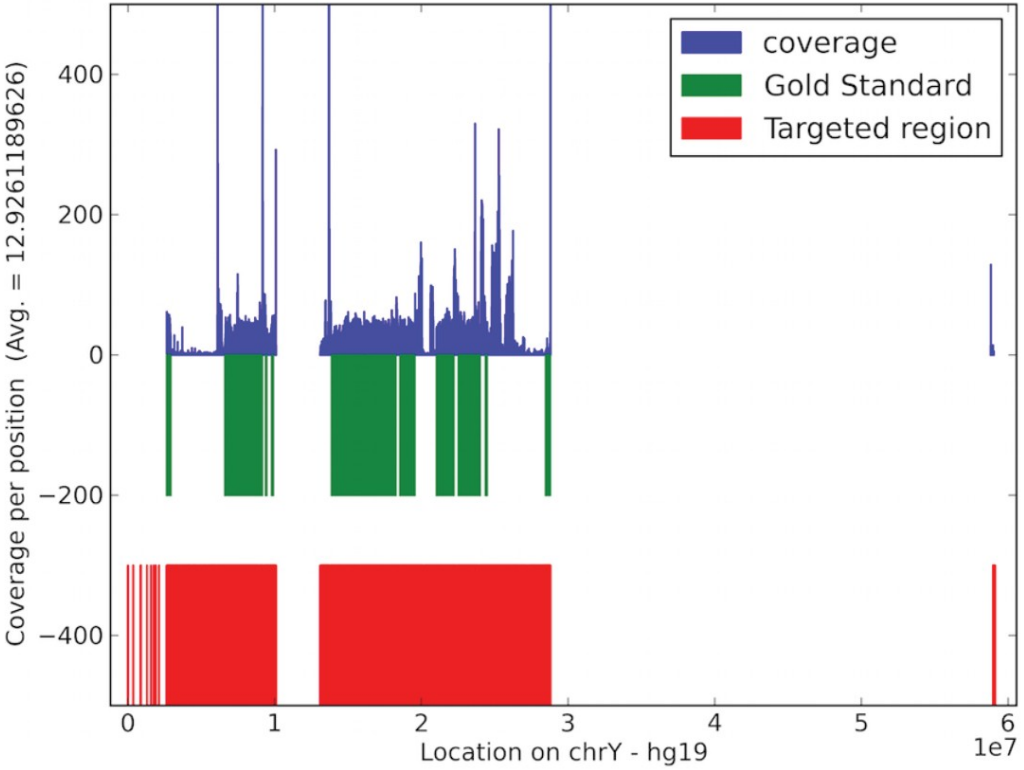


Fig1: Average coverage of 1,000 BigY samples across the Y chromosome. In this figure, the red bars indicate the regions on the Y chromosome that were targeted for sequencing. The green bars indicate the Gold Standard regions. The red and green bars are on the negative Y axis indicate presence only and are not indicative of coverage. Target regions are covered if at least one read covers the position. The blue bars indicate the sequencing coverage across the Y chromosome.

Image is Figure 1 from FTDNA Big Y Whitepaper (2014)
https://www.familytreedna.com/learn/wp-content/uploads/2014/08/BIG_Y_WhitePager.pdf

Costs

- Full retail price of Big Y-700: $ 649
  - Discounts if previous Y DNA tests have been taken
  - Also look for sales
    - e.g. International DNA Day: April 25th
    - sometimes holidays
- Cost per potential marker ~ 5 cents

Competition

- There are other companies offering comprehensive coverage of millions of Y-chromosome markers
  - But not Ancestry and not 23andMe
- FamilyTree DNA is the company offering this kind of service with a Y-chromosome focused, genealogical matching service.
- FamilyTree DNA is the oldest genetic genealogy company and has the largest database of Y-DNA tests.
- FamilyTree DNA has a seasoned network of Project Administrators that help guide members in testing strategy and interpretation.

YSeq

- Offers Whole Genome Sequencing (e.g. all 23 chromosomes + Mitochondrial) for only $ 740 at 15x coverage.
  - https://www.yseq.net
    - Founded by Thomas Krahn, former FTDNA scientest.
    - Uses German NGS lab, extraction & offices in Houston.
  - Practical coverage of Y chromosome has been equal or better than Big Y in my comparisons.

Full Genomes Corp

- https://www.fullgenomes.com/
  - Delaware Corporation
  - DNA is sent to BGI for sequencing in China.
- Y Elite test (Big Y comparable) for only $ 395
- Whole Genome 15x $ 645
- Long Read Whole Genome for $ 2900
  - a new, emerging technology

YFull

- https://www.yfull.com
  - Based in Russia
- Not a lab.
  - Provides Interpretation of tests performed elsewhere.
- Maintains a phylogenetic tree for Y and MtDNA

Next-Gen Products and Coverage

- Coverage Ratio defines minimum number of times number of reads that align with each base.
  - e.g. Each base pair is observed with a value
  - 1000 Genomes Project was done with 4-5X coverage
- Next-Gen Coverage Ratios at current retail products

Current Next-Gen Y-DNA Genetic Genealogy Product Coverage and Pricing			
Company & Product	# Chromosomes	NGS Coverage Ratio	Price (as of 5/28/2019)
Family Tree DNA (FTDNA) Big Y-700	1	55X	$ 649
Full Genomes Corp Y Elite	1	50x at 250bp	$ 395
YSeq *Whole Genome Test*	25	15X	$ 740
Full Genomes Corp *Whole Genome 15x*	25	15X	$ 645

Big Y versus STR Packages

- FTDNA Big Y product includes approximately 700 STR results as well as the SNPs.
  - Helpful at the very end of the tree – among known cousins.
  - Helpful in getting matches with legacy samples which have not been upgraded to Big Y.

SNP and STR Parallel Mutations by Generations

- Integrating Patterns of SNPs and STRs is the correct way to think about mutations and matches.
- Root, Limbs, Branches, Twigs
- Homoplasy
  - STRs the same by coincidence as they are faster at mutating
  - Illustrated at right with two lineages part of R-L21 haplogroup

DYS385b	Lineage A	Lineage B
Generation 0	L21+ M222- DYS390=25	L21+ M222- DYS390=25
Generation 1	L21+ M222+ DYS390=25	L21+ M222- DYS390=25
Generation 10	L21+ M222+ DYS390=25	L21+ M222- DYS390=25 ZZ29+
Generation 20	L21+ M222+ DYS390=26	L21+ M222- DYS390=25 ZZ29+
Generation 30	L21+ M222+ DYS390=26	L21+ M222- DYS390=26 ZZ29+

Always Consider SNP and STR

	Lineage A	Lineage A1	Lineage B
Generation 30	L21+ M222+ DYS390=26	-	L21+ M222- DYS390=26 ZZ29+
Generation 31 (DNA Testing begins)			
STR Only	DYS390=26	DYS390=27	DYS390=26
SNP Only	L21+ M222+	L21+ M222+	L21+ ZZ29+
SNP + STR	L21+ M222+ ZZ29- DYS390=26	L21+ M222+ ZZ29- DYS390=27	L21+ M222- ZZ29+ DYS390=26

SNP vs STR Analogy 1

- SNP's are the root, trunk, and branches
- STRs are the twigs

SNP vs STR Analogy 2

- Early STR testing with 12, 25, 37 markers was kind of like locating something using only the house number portion of the address
  - If your house number was distinct enough, it could pretty confidently tell you who was and was not a close relative
  - But in common house numbers, where lots of people had the same address, very hard to know who was really closely related

SNP vs STR Analogy 2

- Think of the Y-DNA Phylogenetic Tree like the components of a street address:
  - STRs tell you a house address
    - e.g. 2500
  - SNPs tell you everything else
    - Root level: Continent
    - high level: State
    - mid level: City
    - low level: Street
  - With very level resolution testing like Big Y-700, we should be seeing the floor & room number in Y-DNA SNPs
    - Floor
    - Room Number

Address Example:
Marriott Burbank Airport
2500 N Hollywood Way
Burbank, California
North America

3rd Floor, Room 305

Next-Gen Inconsistent Observations

- Two Samples with same surname, from same part of Ireland, and several SNP matches below M-222.
  - Distinction between positive, negative, and NOT OBSERVED
    - Top example not observed for FGC4087
    - Lower example not observed for A738
    - 4 of 8 low level markers NOT OBSERVED (50%) in this real example
  - Cannot resolve the phylogeny due to non-homogenous datasets - inconsistent observations of Next-Gen sequencing
    - Resolution requires ad hoc Sanger sequencing for individual SNP candidates.
      - YSeq provides Sanger confirmation for $ 100 fee on their WGS test.

LastName	Ancestral Geography	Y-14902414 [10-] *M222 Page84 PAGES00084 rs20321*	Y-26078887 [10-] *S7073 FGC462*	Y-22540855 [10-] *S660 DF109 FGC4101 Y2845*	Y-14624294 [10-] *PF682 S569 rs9786370*	Y-17303280 [10-] *A738 BY198*	Y-18028717 [10-] *FGC4087 Y3454*	Y-8157356 [10-] *S7072 FGC449 Y2596*	Y-13686261 [10-]
Larkin	Ireland, Tipperary, , Nenagh	1	1	1	1	1	-	1	1
Larkin	Ireland, Galway, Srahaun	1	1	-	0	-	1	1	-

Big Y vs SNP Packs

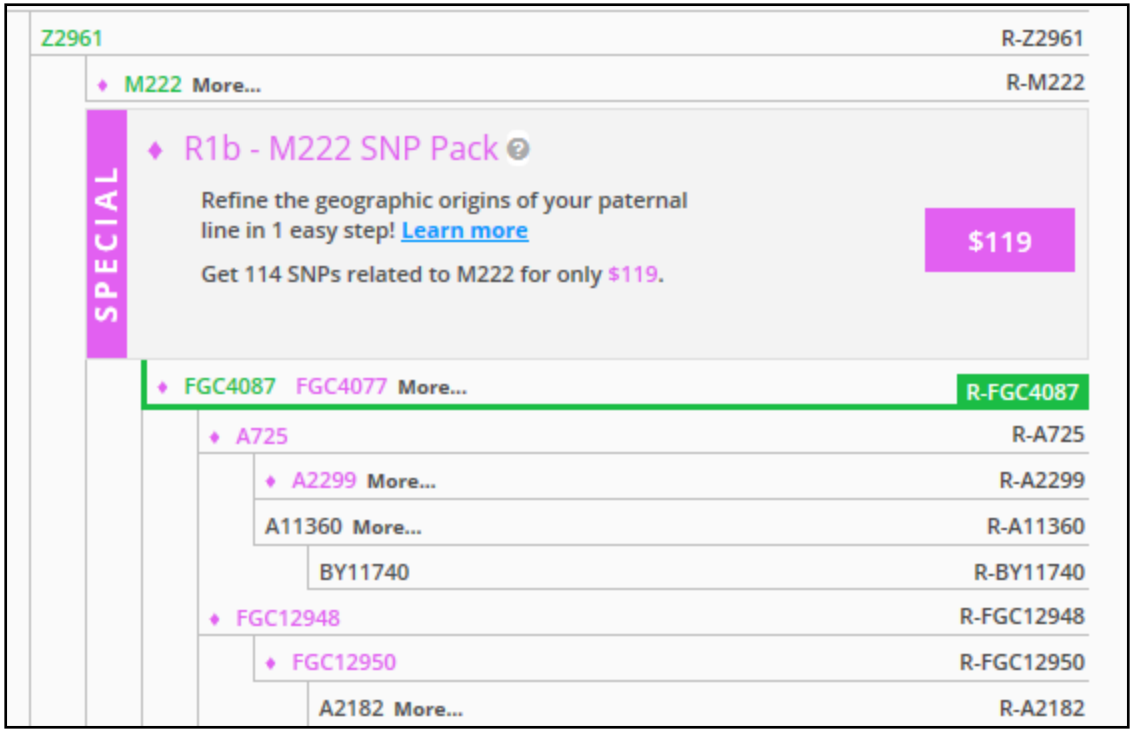
- Compromising on costs and discovery in order to gain incremental information and consistency.

LastName	Ancestral Geography	Y-14902414 [10-] *M222 Page84 PAGES00084 rs20321*	Y-26078887 [10-] *S7073 FGC462*	Y-22540855 [10-] *S660 DF109 FGC4101 Y2845*	Y-14624294 [10-] *PF682 S569 rs9786370*	Y-17303280 [10-] *A738 BY198*	Y-18028717 [10-] *FGC4087 Y3454*	Y-8157356 [10-] *S7072 FGC449 Y2596*	Y-13686261 [10-]
Larkin	Ireland, Tipperary, , Nenagh	1	1	1	1	1	-	1	1
Larkin	Ireland, Galway, Srahaun	1	1	-	0	-	1	1	-

SNP Packs

- SNP Packs are groups of lab-designed sequencing probes of about 100 markers per pack.
  - Consistent - Would expect all markers in the pack to be resolved for every test.

- Designed around specific phylogenetic subclades.
- No new discovery.
  - Depend on the SNPs already being identified.
- Conflicting nomenclature across labs
  - Genetic Homeland DNA Marker Index



Discovery

- Discovery is what the Big Y is all about
  - Millions of markers tested
    - determine which existing markers your family has
    - DISCOVER NEW SNPs unique to your family and your paternal ancestors
      - expanding the Phylogenetic tree in most cases

Privacy

- FamilyTree DNA runs their own laboratory
  - For context, some ‘testing’ companies are actually using 3rd party labs.
  - A number of such laboratories are located overseas.
- FTDNA maintains their own web site of results.
  - Important in context of GDPR and privacy concerns => does not require uploading to a 3rd party site
  - Fairly granular level of privacy and control
    - Can be changed in real-time.

Privacy – Forewarning

- Please do not ever disclose the full name and laboratory kit number of someone else.
- Law enforcement ARE trying to use DNA and genealogical databases to solve cold cases.
  - e.g. 'Golden State Killer' case
  - FTDNA's parent company is selling kits to law enforcement agencies
    - Search warrant not necessary
    - Will accept synthetic and 3rd party data files for import into database and familial match searching.
    - Default settings for opt-in / opt-out may be changed by the company over time.
- Y DNA alone cannot prove identity since all patrilineal family may have the same mutations.
  - Nonetheless we hear anecdotes of law enforcement using Y results for familial searches.

Outline

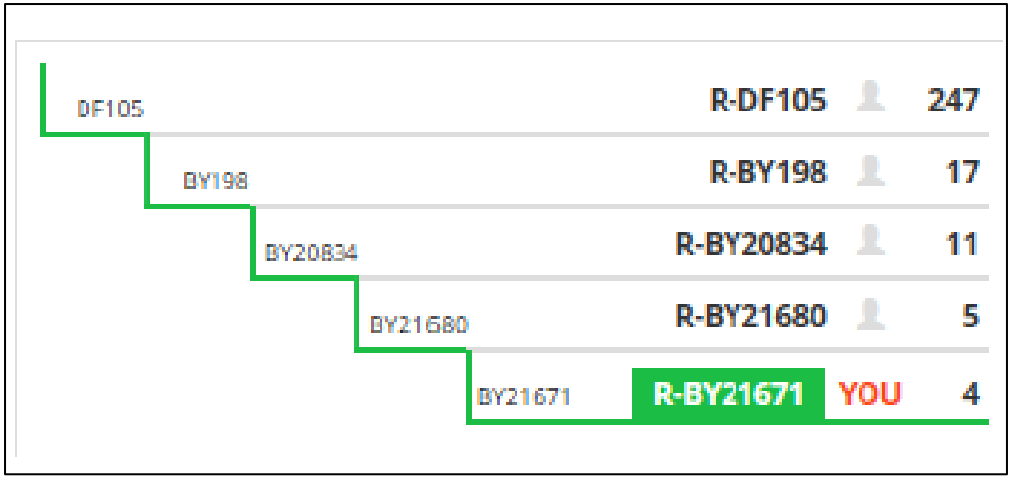
- Intro to Y-DNA Genetic Genealogy
- Phylogenetic Trees
- Comparing Big Y to other Y-DNA tests
- How to Use YOUR Big Y Results
  - Terminal SNP
  - Tree Level
  - Incorporation of non-paternal results in your genealogy
  - Comparing cousins: BAM Viewer

How to use YOUR Big Y Results

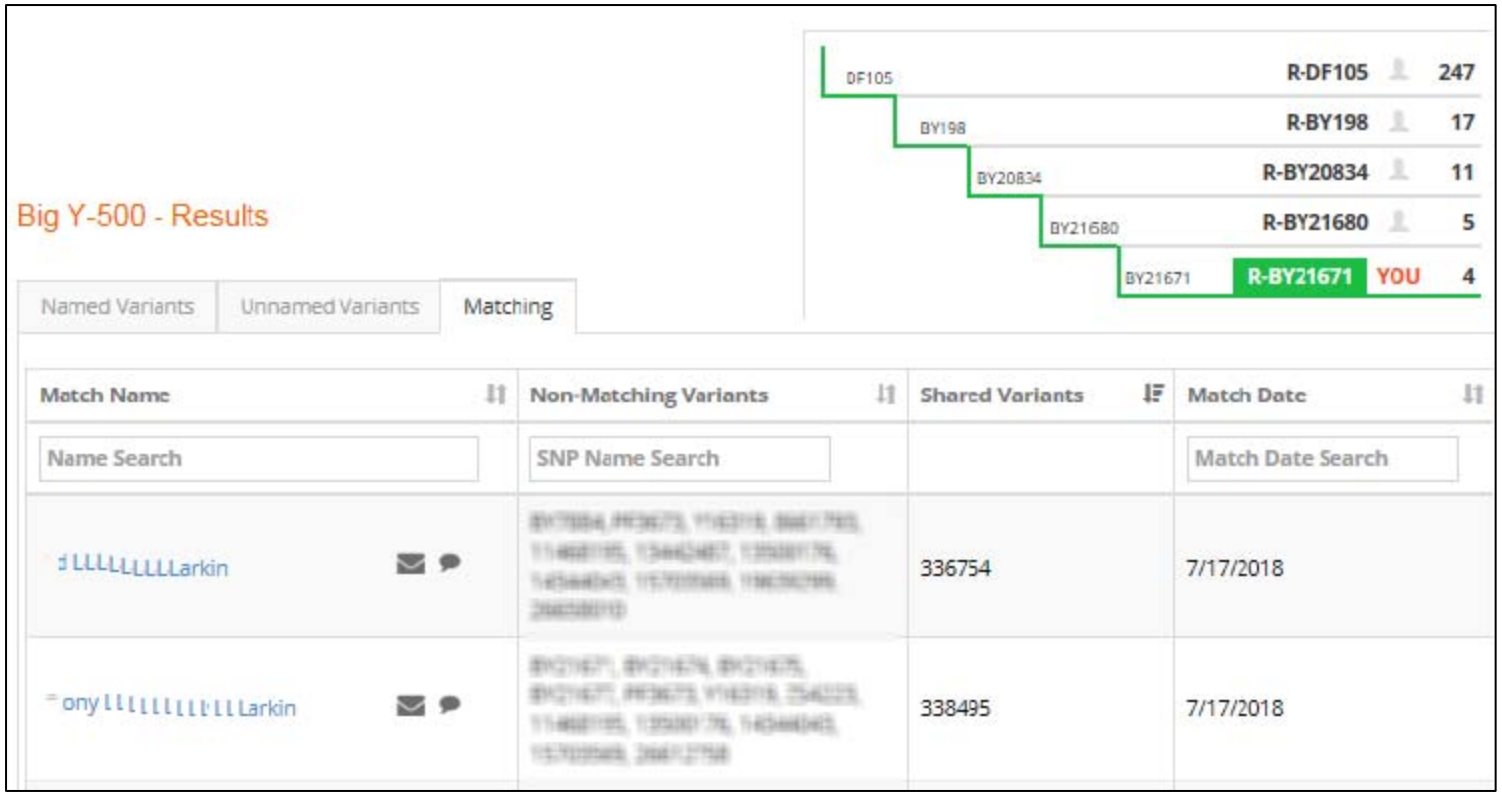
- Terminal SNP
- Tree Level
- Incorporation of non-paternal results in your genealogy
- Comparing cousins: BAM Viewer

Big Y SNP Matching

- Matches Based on your Terminal SNP in Phylogenetic Tree
  - Located on same Branch[1]
  - No False Matches
  - Independent of Surname



[1]As of Jan 30, 2019 FTDNA classifies Big Y matches based on "30 or fewer differences in SNPs with you, and their haplogroup is downstream from your haplogroup or downstream from your four closest parent haplogroups" https://www.familytreedna.com/learn/y-dna-testing/big-y/big-y/



- Some notes on the Big Y matching:
  - The Non-Matching Variants show named SNPs and the HG38 chromosome positions of unnamed variants on which you and the match do not have in common (i.e., these are biology). However, non-calls may be included in this list.
  - The Shared Variants column, however, is a tally count of how many named SNPs you and the match are both positive for. Prior to the development of the Phylogenetic Tree, this was one of the only way we had to match. It is probably an obsolete approach however, and I urge you to focus on the Phylogenetic tree.

Novel Variants

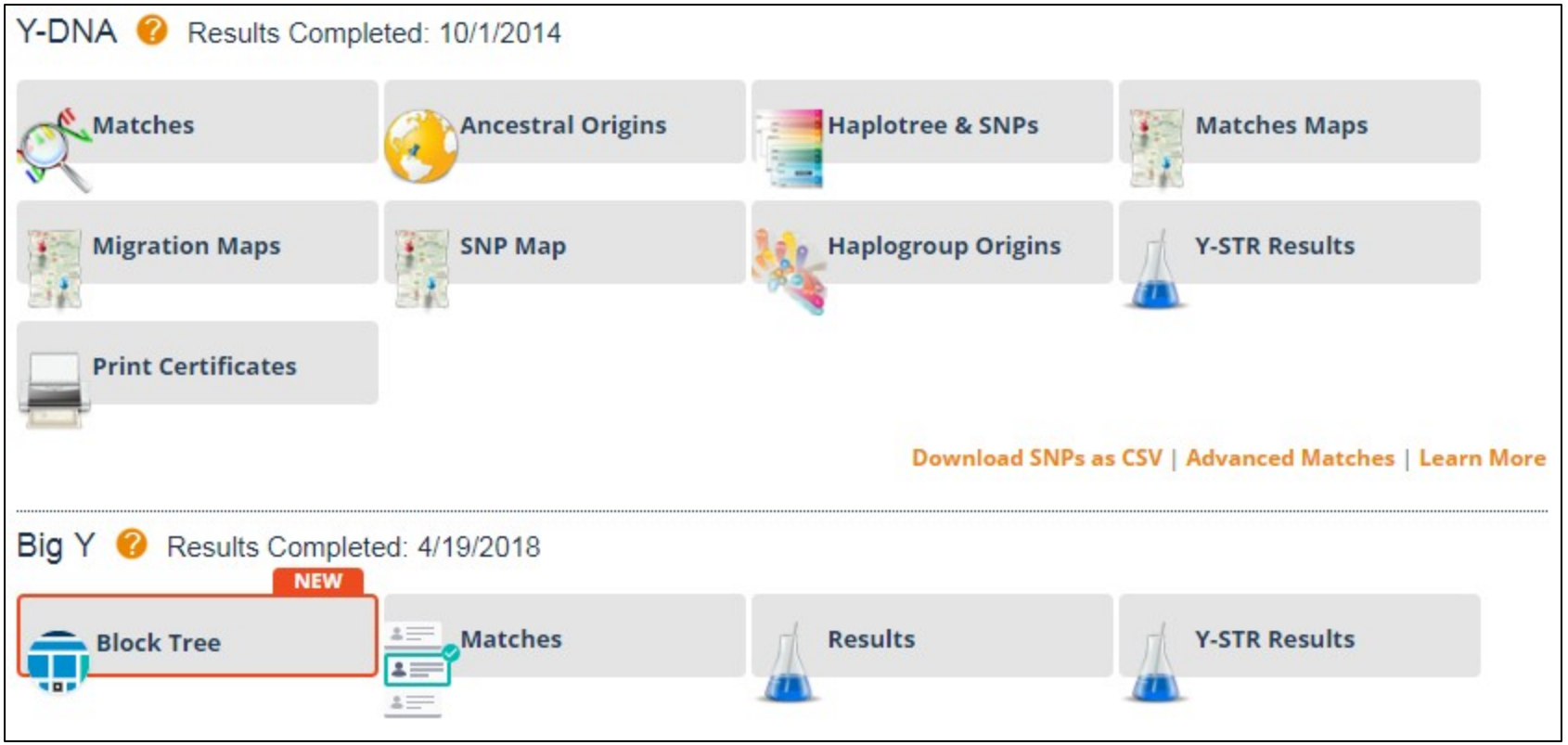
- NOVEL VARIANTS ARE YOUR UNIQUE FAMILY MARKERS!
  - aka *Unnamed Variants*
  - These are your DISCOVERIES
    - Celebrate them
  - As the global phylogenetic tree grows over time, these variants will become named SNPs.
    - Many of those SNPs will become phylogenetic tree branches – YOUR TREE BRANCHES
    - In your results, the number of unnamed variants should shrink over time as they receive SNP names
      - e.g. 78 novel variants in 2014 -> 0 unnamed variants in 2019

What you get with Big Y Test

- Original core Y-111 STR markers used in hundreds of legacy Y DNA studies and surname projects.
  - STR-based matching
- Over 11 million base pairs of potential SNPs tested.
  - SNP-based matching

Result Tools

- These standard FTDNA result tools
- Plus ability to download raw data in several formats such as CSV file, BAM file, etc.



Kinship – Comparing Terminal SNPs

- With millions of markers tested and the phylogenetic tree emerging, we can now make clear distinctions on where the ancestry of two individuals diverged thanks to Big Y.
  - FTDNA Block Tree
  - Genetic Homeland Pedigree Comparison

DNA Pedigree

Navigate to...

GeneticHomeland.com Mapping Technology for DNA, Surname & Genealogy Research

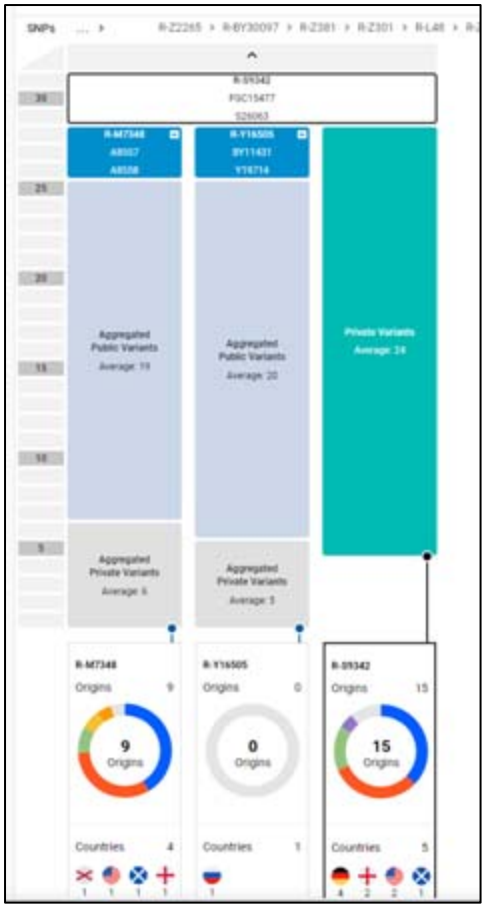
Ancestral DNA Marker Pedigree Display

Phylogenetic Ancestral Tree for [BY21680] on Chromosome Y (Tree 1226825) Compared to [A725 rs1042558592] (Tree 1107410)

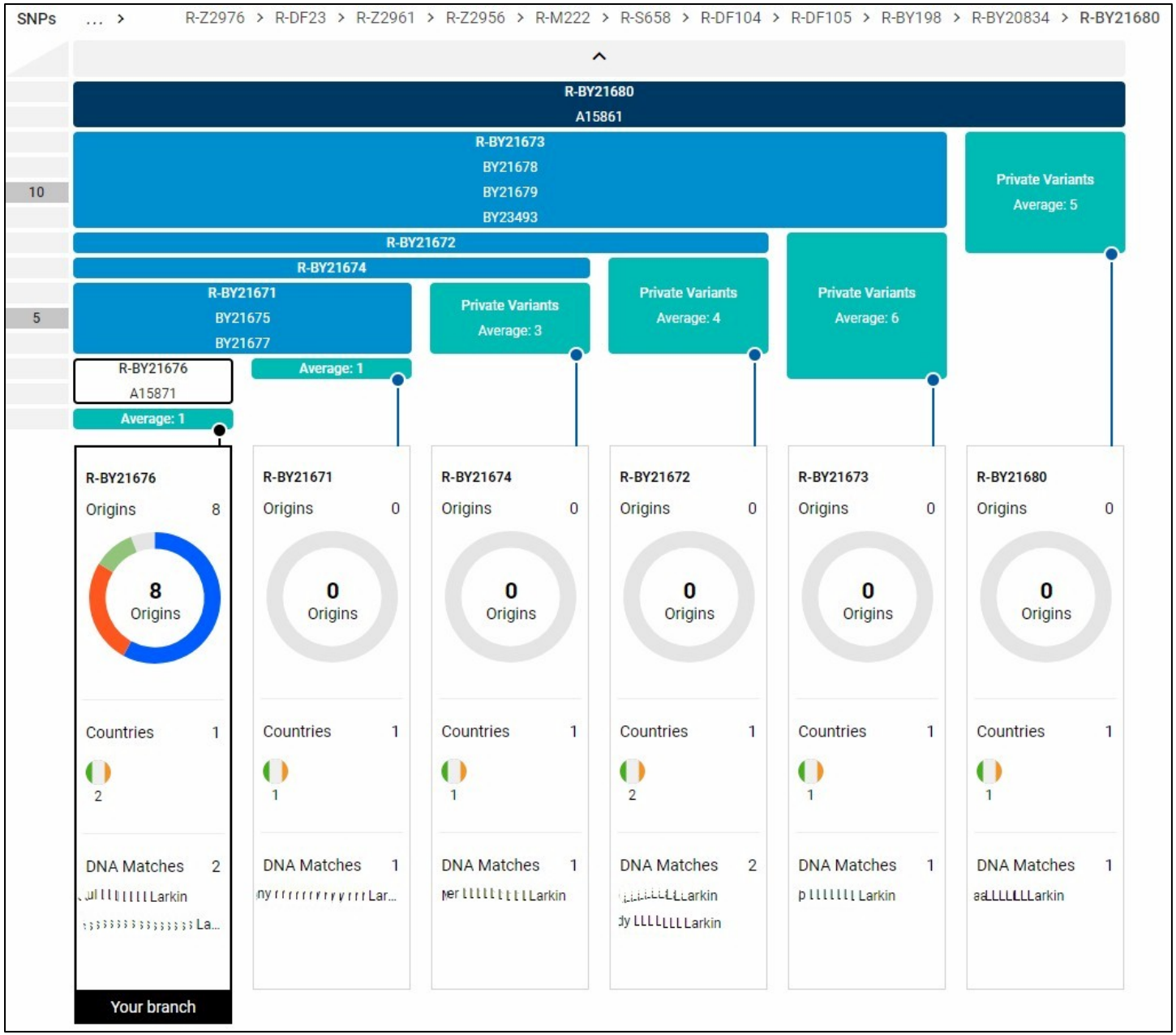
Tree Level	Marker / Branch Name	Comparison [A725]
38	Z2961 rs771631896	-same-
39	Z2956 AM01923 rs1006912931	-same-
40	M222 Page84 PAGES00084 rs20321 USP9Y+3636	-same-
41	S658 DF106 FGC4100 Y2841 rs747839864	DIVERGENCE: BY35297
42	DF104 S661 Y2842 FGC4099 rs752415261	FGC4077 Y3455
43	DF105 S659 Y2843 FGC7927 rs755714899	A725 rs1042558592
44	BY198 A738	-
45	BY20834 S27575 rs3097069	-
46	BY21680	-

Big Y Block Tree

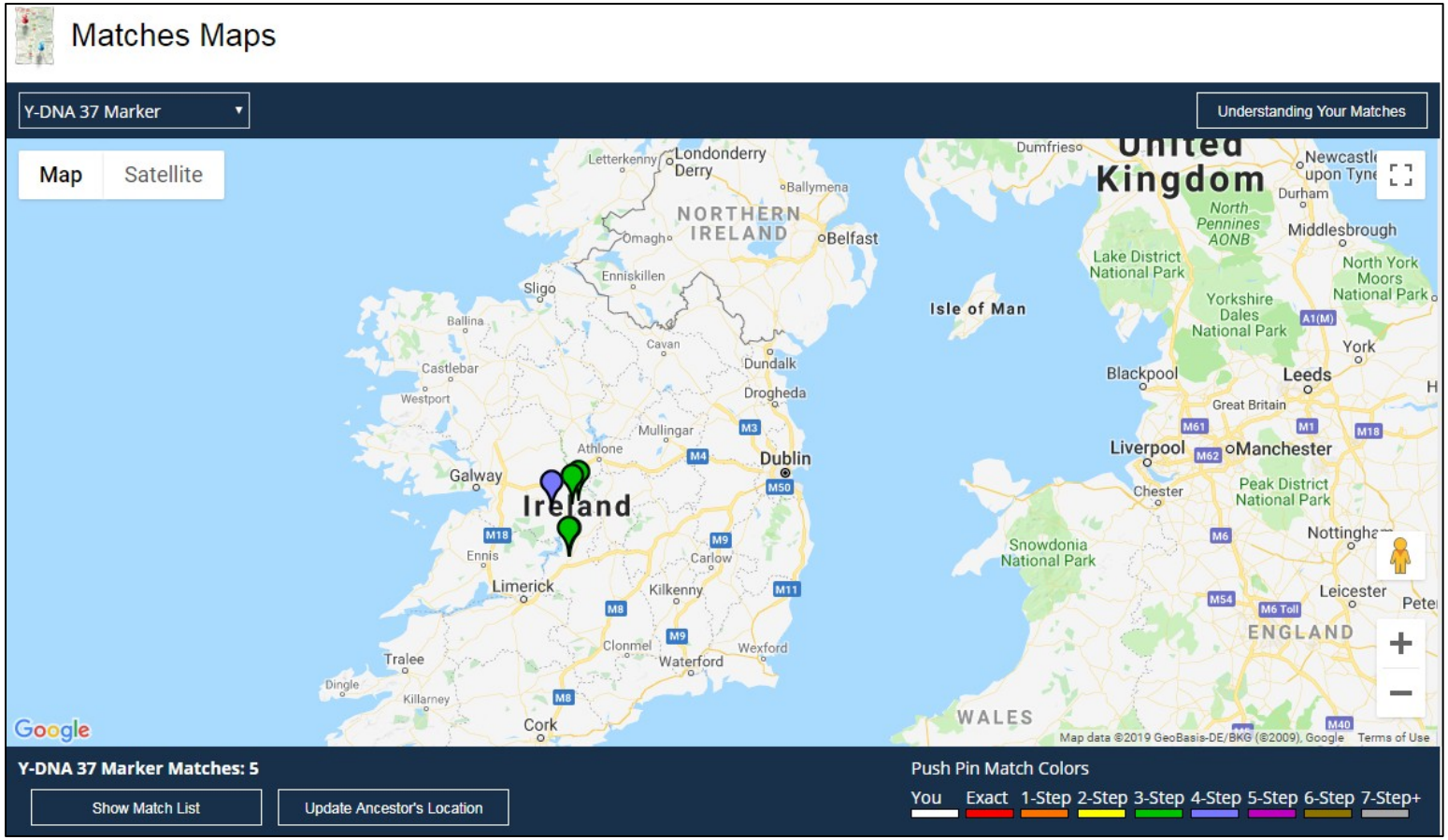
- Adaptation and Extension of Alex Williamson's YTree.net site.
  - Shows phylogenetic tree path
  - Shows context of your terminal SNP markers
  - Gives a sense of time (TMRCA) based on block-length being proportional to number of mutations on that branch.



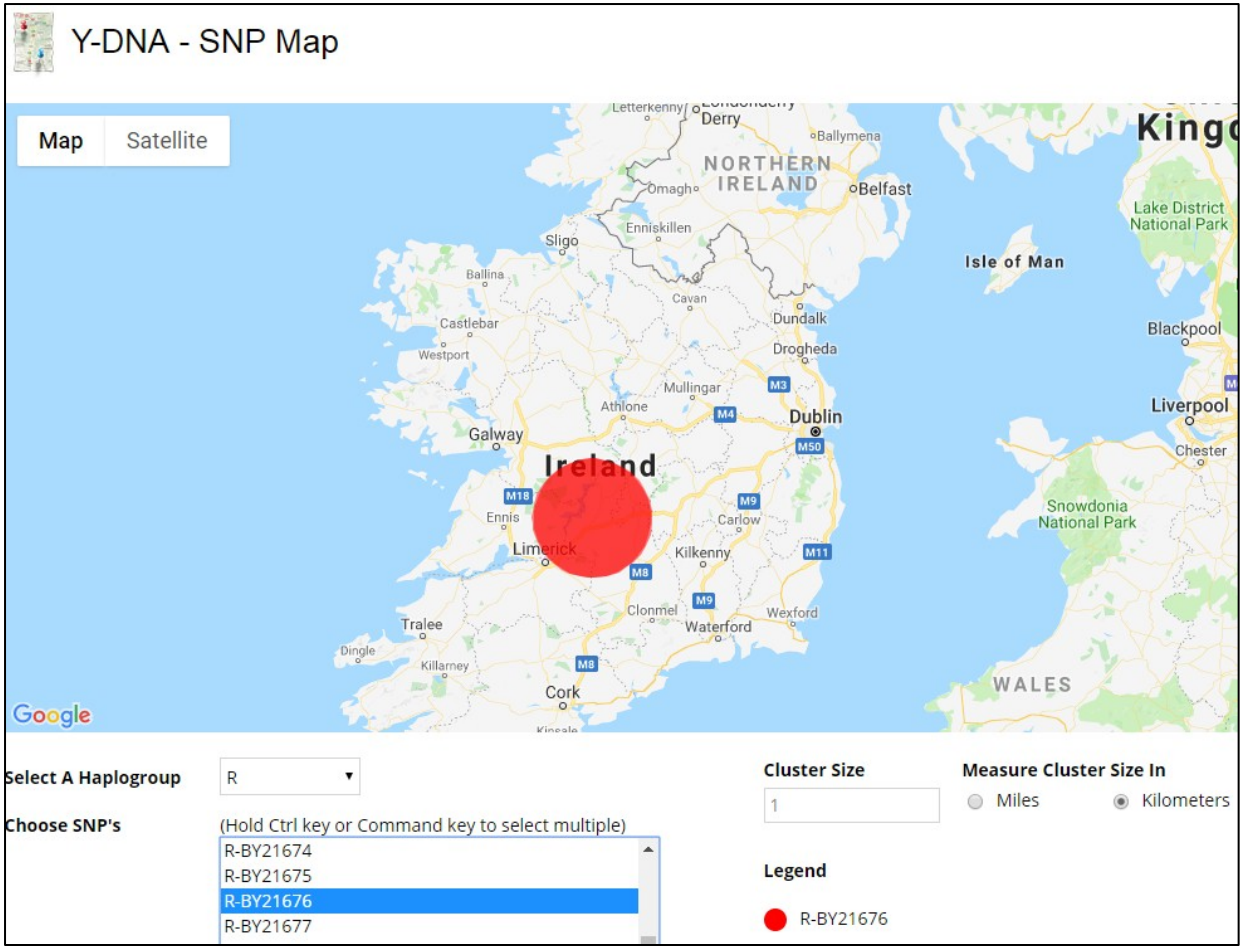
Why get the Big Y DNA Test (2019 Brad Larkin)



Matches Map

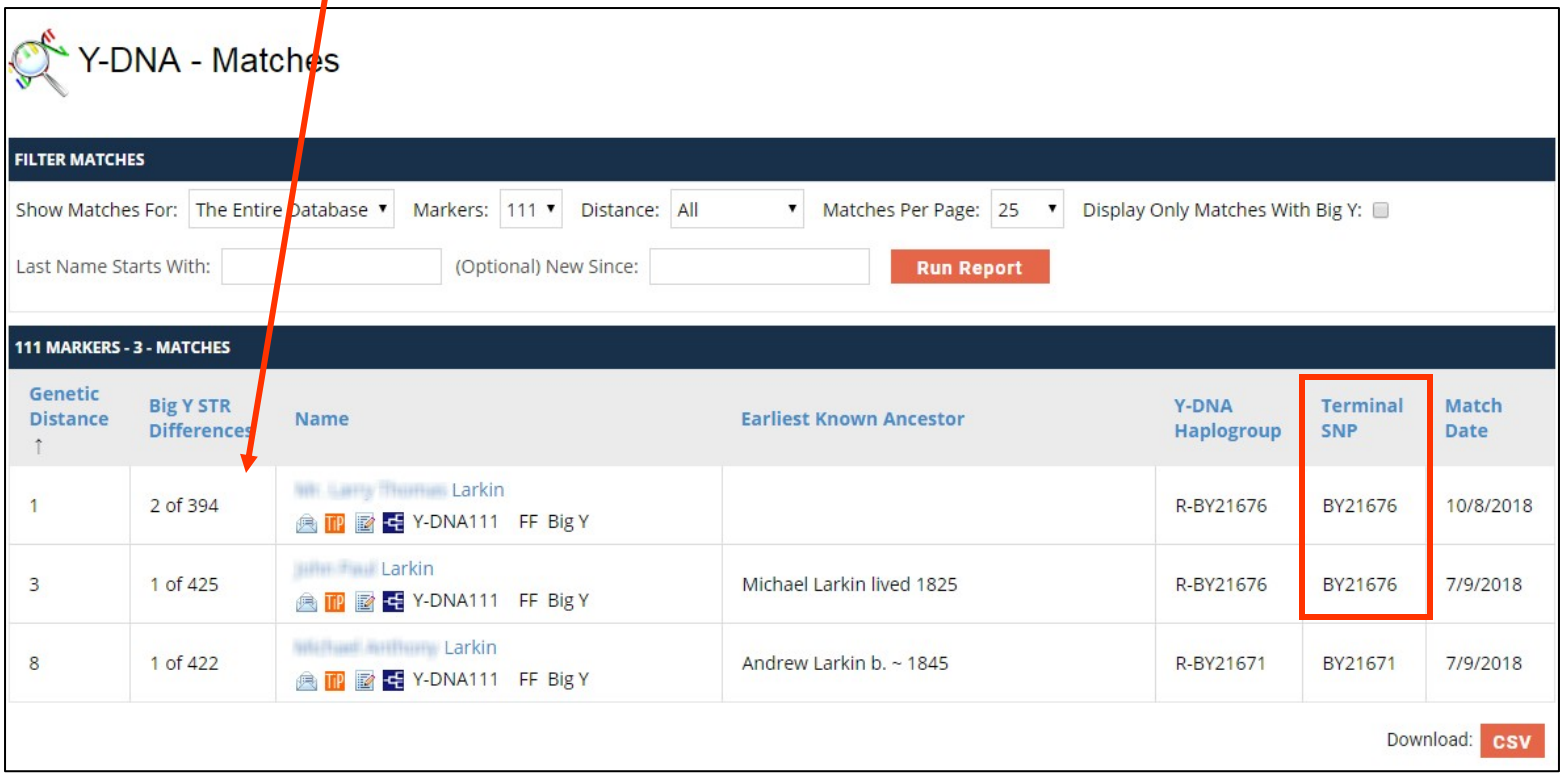

Matches Maps
Y-DNA 37 Marker
Understanding Your Matches
Map
Satellite
Ireland
NORTHERN IRELAND
Isle of Man
United Kingdom
Dublin
Belfast
Londonderry
Derry
Galway
Limerick
Cork
Dundalk
Drogheda
Athlone
Mullingar
Sligo
Enniskillen
Omagh
Ballina
Castlebar
Westport
Tralee
Dingle
Killarney
Ennis
Carlow
Kilkenny
Clonmel
Waterford
Wexford
Liverpool
Manchester
Leeds
Blackpool
Chester
Newcastle upon Tyne
Durham
Middlesbrough
Dumfries
Lake District National Park
North Pennines AONB
Yorkshire Dales National Park
North York Moors National Park
York
Great Britain
Peak District National Park
Snowdonia National Park
Leicester
ENGLAND
WALES
Google
Map data ©2019 GeoBasis-DE/BKG (©2009), Google
Terms of Use
Y-DNA 37 Marker Matches: 5
Show Match List
Update Ancestor's Location
Push Pin Match Colors
You
Exact
1-Step
2-Step
3-Step
4-Step
5-Step
6-Step
7-Step+

(Terminal) SNP Map



SNP & STR Match Integration

- When you have a Big Y test, will now include Big-Y STR mutational differences.



Y-DNA - Matches

FILTER MATCHES

Show Matches For: The Entire Database | Markers: 111 | Distance: All | Matches Per Page: 25 | Display Only Matches With Big Y:

Last Name Starts With: | (Optional) New Since: | Run Report

111 MARKERS - 3 - MATCHES

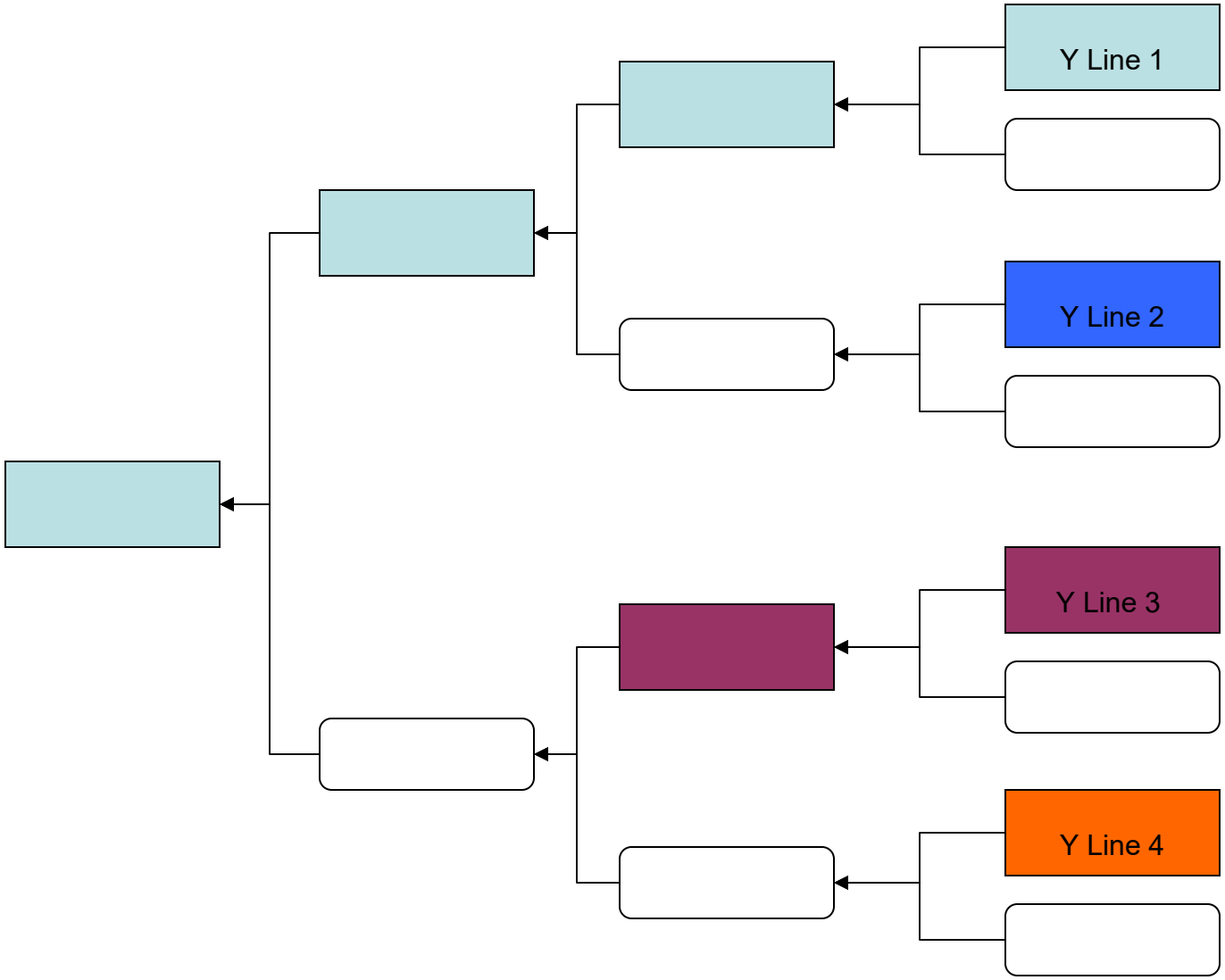
Genetic Distance ↑	Big Y STR Differences	Name	Earliest Known Ancestor	Y-DNA Haplogroup	Terminal SNP	Match Date
1	2 of 394	[redacted] Larkin Y-DNA111 FF Big Y		R-BY21676	BY21676	10/8/2018
3	1 of 425	[redacted] Larkin Y-DNA111 FF Big Y	Michael Larkin lived 1825	R-BY21676	BY21676	7/9/2018
8	1 of 422	[redacted] Larkin Y-DNA111 FF Big Y	Andrew Larkin b. ~ 1845	R-BY21671	BY21671	7/9/2018

Download: CSV

Non-Paternal Ancestry

- While your own father's Y-DNA results define your patrilineal ancestry, we can also think of our other ancestral lines as being collections of those Phylogenetic Tree lineages.
  - Testing a male relative from each of your 4 great grandfather's Y lineages is likely to yield much new insight to your family knowledge and lore.

Big Y all your Paternal Lineages



Flipping the Surname Identification Paradigm

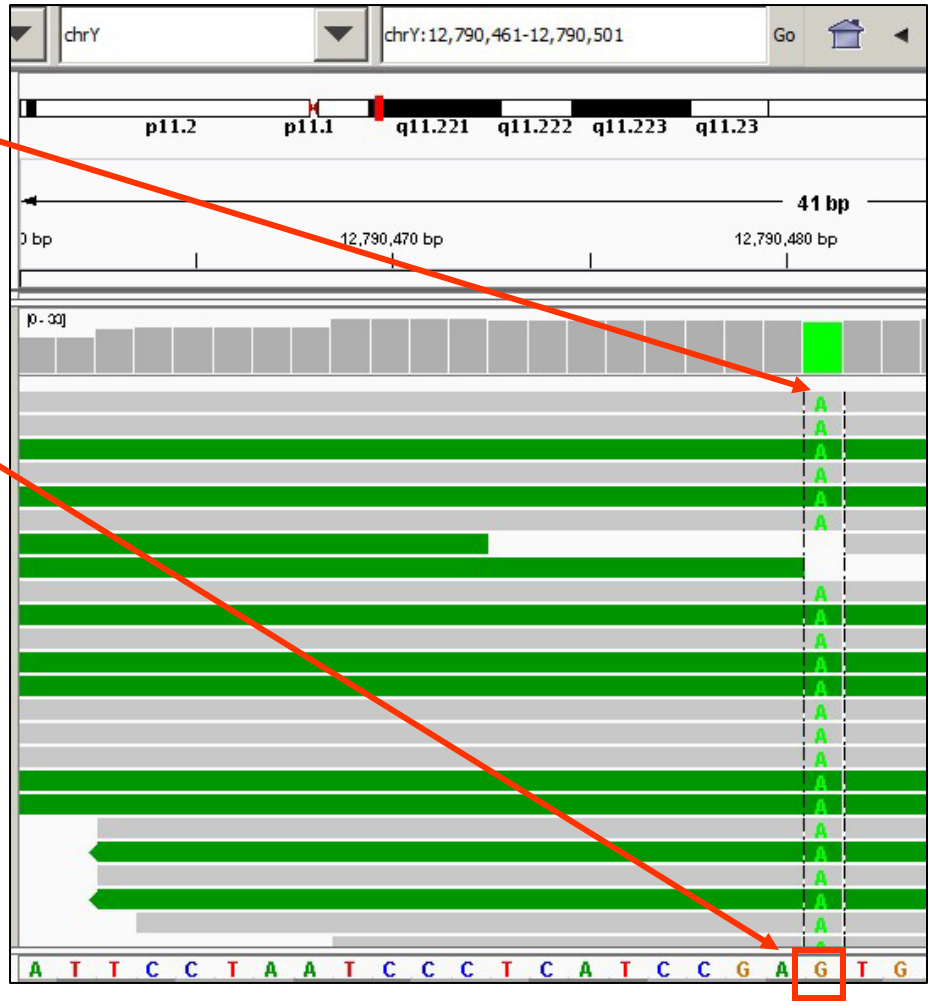
- Rather than thinking about, “I have surname [xxxx] and that must be my paternal ancestry”
- Instead, consider thinking about, “I have SNP markers [A], [B], [C], [D] and so those DEFINE my paternal lineages.

Advanced: BAM Viewer

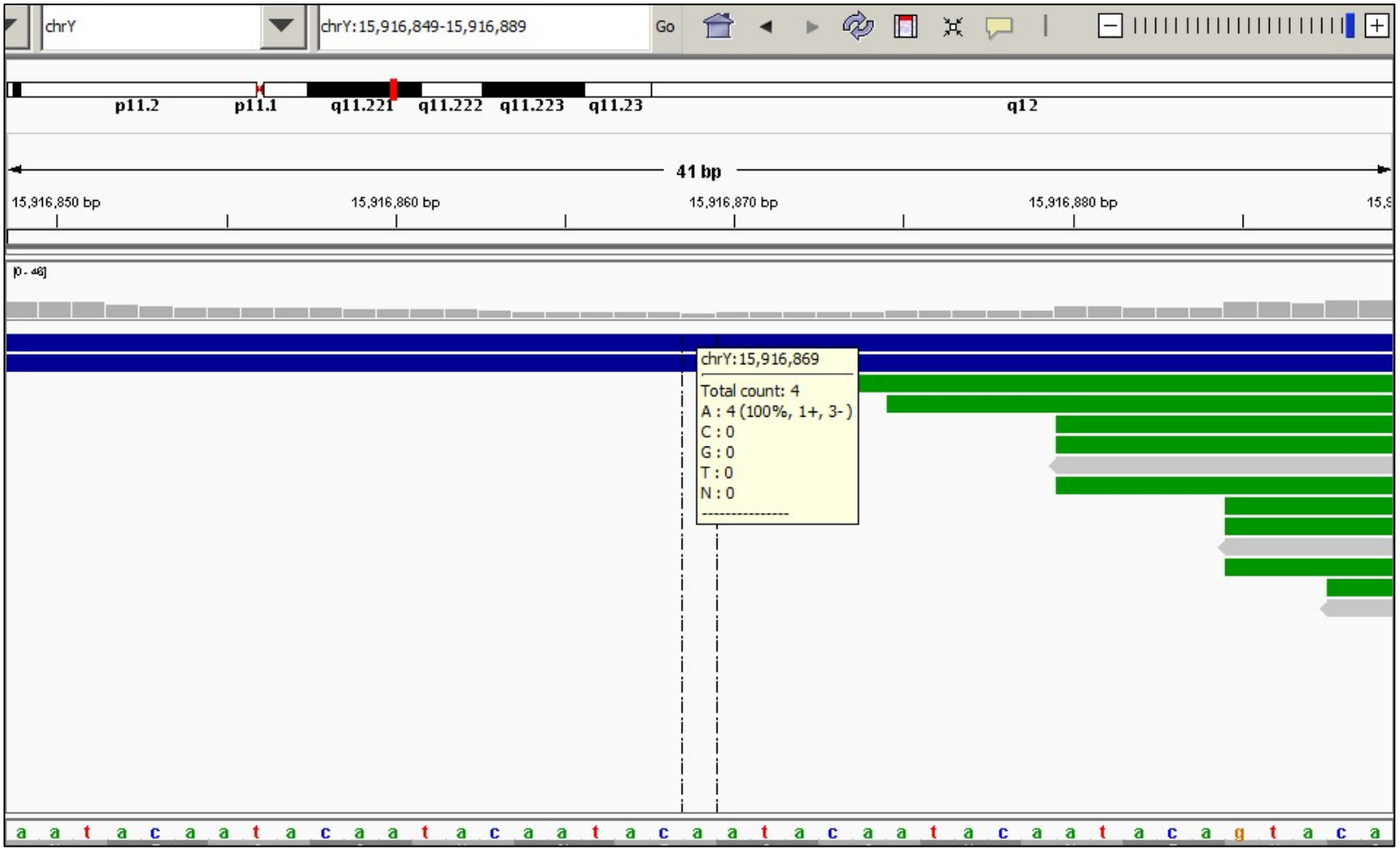
- Downloading your raw BAM data file of Big Y or comparable data from your testing laboratory.
- Download a free BAM file viewer.
  - http://software.broadinstitute.org/software/igv/download
- Can also compare results from multiple laboratories at the same time.
- Can compare results from multiple individuals at same time.
  - e.g. close family members or known cousins
- When using a BAM Viewer or chromosome browser, one must often use a nomenclature like *assembly:chromsome:startposition..endposition*
  - "hg38:chrY:2844421.. 2844421"
  - This is why cross-references like the DNA Marker Index are helpful for comparing lab results to the BAM data file.

What an SNP Looks Like in BAM Viewer

- The green letter "A"s indicate a read that is different from the reference sequence ("G")
- Where we see this mutation occur in one group of people and not occur in another, it is thus an SNP for genetic genealogy.

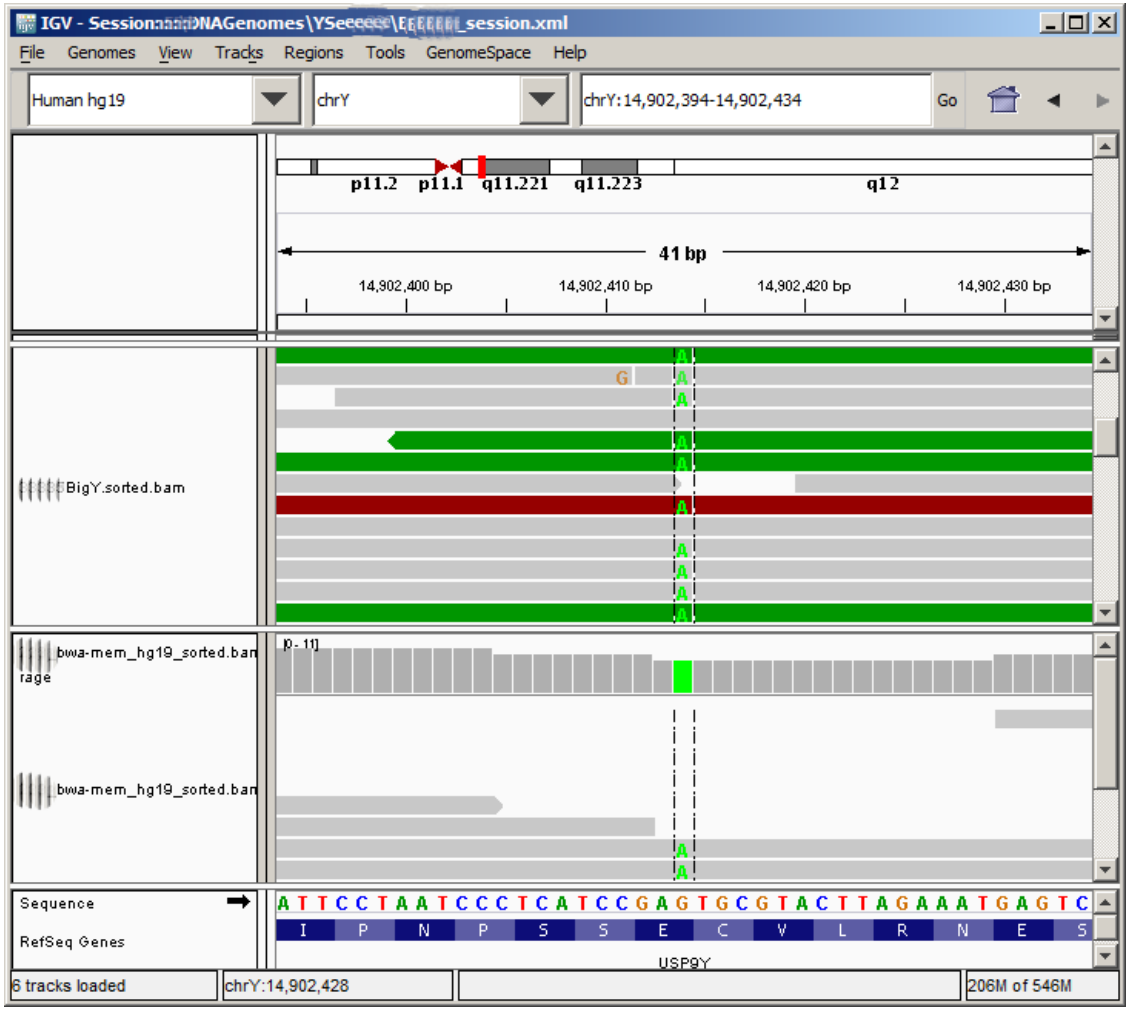


What a No-Call Looks Like in BAM Viewer



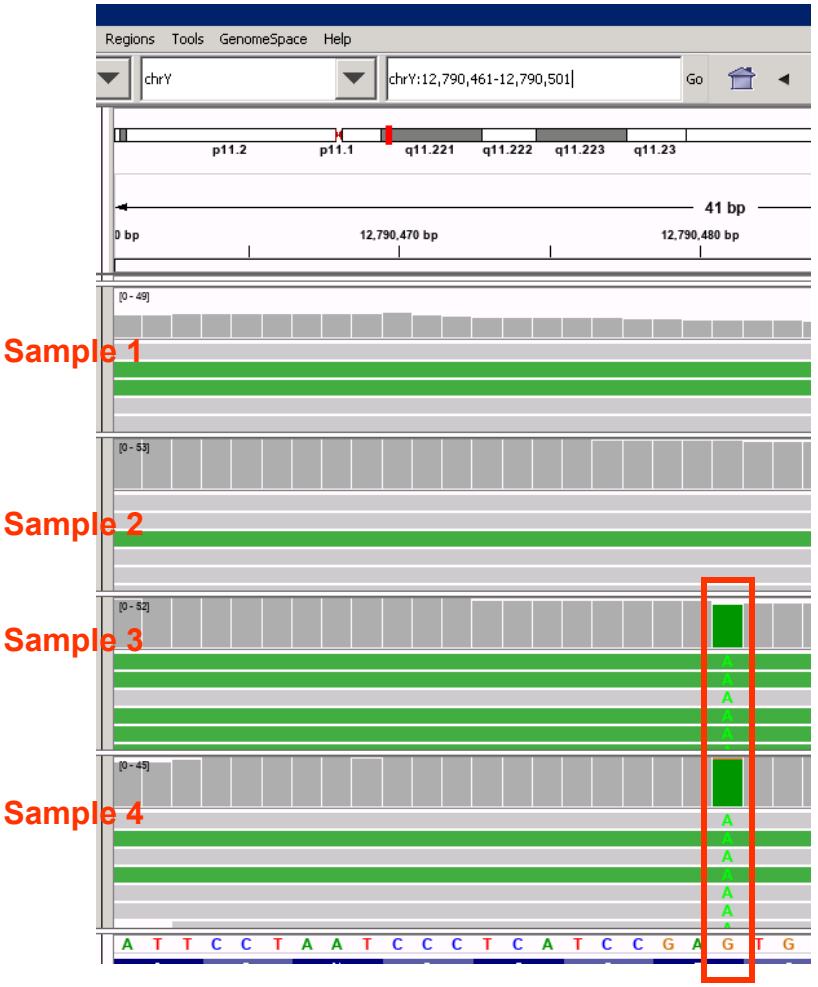
Shotgun Coverage - BAM

- The individual reads are stored in a binary format called a 'BAM' file.
  - Example from real BAM file from Next-Gen sequencing.
    - G->A mutation at Y-14902414 = M222



BAM: Comparing Multiple Kits

- One of the advantages of using a BAM viewer is that we can visually compare multiple samples at once.
  - In this example we see 4 samples compared.
  - The lower two (#s 3 and 4) having the *A* mutation at position 12790481 (M222)



Summary

- The most important reason to get the Big Y-700 DNA test is that the results will locate your family on the emerging phylogenetic tree of mankind in ways that are both distinct yet provide a measurable degree of paternal kinship to others, regardless of an individual's surname.

Surname-Related Research Questions

- Y-DNA focused, surname-related research questions can include:
  - Classifying worldwide linkages in diaspora populations
    - All Larkin's living today in Shannon River Valley in Ireland
    - Ashkenazi descendants of a particular 18th century Eastern European Rabbi
  - Connecting American families with common surnames to colonial roots
    - Relationship of all Reynolds families living in Texas in 1860.
  - Connecting genealogical lineages for surnames which have highly-variable spelling
    - All Robinson / Robertson / Roberson families living in Charleston South Carolina area today.
      - Y-DNA can go to much higher resolution than surname spelling.

Y-DNA Genealogy Transition

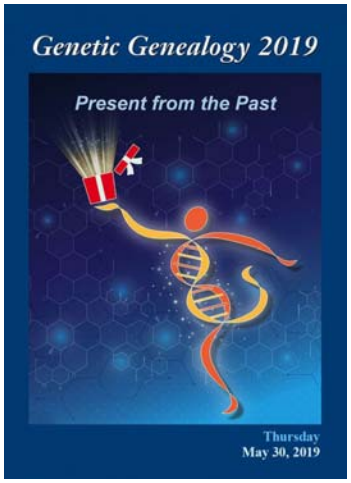
	Old Way	New Way
Matching	Compare 37 marker STRs and then use probabilities to estimate TMRCA	Millions of SNPs in combination with emerging ancestral tree.
Testing Strategy	Order minimalists tests and then embark on a series of upgrades & guesses.	Buy the Big Y once, then let matches and tree develop for you.
Surname Dependency of Results Interpretation	Guess that common surname & STR similarity were proxies for having SNP match.	Get definitive SNP matches and an ancestral tree, regardless of surname.

The Answer

- The Y-DNA Phylogenetic Tree
  - The most important reason to get the Big Y DNA test is that the results locate your family on the phylogenetic tree of mankind.
    - Distinctness
      - With millions of markers tested, every family will have unique marker(s)
    - Defined Kinship
      - Kinship between Y lineages on the tree is clear and getting more refined as more samples are tested.
      - Based on absolutes rather than probabilities.
        - » Either the two samples have the same mutation, or they do not.

Audience Questions

Why get the Big Y DNA Test?



Brad Larkin

Southern California Genealogical Society
Jamboree 2019